CRC Press
Taylor & Francis Group

□ 应用统计学丛书 23

Quantitative Trading:
Algorithms, Analytics, Data, Models, Optimization

量化交易:
算法、分析、数据、模型和优化

Xin Guo
Tze Leung Lai
Howard Shek
Samuel Po-Shing Wong 著
冯玉林 刘庆富 译

高等教育出版社·北京

infoway.io

实时金融行情数据接口

低延迟 • 高可用 • 全覆盖 **专为量化交易设计的行情API**

全市场覆盖 一个接口查所有

股票API
涵盖A股、港股、美股等多个市场。

外汇API
包含56组主要货币对，高频更新。

加密货币API
超过100个币种，全天24小时实时报价。

商品期货API
贵金属、能源天然气、期货指数等

Infoway API是一款专门为量化交易设计的行情接口。支持逐笔成交/订单簿快照等颗粒度，兼容RESTful API和WebSocket，满足高频量化、资产管理系统、行情终端等场景需求。

全域市场一网打尽

每日处理4300万条市场数据流，开发者无需对接多个数据源即可构建多资产策略系统。

简单易接入

我们的API设计注重易用性，支持多种编程语言，确保无论您使用什么开发环境，都能快速将实时行情数据集成到您的应用中。

免费试用！

注册后自动获得试用资格，无需实名认证。（**本活动长期有效**）

联系我们

官网地址：https://infoway.io
客服邮箱：hi@infoway.io



实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

图字：01-2017-5458 号

Quantitative Trading: Algorithms, Analytics, Data, Models, Optimization, 1st Edition, authored/edited by Xin Guo, Tze Leung Lai, Howard Shek, Samuel Po-Shing Wong

© 2017 by Taylor & Francis Group, LLC
Authorized translation from the English language edition published by Chapman and Hall/CRC, an imprint of Taylor & Francis Group LLC. All Rights Reserved. 本书原版由 Taylor & Francis 出版集团旗下 Chapman and Hall/CRC 出版公司出版，并经其授权翻译出版。版权所有，侵权必究。

Higher Education Press Limited Company is authorized to publish and distribute exclusively the **Chinese (Simplified Characters)** language edition. This edition is authorized for sale throughout **Mainland of China**. No part of the publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher. 本书中文简体翻译版授权由高等教育出版社有限公司独家出版并限在中国大陆地区销售。未经出版者书面许可，不得以任何方式复制或发行本书的任何部分。

Copies of this book sold without a Taylor & Francis sticker on the cover are unauthorized and illegal. 本书封面贴有 Taylor & Francis 公司防伪标签，无标签者不得销售。

图书在版编目（CIP）数据

量化交易：算法、分析、数据、模型和优化 / 郭新等著；冯玉林，刘庆富译. -- 北京：高等教育出版社，2020.1

书名原文：Quantitative Trading: Algorithms, Analytics, Data, Models, Optimization

ISBN 978-7-04-051508-4

Ⅰ. ①量… Ⅱ. ①郭… ②冯… ③刘… Ⅲ. ①股票交易 Ⅳ. ① F830.91

中国版本图书馆 CIP 数据核字（2019）第 040840 号

策划编辑 李华英　　责任编辑 李华英 李 鹏　　封面设计 张雨微　　版式设计 范晓红
责任校对 马鑫蕊　　责任印制 尤 静

出版发行	高等教育出版社	网　址	http://www.hep.edu.cn
社　址	北京市西城区德外大街4号		http://www.hep.com.cn
邮政编码	100120	网上订购	http://www.hepmall.com.cn
印　刷	涿州市星河印刷有限公司		http://www.hepmall.com
开　本	787mm×1092mm　1/16		http://www.hepmall.cn
印　张	22.25		
字　数	420 千字	版　次	2020 年 1 月第 1 版
购书热线	010-58581118	印　次	2020 年 1 月第 1 次印刷
咨询电话	400-810-0598	定　价	89.00 元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换
版权所有　侵权必究
物 料 号　51508-00

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

应用统计学丛书

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

致我们爱的人

Xin Guo: 感谢我的丈夫 Ravi Kumar 在我与合作者日夜工作时的支持和理解, 感谢我的儿子 Raman 和 Kannan 为我带来的快乐, 同时谨以此书纪念我的父母。

Tze Leung Lai: 感谢我的妻子 Letitia 基于自己在银行、投资和分析方面的丰富业界经验以及对于数据和模型的热爱, 为本书做出的卓越贡献。

Howard Shek: 感谢我的父母和已故的祖父母在这些年来的坚定支持, 感谢我的教女 Isabelle 在每次见面时为我带来的无尽快乐。

Sam Wong: 感谢我的母亲 S. C. Lee、我的妻子 Fanny 和我的儿子 Theodore 对我的爱与支持。

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

前　言

在经历了 2007—2008 年的金融危机以及 2009 年的大衰退这一段混乱的时期之后, 金融行业进入了一个新时代。量化投资策略, 连同用于其开发和执行的统计模型和方法、知识表达和数据分析、以及算法和信息, 在这个新时代变得日益重要。这个新时代的开端以两项改变了现代生活和商业的 “革命” 为标志。其中一项是技术方面的,《经济学人》杂志 2015 年 5 月 9 日刊文称该变革为面向金融服务的 “金融科技革命”, 该文写道: “自 2007—2008 年金融危机以来, 决策者开始致力于增强金融的安全性。在监管机构的关注焦点之外, 另一项革命正在发生, 从支付到理财, 从 P2P 借贷到众筹, 一批新的创业公司瞄准了金融行业的核心。根据高盛集团测算, 这个领域的收益价值 4.7 万亿美元。像其他规则破坏者一样, 金融科技公司的发展十分迅速。” 另一项革命被称为 “大数据革命”。2014 年 8 月, 联合国秘书长任命了一个独立顾问团就 “发起可持续发展的数据革命” 提出建议。《哈佛商业评论》2012 年 10 月刊发表了一篇题为 “大数据: 管理改革” 的文章。2015 年 8 月 20 日, 中国国家总理要求不同的政府部门共享数据并执行大数据操作计划。紧接着中国国务院又于 2015 年 9 月 5 日发布了在经济规划、金融、国土安全、交通、农业、环境和医疗等领域发展和推广大数据应用的行动计划。

在这个新时期, 大数据和金融科技革命所带来的机遇与挑战深深地吸引了学生们, 对此, 作者团队里恰好在硅谷地区任教的两位学者 (Xin Guo 和 Tze Leung Lai) 在过去的三年里在斯坦福大学和加州大学伯克利分校开发并教授了金融工程和金融数学课程, 并更新了课程的教材。他们还邀请了业界人士,

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

尤其是本书的另外两位合著者 (Howard Shek 和 Samual Wong) 为这些课程举行客座演讲和研讨会。这些日常的合作迅速转化成写作本书的紧密的合作。本书的素材不仅可用于有效地教授相关课程, 也可用于相关的简短课程和培训项目。目前, 上海高级金融学院、清华大学、香港中文大学、香港科技大学、新加坡国立大学、台湾大学、首尔国立大学等机构已经在课程和培训中使用本书。在斯坦福大学, 这些课程的先修课是 STATS 240 (金融统计方法), 涵盖了 Lai 和 Xing (2008) 一书的前六章内容。因此, 我们在本书中对 Lai 和 Xing (2008) 一书的前六章内容进行了大量的引用, 仅对主要结论进行归纳, 以避免细节上的重复。

本书的网址为 http://lait.web.stanford.edu/quantstratbook/。书中的练习和案例的数据可以从该网站下载。我们在本书中强调了算法交易和量化投资策略的开发与执行的跨学科方法。该方法横跨计算机科学与工程、金融与经济、数学与统计、法律等多个学科。跨学科特征不仅反映在加州大学伯克利分校的金融工程和斯坦福大学金融数学项目 (最近已变成更广泛的数学和计算金融项目) 所提供的课程中, 以进一步强调数据科学、统计建模、高级编程、高速计算, 还体现在斯坦福大学新成立的金融与风险建模研究中心 (Financial and Risk Modelling Institute, FARM) 所开展的研究活动中。除了跨学科性之外, 本书的另一大特点是弥合学术研究和教育与金融实务之间的差距, 这也是 FARM 的使命之一。本书的部分内容也可以用于面向从业人员的短期课程, FARM 正在开发这些短期课程。

Xin Guo, Tze Leung Lai, Howard Shek, Samuel Wong

2017 年 3 月

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

目　　录

第一章　概论 **1**

1.1　交易基础结构的演变 1

1.2　量化交易策略及时间跨度 5

1.3　统计套利及关于有效市场假说的争论 6

1.4　量化基金、公募基金、对冲基金以及基金表现 7

1.5　数据、分析、模型、优化和算法 9

1.6　量化交易的跨学科性及本书的使用方法 10

1.7　补充材料和习题 12

第二章　量化交易中的统计模型与方法 **16**

2.1　股票价格的特征 17

2.1.1　低频交易数据的时间序列分析 17

2.1.2　高频交易数据的离散化价格变动 17

2.2　巴黎证交所的布朗运动和华尔街的随机游走 20

2.3　现代投资组合理论是在有效市场假说下的华尔街随机漫步靴 . . 21

2.4　现代投资组合理论的统计学基础 22

2.4.1　多因子定价模型 23

2.4.2　贝叶斯方法、缩减技术及 Black-Litterman 估计方法 23

2.4.3　Bootstrap 方法和重抽样方法前沿 25

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

2.5 一种考虑了参数不确定性的新方法 · 25
2.5.1 优化问题的求解 · 26
2.5.2 最优权重向量的计算 · 27
2.5.3 收益率的 Bootstrap 估计和 NPEB 法则 · · · · · · · · · · · · · · · · · · 28
2.6 从随机游走模型到更符合事实的鞅模型 · · · · · · · · · · · · · · · · · · · 29
2.6.1 从高斯形式到帕累托形式的随机游走 · · · · · · · · · · · · · · · · · · 30
2.6.2 包含最优抽样时间的随机游走 · 31
2.6.3 从随机游走模型到 ARIMA、GARCH 和鞅回归等时间序列模型 · · · · 34
2.7 新现代投资组合理论下的鞅回归模型 · 37
2.7.1 在 NPEB 中加入时间序列效应 · 37
2.7.2 有效前沿上的最优信息比率 · 38
2.7.3 新现代投资组合理论的实证检验 · 38
2.8 有效市场假说之外的统计套利和策略 · 40
2.8.1 非参数回归和变点模型中的技术准则和统计背景 · · · · · · · · · · · 41
2.8.2 时间序列、动量策略和配对交易策略 · · · · · · · · · · · · · · · · · · 42
2.8.3 逆向投资策略、行为金融及投资者的认识偏差 · · · · · · · · · · · · 43
2.8.4 从价值投资到全球宏观策略 · 44
2.8.5 投资策略的样本内评估和样本外评估以及多重检验方法 · · · · · · · 44
2.9 补充材料和习题 · 45

第三章 积极型投资组合管理和投资策略 · · · · · · · · · · · · · · · · · · · 57
3.1 积极型投资组合管理中的 α 和 β · 58
3.1.1 超额收益率 α 的来源 · 59
3.1.2 积极的 α 之外的奇异 β · 59
3.1.3 一种积极型投资组合优化的新方法 · · · · · · · · · · · · · · · · · · · 60
3.2 交易成本和买卖限制 · 62
3.2.1 交易成本的组成 · 62
3.2.2 买卖限制和其他约束条件 · 63
3.3 多期投资组合管理 · 64
3.3.1 Samuelson-Merton 基于随机控制理论的 “终身投资组合选择理论” · 65
3.3.2 包含了交易成本的 Merton 问题 · 67
3.3.3 多期资本增长和波动率脉动 · 68
3.3.4 多期均值–方差投资组合再平衡 · 69
3.3.5 存在交易成本情况下的动态均值–方差投资组合优化 · · · · · · · · · 70
3.3.6 参数不确定性情况下的动态投资组合选择 · · · · · · · · · · · · · · · 72
3.4 补充材料和注释 · 74

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

3.5 习题 · · · · · 96

第四章　电子交易中的计量经济学 · · · · · 97

4.1 电子交易和交易数据 · · · · · 98
4.2 高频数据建模 · · · · · 99
4.2.1 Roll 提出的买入–卖出反弹模型 · · · · · 99
4.2.2 包含噪声项的市场微观结构模型 · · · · · 99
4.3 X_t 的积分方差的估计 · · · · · 101
4.3.1 稀疏抽样方法 · · · · · 102
4.3.2 基于子样本的平均法 · · · · · 103
4.3.3 两种时间尺度的方法 · · · · · 103
4.3.4 核平滑方法: 已实现核估计量 · · · · · 104
4.3.5 预平均方法 · · · · · 105
4.3.6 从波动率的 MLE 估计量到 $[X]_T$ 的 QMLE 估计量 · · · · · 106
4.4 多资产协变差估计 · · · · · 107
4.4.1 异步性和 Epps 效应 · · · · · 107
4.4.2 同步技术 · · · · · 108
4.4.3 协方差和相关系数估计的 QMLE 方法 · · · · · 109
4.4.4 多元已实现核和双尺度的估计量 · · · · · 110
4.5 傅里叶方法 · · · · · 111
4.5.1 $[X]_T$ 和瞬时波动率的傅里叶估计 · · · · · 112
4.5.2 傅里叶估计量的统计特征 · · · · · 113
4.5.3 瞬时协方差的傅里叶估计量 · · · · · 115
4.6 与 TAQ 相关的其他计量经济学模型 · · · · · 116
4.6.1 交易之间间隔的 ACD 模型 · · · · · 117
4.6.2 自激励点过程模型 · · · · · 118
4.6.3 D_i 的分解和广义线性模型 · · · · · 119
4.6.4 McCulloch 和 Tsay 分解方法 · · · · · 120
4.6.5 点过程和其标值的联合建模 · · · · · 121
4.6.6 结合低频波动率和高频波动率的已实现 GARCH 等预测模型 · · · · 122
4.6.7 有效价格过程中的跳跃和幂变差 · · · · · 124
4.7 补充材料和评论 · · · · · 125
4.8 习题 · · · · · 133

第五章　限价指令簿: 数据分析和动态模型 · · · · · 136

5.1 从市场数据到限价指令簿 (LOB) · · · · · 137

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

5.2 LOB 数据的特征 · · · · · 138
5.2.1 订单簿的价格调整机制 · · · · · 138
5.2.2 交易量不平衡和其他指示量 · · · · · 140
5.3 LOB 数据的多元点过程拟合 · · · · · 142
5.3.1 市场化订单的多元点过程 · · · · · 143
5.3.2 实证结果 · · · · · 145
5.4 基于机器学习的 LOB 数据分析 · · · · · 149
5.5 LOB 动态的排队论建模 · · · · · 151
5.5.1 价格水平为 1 的 LOB 数据的简化形式模型的扩散极限 · · · · · 151
5.5.2 订单位置的流动极限 · · · · · 154
5.5.3 基于 LOB 的序列反应模型 · · · · · 158
5.6 补充材料和习题 · · · · · 161

第六章 最优执行与配置 · · · · · 172
6.1 单资产最优执行问题 · · · · · 173
6.1.1 问题 (6.2) 的动态规划求解 · · · · · 174
6.1.2 连续时间模型和变分法 · · · · · 176
6.1.3 最优确定性策略 · · · · · 178
6.2 乘数价格影响模型 · · · · · 179
6.2.1 模型设定和随机控制问题 · · · · · 179
6.2.2 有限时间情形下的 HJB 方程 · · · · · 180
6.2.3 无限时间情形 $T=\infty$, 且 $0\leqslant\mu<\delta$, $C_{\mathrm{s}}>0$ · · · · · 182
6.2.4 价格操纵与传导性价格影响 · · · · · 184
6.3 基于 LOB 数据的最优执行问题 · · · · · 184
6.3.1 成本最小化 · · · · · 187
6.3.2 模型 1 的最优策略 · · · · · 190
6.3.3 模型 2 的最优策略 · · · · · 191
6.3.4 块状 LOB 的解析解 · · · · · 192
6.4 投资组合的最优执行 · · · · · 192
6.5 最优配置 · · · · · 195
6.5.1 包含均值回归的马尔可夫随机游走模型 · · · · · 196
6.5.2 连续时间马尔可夫链模型 · · · · · 199
6.6 补充材料和习题 · · · · · 203

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

第七章　做市和智能订单路径 · · · · · 209
7.1 Ho 和 Stoll 模型以及 Avellanedo-Stoikov 决策 · · · · · 210
7.2 HJB 方程的求解和后续的拓展 · · · · · 211
7.3 包含限制和市场订单的脉冲控制 · · · · · 213
7.3.1 做市商的脉冲控制 · · · · · 213
7.3.2 最优限价和市价订单策略的控制公式 · · · · · 214
7.4 智能订单路径和暗池 · · · · · 216
7.5 基于交易所 SOR 的最优订单分拆 · · · · · 218
7.5.1 成本函数与优化问题 · · · · · 218
7.5.2 K 个交易所的最优下单策略 · · · · · 220
7.5.3 一个随机逼近方法 · · · · · 220
7.6 暗池的删失探索和开发 · · · · · 221
7.6.1 贪心配置算法 · · · · · 221
7.6.2 T_i 的改进 Kaplan-Meier 估计量 $\hat{T}_i^t$ · · · · · 222
7.6.3 探索、开发和 ε 最优配置 · · · · · 224
7.7 暗池中的随机拉格朗日优化 · · · · · 225
7.7.1 基于随机近似的拉格朗日方法 · · · · · 226
7.7.2 拉格朗日递归到优化函数的收敛性 · · · · · 227
7.8 补充材料和注释 · · · · · 228
7.9 习题 · · · · · 235

第八章　信息学、监管和风险管理 · · · · · 237
8.1 量化策略 · · · · · 238
8.2 交易基础设施 · · · · · 240
8.2.1 订单网关 · · · · · 242
8.2.2 匹配引擎 · · · · · 243
8.2.3 市场数据传播 · · · · · 244
8.2.4 订单的收费结构 · · · · · 245
8.2.5 主机托管服务 · · · · · 246
8.2.6 清算和结算 · · · · · 246
8.3 策略信息和基础设施 · · · · · 247
8.3.1 市场数据处理 · · · · · 247
8.3.2 Alpha 引擎 · · · · · 249
8.3.3 订单管理 · · · · · 249
8.3.4 订单类型和订单限定条件 · · · · · 250

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

8.4 交易所规章制度 · · · · · 252
8.4.1 证券信息处理程序和国家市场体系管理规则 · · · · · 252
8.4.2 证券卖空规则 · · · · · 255
8.4.3 其他交易规则 · · · · · 256
8.4.4 熔断机制 · · · · · 256
8.4.5 市场操纵 · · · · · 256
8.5 风险管理 · · · · · 257
8.5.1 操作风险 · · · · · 257
8.5.2 策略风险 · · · · · 259
8.6 补充材料和注释 · · · · · 260
8.7 习题 · · · · · 269

附录 A 鞅理论 · · · · · 274
A.1 离散时间鞅 · · · · · 274
A.2 连续时间鞅 · · · · · 277

附录 B 马尔可夫链 · · · · · 282
B.1 CTMC 生成元 $\boldsymbol{Q}$ · · · · · 282
B.2 马尔可夫链的势论 · · · · · 283
B.3 马尔可夫决策理论 · · · · · 283

附录 C 双随机自激点过程 · · · · · 285
C.1 鞅理论和多元计算过程的补偿 · · · · · 285
C.2 双随机点过程模型 · · · · · 286
C.3 点过程模型的似然估计 · · · · · 287
C.4 双重随机 SEPP 的模拟 · · · · · 290

附录 D 排队系统的弱收敛和极限定理 · · · · · 292
D.1 Donsker 定理及其扩展 · · · · · 293
D.2 排队系统和极限定理 · · · · · 294

参考文献 · · · · · 296

索引 · · · · · 329

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

插 图 目 录

图 1.1 芝加哥证券交易所的公开喊价系统中的手势说明 2

图 1.2 西联电报公司在 19 世纪 70 年代制造的证券报价机. 最初, 交易单上只有交易价格和交易公司的缩写, 到 20 世纪 30 年代以后, 交易额也会出现在交易单上. 该交易机目前在美国加州山景城的计算机历史博物馆展出. 照片来源: 维基共享资源/Don Debold 3

图 2.1 2014 年 10 月 10 日, 代码为 388 的股票的逐笔交易价格 P_t 的时序图 18

图 2.2 逐笔交易价格分布 (上图); 上午交易时段价差 (下图) 19

图 2.3 2014 年 10 月 10 日上午交易的代码为 388 的股票的交易价差的 ACF 函数. 虚线表示 5% 水平下的拒绝域临界值 20

图 2.4 2005 年 1 月 1 日—2014 年 12 月 31 日 Pfizer 公司收盘价 P_t 的时间序列图 (上图) 和 Δ_t 的时间序列图 (下图) 32

图 2.5 正态 QQ 图: $z_t^{(a)}$ (上图)、$z_t^{(b)}$ (下图), 其中 $z_t^{(a)}$ 和 $z_t^{(b)}$ 由 (2.18) 定义 33

图 2.6 Δ_t 的正态分布拟合结果的 QQ 图 (上图) 和对称稳定分布拟合结果的 QQ 图 (下图) 34

图 4.1 上图为根据模型 (4.21) 生成得到 $n = 1000$ 个样本的似然函数, 等价于模型 (4.69) 的生成样本似然函数图; 下图为根据 MA(1) 模型 (4.70) 的生成样本似然函数图 127

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

图 5.1 10 级的限价订单簿在英国石油公司 (BP) 的交易发生前 (灰色线) 和发生后 (黑色线) 演变过程的快照. 虚线表示最优买入和卖出价格. 实线表示中间价格的平均值. 在订单簿中, 柱状图按照最大序列长度进行比例缩减, 以深灰色和浅灰色表示, 便于区别订单簿在交易发生前后的变化 139
图 5.2 2010 年 6 月 25 日, BP 交易提交之后能在 5 秒内完成的概率. 方块表示基于实证数据的相对频率, 实线则为基于 Bouchaud 等 (2002) 建议的幂函数的拟合曲线 140
图 5.3 BP 公司在 2010 年 6 月 25 日的概率权重累积交易量之差 $\bar{v}(t,\tau,L;1)-\bar{v}(t,\tau,L;2)$ 的时间序列 141
图 5.4 在市场买入侧的新订单提交后, 20 毫秒至 500 毫秒时间范围内买入侧和卖出侧订单数的条件密度图. 数据来源为 2010 年 6 月 25 日 BP 公司的逐笔数据 144
图 5.5 2010 年 6 月 25 日, 订单簿中 BP 股票的最优现行买入价格 (上图) 和买入–卖出价差 (下图) 146
图 5.6 订单簿买方侧和卖方侧有限订单 (上图) 和市场订单 (下图) 的到达时间强度, 数据来源于 2010 年 6 月 25 日 BP 公司数据, 利用一分钟的重叠窗宽进行估计 146
图 5.7 四种模型对于 2010 年 6 月 25 日 BP 股票数据的拟合 QQ 图 . 148
图 5.8 最优买入序列中的订单位置 155
图 5.9 一元 Hawkes 过程的模拟强度 161
图 5.10 二元 Hawkes 过程的模拟强度, 其中上图中用 $\lambda^{(1)}$、下图中对 $\lambda^{(2)}$ 进行了反置以便查看 162

图 6.1 时刻 0 时最优策略中的 “等待区域” “全部卖出” 和 “部分卖出” 区域 183
图 6.2 当 $X_t=1$ 时对于 $1<t\leqslant T-1$ 的最优配制策略 198
图 6.3 当 $X_t=-1$ 时对于 $1<t\leqslant T-1$ 的最优配制策略 198

图 7.1 智能订单路径 216
图 7.2 第 k 家交易所限价订单的演化过程. 执行规模 L_k 由 $A_k=(\eta_k-Q_k)^+-(\eta_k-Q_k-L_k)^+$ 给出. 此处 L_k 被部分填充, 因此 $A_k=\eta_k-Q_k$ 219

图 8.1 交易支柱和策略支柱的图解 237

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

图 8.2 典型交易平台的组件 241
图 8.3 iLink 订单路由接口的原理图和 CME Globex 的匹配引擎. CGW1, CGW2, CGW3, CGW4 是 4 个网关节点; 交易引擎 1, 交易引擎 2, 交易引擎 3 是 3 个匹配的引擎. CGW1 正在处理 4 个会话: AB1:9000, AB2:9001, . . ., AB4:9003, 其中 9000 到 9003 是会话连接的端口. 会话 AB1 只能在一个网关上存在; AB1 有一个单一的序列流, 由来往所有市场的消息组成 (在 CME 的许可下使用) 242
图 8.4 包含两个 FIX 消息的 CME MDP3.0 包 (经 CME 许可使用) . . 245
图 8.5 典型策略基础设施的主要处理单元. 虚线箭头表示三个处理单元之间信息流的方向 247
图 8.6 买方激进交易的 MDP3.0 信息, 该交易出现在 LOB 卖出的三个级别中 (由 CME 授权使用) 262

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

表 格 目 录

表 2.1 四种准则下由 $m = 4$ 个资产组成的投资组合的收益 29
表 2.2 S&P 500 作为基准投资组合时, 五种投资组合的信息比率和平均已实现超额收益率 40
表 3.1 价值加权的投资组合的年平均已实现收益率的均值和标准差 . . 61
表 3.2 S&P 500 指数的各部分及其在 2016 年 10 月 17 日的权重 . . . 75
表 4.1 买入 150 股限价订单, 限价为 \$114.50, GTC, 产生 100 股在 \$114.50 价格上交易的限价订单, 而剩下的 50 股成为 LOB 中的最优买入股 98
表 5.1 (5.3) 中四种设定下的 MLE. 模型 1: $\bar{v}_{1t} = \bar{v}_{2t} \equiv 1$, 且 $w_{1i} = w_{2i} \equiv 1$; 模型 2: $w_{1i} = w_{2i} \equiv 1$; 模型 3: $\bar{v}_{1t} = \bar{v}_{2t} \equiv 1$; 模型 4: 参数 $\bar{v}_{1t}, \bar{v}_{2t}, w_{1i}, w_{2i}$ 不受约束. 标准差在括号中给出, † 指该值在 0.05 的置信水平下并不显著不等于 0 147
表 8.1 2015 年 11 月 29 日, 美国主要交易所、其主要交易产品及持仓量或日交易量. 例如, 那天苹果公司的交易量是 72056345 股, 而 ES 的持仓量是 2920713 份合约 240

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

表 8.2 在 CME Globex 上的十大算法交易期货类别占总量的百分比. 数字是基于交易指示器 (FIX Tag 1028). “自动” 指令由计算机算法提交, 无需人工干预; “人工” 指令是由人工干预所提交的交易; “其他” 指令包括在场内或场外进行的交易. 请参考 Haynes 和 Roberts (2015) 在 2012 年 11 月 12 日至 2014 年 10 月 31 日的芝加哥商品交易所交易数据集的完整分类和描述. 241
表 8.3 2015 年 10 月 5 日的收费结构. 括号里的费用意味着向流动性提供者提供回扣: 纽约证券交易所收取和提供显示流动性的费用; 纳斯达克通过单一的 MPID 获取流动性, 并提供显示指定的零售流动性; 从公司拿走至少 0.20% 的 TCV, 并提供显示和非显示的流动性, 从公司获得流动性的回扣; CME 作为非交易所成员, 为 E-mini 标准普尔承担和增加流动性 246
表 8.4 半价单指令购买 200 股股票. 纳斯达克在全国最佳报价和要约市场中. 上面部分: 半价单指令购买订单的结果是, 作为隐含的订单, 在中间市场定价为 114.495 美元. 然后, 在第 1 阶段的订单被取消, 一个新的一级请求级别被确定, 它将新中间价重新定价为 114.50 美元. 下面部分: 半价单指令购买订单结果, 立即以 114.495 美元的价格执行 251
表 8.5 购买 200 股被动委托订单. 上面部分: 限制价格为 114.50 美元, 导致订单被添加到订单簿中, 其与卖价相差 1 个单位, 并被定价为 114.49 美元. 下面部分: 限定价格为 114.51 美元, 订单将以 114.50 美元的价格填充 252
表 8.6 在交易所 A、交易所 B 和交易所 C 的 XYZ 订单的快照; NBBO 的买价为 114.90 美元, 报价为 114.50 美元 254
表 8.7 ITCH5.0 添加订单消息规范. 此消息将为纳斯达克系统接受的未归类订单而生成. 添加订单消息的固定长度为 36 字节 270
表 8.8 ITCH5.0 订单执行消息规范. 只要书上的订单全部或部分执行, 就会发送此消息 270
表 8.9 ITCH5.0 添加订单消息. “A” 表示增加消息, “B” 表示买单, “S” 表示卖单; “股票定位” 采取整数值, 是识别证券的定位代码; “追踪号码” 是指纳斯达克 OMX 内部追踪号码; “时间戳” 是 0 点以来的纳秒数 271

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

第一章　概　　论

本章对于量化交易的历史背景进行回顾, 并对量化交易领域进行综述. 在 1.1 节中, 我们对于量化交易基础结构从口头交流、手势交流到电子交易平台的演变历史进行回顾. 1.2 节对于量化交易策略进行综述, 并分析了不同的策略的时间跨度. 1.3 节简要回顾了从有效市场假说到利用量化交易进行套利的模式转变历史. 1.4 节对于使用量化交易策略的量化基金进行了介绍, 并介绍了与此紧密相关的公募基金和对冲基金. 1.5 节对于本书副标题提到的量化交易中的 “算法、分析、数据、模型和优化” 五个主题分别进行了介绍. 1.6 节讨论了本书的跨学科性背景和多样化的目标读者人群, 并对于不同类型的读者如何使用该书给出了建议. 1.7 节给出了补充材料和习题.

1.1　交易基础结构的演变

从经济学角度来说, 证券交易本质上是一种双向拍卖. 在证券交易中, 买方和卖方可以对于证券的价格进行竞价, 提供最优价格的买方和卖方成交, 这就是拍卖中的 “有效价格发现” 的过程. 在电子交易平台出现之前, 买方和卖方只能到证券交易所进行场内交易. 在场内交易时, 一种常见的沟通方式是公开喊价, 即买方和卖方口头喊出或用手势来表示自己证券买卖的数量. 如图 1.1 所示, 不同的手势代表不同的行为, 手心向外、手向外伸表示卖出证券; 而手心向内、手向内合表示买入证券. 与传统的拍卖系统不同, 公开喊价系统允许多个交易者同时对于一个报价进行响应, 并且他们可能各自完成一部分交易, 并不像典型拍卖制度那样具有清晰的价格–时间优先顺序.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)



图 1.1　芝加哥证券交易所的公开喊价系统中的手势说明

在 20 世纪 80 年代, 信息技术的进步对于电子交易平台发展起到了重要的推进作用. 与公开喊价系统相比, 电子交易平台使得买卖交易指令更易于完成, 也更易于进行监测、清算和结算.1971 年由国家证券业者自动报价系统协会成立的纳斯达克交易所 (NASDAQ) 是第一个电子证券交易市场. 在成立初期, 纳斯达克只是一个具有自动交易功能的电脑公告板, 并不具备联系证券买卖双方的功能. 在 1987 年前, 纳斯达克的大多数交易还都是通过电话来进行的. 成立于 1969 年的 Instinet 公司是第一家电子通信网络公司, 服务对象是机构投资者.Instinet 公司的交易平台在 80 年代中期发展迅速. 在 90 年代末期美国证券交易委员会 (SEC) 引入委托处理规则 (OHR) 和另类交易系统 (ATS) 时, Instinet 公司已成为市场中占主导地位的电子通信网络公司. 在 1986 年, 伦敦证券交易所也引入了电子交易系统. 芝加哥商品交易所 (CME) 在 1987 年开始构思一个对于国债、外汇、商品等多个金融市场同时开放的电子交易平台 Globex, 并于 1992 年将该交易平台正式发布. 作为对芝加哥商品交易所开设的平台的回应, 芝加哥期货交易所 (CBOT) 成立了一个基于电子交易平台的 “电子公开喊价系统”, 该系统允许电子交易与传统的交易池交易同时进行. 1997 年, 多伦多商品交易局引入了电子交易系统. 在原油期货交易市场的份额

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

落后于洲际交易所 (ICE) 后, 纽约商品交易所 (NYMEX) 于 2006 年也引入了电子交易系统. 同年, 纽约证券交易所 (NYSE) 也引入了电子交易系统.

在电子交易平台被广泛应用之前, 由于交易所会进行清算和结算, 交易市场中最精确的数据是交易价格, 而不是买卖价格. 实时的交易数据通过证券报价机向公众公开, 报价机会记录交易公司的名称缩写、交易价格和交易额等信息. 根据 Powers (1996) 对于 "旧世界的交易池" 的描述, 在当时的交易中, 交易者喊出交易的报价以及是否提受现有的报价. 交易者会将交易的细节信息记录在特制的卡片上, 并随后输入到交易所的清算系统中. 如果买卖双方输到清算系统中的数据不匹配的话, 该交易会被当成错误交易. 在交易双方清除分歧之前, 该交易不能完成. 错误交易对于交易来说, 是一种额外的成本. 如果清算机构出现错误交易的比例超过一定水平, 一些交易所会对于清算机构进行处罚.

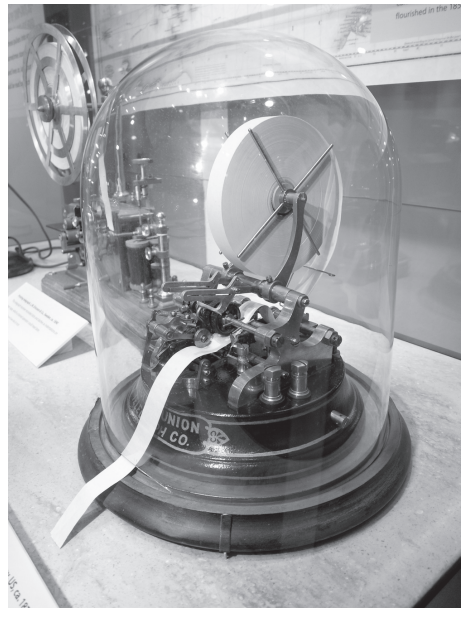


图 1.2 西联电报公司在 19 世纪 70 年代制造的证券报价机. 最初, 交易单上只有交易价格和交易公司的缩写, 到 20 世纪 30 年代以后, 交易额也会出现在交易单上. 该交易机目前在美国加州山景城的计算机历史博物馆展出. 照片来源: 维基共享资源/Don Debold

电子交易平台 (E-Trading) 是一种典型的在商业现货 (COTS) 操作系统上运行的专利软件, 平台通常使用一些常见的网络通信协议, 或者使用类似的 TCP/IP 等协议. 交易所通常会开发自己的电子交易系统, 并提供两种方式来登录系统: (a) 使用交易所提供的桌面图形化用户界面 (GUI) 系统来进行登

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

录. (b) 使用一个应用程序界面 (API) 来进行登录. 大多数交易所通过建立在公司网络上的网关来与交易所的交易系统连接. 买卖双方的投资公司的交易系统比交易所的交易系统更复杂, 一方面是因为投资公司需要同时与多个交易所进行交易, 对于交易系统的要求更高; 另一方面, 投资公司减少了股票、债券、外汇、金融衍生品交易员, 而是更多地基于算法来分析金融市场, 并自动执行相关的交易, 因此在电子交易平台科技上的投资也在逐渐增加.

电子交易的一个很大的优势在于可以减少交易成本. 尽管在经典的投资理论和期权定价理论中, 交易成本常常被认为是可以忽略的, 但是在交易策略的实际制定中, 交易成本是一个重要的考虑因素. (参考 Lai 和 Xing (2008, 第 183 和 300 页)). 当考虑交易成本时, 投资理论与期权定价理论的复杂程度大大增加, 本书第三章对于该领域的最新进展进行了介绍. 利用电子交易平台的直通式处理系统 (STP), 可以降低交易成本, 从而在增加交易量时不会显著增大交易成本, 以增大交易的回报. 与传统交易平台相比, 电子交易平台的优势还表现在随着金融市场中交易者数量的增加, 金融市场更具有流动性. 其次, 电子交易平台中证券价格信息以电子化的形式在全球范围内传播, 可以增加市场透明性. 在利用电子交易平台进行交易时, 交易者通过一次点击就可以完成交易, 并不需要通过经纪人或者交易池中的交易员, 避免了投资中的一些主要障碍, 且其影响是全球性的. 另外, 电子交易平台的另一个重要的影响是此时做市商的买卖价差更小, 即此时做市商的最优卖出价和最优买入价的差价更小, 因而做市商的收益也更小.

电子交易平台的发展不仅仅体现在一些大型交易所中, 也体现在 B2B 和 B2C 交易中. 在 B2C 交易中, 证券经纪人充当投资者与 B2B 市场的中介, 散户和机构投资者与经纪人进行交易. 机构投资者包括交易大量证券的对冲基金、基金经理和保险公司. 散户则通常是指交易量较小的一些个人. 在 20 世纪 90 年代末期, 大量的散户 (如嘉信理财、亿创理财等) 开始通过电子交易平台进行交易. 目前, 大部分美国的散户都已经从电话交易转到使用电子交易平台进行交易. 大型的机构投资者通常使用专有的电子交易平台进行交易, 如 Bloomberg Terminal, Reuters 300 Xtra, Thomson TradeWeb, 或者自己开发的软件. 在 B2B 交易中, 大型投行和经纪人直接与对方进行交易, 而通过电子交易平台的联系也符合金融信息交换 (FIX) 协议的标准.FIX 协议是行业内的交易前数据通信和交易执行的标准, 该标准最初用于证券交易, 后来逐渐应用于商品、货币、衍生品、固定收益证券等交易.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

1.2 量化交易策略及时间跨度

量化交易策略是基于对金融市场的量化分析和对未来的预测的投资策略. 对于未来的预测及策略的制定都与投资的时间跨度有关, 具体可分为以下几类:

- 基本面驱动的交易策略 (FMQ): 该策略是指当证券的价格相对于其基本价值被低估时, 进行买入; 而当证券的价值被高估时, 则进行卖出. 由于证券的基本面价值是由按季度发布的收益报告和预测报告所决定的, 因此该策略的时间跨度为一个季度. 在进行基本价值分析时, 常用的一些指标是收益质量、公司价值以及投资者情绪.

- 全球宏观策略: 宏观策略利用对市场事件和趋势的宏观分析来识别投资机会. 宏观策略可以分为酌情策略和系统性策略, 前者是指由投资经理发现和选择投资机会, 而后者是指利用模型和软件替代人工来进行交易. 举例来说, 当宏观分析表明美元即将上涨时, 则买入美元. 美国国会规定美联储必须持有相应的储备, 且储备的主要形式是国债. 由于国债的周期为 28 天 (约一个月), 因此该全球宏观策略的时间跨度是一个月.

- 收敛策略、相对价值策略和其他统计套利 (StatArb) 策略: 收敛性策略是指同时投资于一组相似的资产, 且这些资产的价格存在收敛的趋势. 举例来说, 由于流动性的差异, 正在发行的 30 年期美国国债的价格通常会比 29.5 年期国债的价格更高. 但当证券发行一段时间后, 这两种证券的价格都会下降, 因此两种证券之间的流动性溢价会逐步减小. 因此如果在证券新发行时选择做空 30 年期的国债, 并买入 29.5 年的国债, 则在几个月后可以通过以相近的价格买入 30 年期国债并卖出 29.5 年期国债的方法平仓, 从而从中获利. 与收敛策略类似, 相对价值策略是指同时做空被高估的资产, 并做多被低估的资产, 然后在价差降低时反向操作从而获利. 该策略可以应用于股票、证券、衍生品等不同的资产. 相对价值策略旨在在不同的市场情况下收益都能够保持一致性. 收敛策略和相对价值策略都属于统计套利策略, 这种策略的时间跨度范围从一分钟到几个月不等. 在下一节中我们会对统计套利进行概述.

- 高频交易 (HFT): 在 1.5 节中会对高频交易进行具体阐述, 高频交易的时间跨度为毫秒量级, 且高频交易的资产持有时间通常小于一秒.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

1.3 统计套利及关于有效市场假说的争论

有效市场假说 (EMH)在 20 世纪 60 年代由 Eugene Fama 基于他在芝加哥大学的博士论文提出的. 该假说的理论基础是在 20 世纪初 Louise Bachelier 基于他在巴黎–索邦大学的博士论文《投机理论》而提出的布朗运动模型 (离散情况下的随机游走). Cowles (1933, 1944) 的实证研究表明专业的投资者常常很难得到比市场收益更高的收益, 从而支持了随机游走模型. 随着 Fama 发表其毕业论文, Samuelson (1965) 提出有效市场假说的另一种阐述, Fama (1970) 对于有效市场假说的理论和存在的证据进行了综述, 有效市场假说在 20 世纪 70 年代成为金融经济学中的重要理论. 在有效市场假说中, 除了假定市场中的个体以效用最大化为目标外, 还假定个体对于未来事件的发生有理性预期, 且在新的信息出现时会更新现有的理性预期. 一般来说, 个体的预期应该是准确的. 尽管有些个体对于市场的预期可能是错误的, 但他们对于市场的反应是随机的, 且并不能通过对信息的可靠利用而获得超常的收益. 在存在交易成本的情况下, 超常的收益更难获得. 目前对于有效市场假说的实证检验结果纷杂, 对于实证结果的解释也并没有统一的结论.

行为金融理论是有效市场假说的替代理论. 行为金融理论中认为由于人体的认识偏差, 如过度自信、反应过度、信息偏差及其他的人为偏差, 会使得金融市场并不是一个有效的市场. 行为金融理论可以参考 Shiller (2015). Shiller 在 2013 年与 Fama 分享了当年的诺贝尔经济学奖. 另一方面, 有效市场假说的拥护者认为现实市场与有效市场的偏差在于个体的认知偏差, 而不是在于市场本身 (参见 Malkiel (2013, 第 11 章)). 为了协调有效市场假说与行为金融理论的冲突, Lo (2004) 在有效市场假说和行为金融理论的基础上, 基于进化原则对于有效市场假说进行了改进, 提出了适应性市场假说. 而在此之前, Lo 和 Mackinlay (1999) 也曾对于有效市场假说中的随机游走模型进行了质疑. 在适应性市场中, 根据进化原则, 价格反映的是由市场的经济情况和退休基金经理、对冲基金经理、做市商、散户投资者等不同群体的市场参与者的天性共同决定的信息. 在一个群体中, 如果有很多成员共同在为了稀缺资源而竞争, 则市场可能会更有效率 (如美国中期国债市场). 而如果市场中只有少量参与者在为充足的资源而竞争的话, 市场更可能呈现没有效率的情况. 市场的有效程度会随着时间而改变, 且受到一些市场环境因素的影响. 这些环境因素包括市场中竞争者的数量、盈利机会的多少、以及市场参与者对于变化的环境的适应能力等, 这些因素共同刻画了 “市场生态”. 在适应性市场假说下, 行为金融学中认为与

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

有效市场假说中的经济理性假说相违背的例子与经济行为中的竞争、适应、自然选择等进化原则相符合 (参考 Lo (2014)).

在有效市场假说下, 市场价格反映了能获得的所有信息, 因此在此类市场中, 基于所获得的信息集进行交易并不能获利. 有效市场假说可以分为三类: (1) 弱有效市场: 市场价格只能反映证券的历史价格的收益的信息; (2) 半强式有效市场: 市场价格反映了所有公开的证券信息; (3) 强有效市场: 市场价格反映了所有公开及未公开的信息. 和有效市场假说相反, 统计套利 (StatArb) 理论认为, 市场中存在统计套利机会 (SAOs), 如金融市场中可能会存在对风险资产的错误定价. 与通过对不同的证券分别做多或做空, 以获得确定性收益的策略 (这样的策略在有效市场中并不存在) 相反, 统计套利策略通过利用统计套利机会来制定相应的策略, 从而获得正向的期望收益. 统计套利策略中可以允许存在获得负收益的可能性, 但期望收益为正. Bondarenko (2003) 指出如果排除市场中的统计套利机会, 则会对股票的动态价格产生一个鞅式的影响.

统计套利通常是指在投资套利机会出现时, 一些特定类型的对冲基金制定统计套利策略来获利. 在进行套利时, 这些对冲基金通常会使用大量的证券、高级的统计模型、数据分析方法和高性能的计算技术和信息技术来制定策略. 在构建投资组合时, 通常需要两个阶段: 在第一个阶段中会对于每一个证券根据其综合指标进行评分, 评分高的证券表示建议进行多头交易, 而评分低的证券则表示建议进行空头交易; 在第二个阶段, 这些证券会被组合成一个投资组合, 从而降低市场风险和因素风险. 在第一个阶段中的评分方法通常是基金公司内部专利持有的, 但在第二个阶段中的风险评估模型, 如 MSCI/Barra 模型和 Northfield 信息系统等商用模型, 通常是可以在市场中购买的.

1.4 量化基金、公募基金、对冲基金以及基金表现

量化基金是指利用模型和计算机替代人为决策来进行投资决策的投资基金. 量化基金的投资决策过程包含三个关键组成部分: (a) 输入模块, 用于提供决策需要的交易规则及数据; (b) 预测模块, 用于估计未来证券价格、证券收益率和评估风险的模型的参数; (c) 构建投资组合模块, 利用优化算法和启发式算法来生成建议的投资组合.

近年来, 量化投资决策成了新成立的公募基金非常常见的一种决策方法. 公募基金是指将众多投资者的余额集中起来, 由专业的投资经理来集中投资. 监管公募基金的制度严谨, 在美国成立公募基金需要在证监会 (SEC) 注册, 由董事会监督 (如果公募基金是以信托基金方式成立而不是公司或合伙人形式成

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

立的话, 则由信托受托人监督), 由注册的投资顾问运营, 并向公众出售以便投资者随时购买. 公募基金在家庭理财方面起着重要的作用, 尤其是在退休金管理方面. 公募基金通常根据其主要的投资方向来进行分类, 主要类型包括货币型基金、债券及固定收益型基金、证券和股票型基金、混合型基金等. 公募基金的投资者需支付基金的各项费用, 包括管理费、交易费及其他费用. 公募基金既包括开放型基金也包括封闭型基金, 其中开放型基金是指基金公司在每个工作日都不断地接受新投资, 同时发行新股. 与公募基金紧密相连的一种类型的基金是交易所交易基金 (ETF). ETF 基金在交易日中持有资产和进行交易的资产的价值与资产的净值密切相关. 与 ETF 基金签署协议的大型的证券公司可以直接与 ETF 基金进行 ETF 股票的买卖交易. 个人则可以通过散户经济人在二级市场上进行 ETF 股票的交易.

对冲基金是一种集合投资基金, 它通常是以私人合伙的形式募集资金而组建. 对冲基金通常是一种开放式基金, 允许投资者增减其投资额. 对冲基金的价值随着基金投资策略回报的变动而变动, 且价值反映在投资者可以从基金中提取的金额上. 对冲基金的投资经理通常会将自己的钱也投入到所管理的基金中, 因此基金的收益也与他们自身的收益直接相关. 对冲基金并不向大众开放购买, 也不向散户投资者开放. 对冲基金投资者需要向基金经理支付年费, 年费通常是基金的净值的一个百分比, 且当基金的价值在当年上升时, 需要向基金经理另外支付一笔绩效费用. 对冲基金的投资策略众多, 大致可以分为全球宏观策略、方向性策略、事件驱动策略、相对价值策略等. 交易策略既可以是 “酌情策略”, 即策略可以由基金经理制定; 也可以是 “定性策略”, 即通过计算机程序来确定. “定性策略” 也被称为 “系统性策略”. 全球宏观策略使用宏观趋势和宏观事件来识别全球市场中的投资机会. 方向性策略利用市场变动或市场的不一致性, 从科技、医疗、制药、能源、基础材料、新兴市场等一系列的市场中选择投资机会. 事件驱动型策略从公司的交易事件中获取投资机会, 这些交易事件包括合并、兼并收购、资产重组、破产、清算、诉讼等. 相对价值策略利用不同证券的相对价格差异来进行投资获利.

对于基金经理或投资组合经理的业界的评估通常是用业绩指标来衡量. 假设市场中存在一个无风险资产, 其利率为 r_f. 存在一个风险资产, 其收益率为 r, 且 r 的均值为 μ, 方差为 σ^2. 根据资本资产定价模型 (CAPM) 可知风险资产的 Sharpe 比率 (SR) 为 $SR = (\mu - r_f)/\sigma$ (参考 Lai 和 Xing (2008, 3.3 节)). 另一个衡量业绩的指标是风险资产的 Treynor 比率 $(\mu - r_f)/\beta$, 其中 $\beta = \mathrm{Cov}(r, r_M)/\sigma_M^2$, r_M 和 σ_M^2 分别是市场组合 M 的收益率和方差. 与 Sharpe

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

比率中使用的无风险利率 r_f 不同, 一个更常用的评估基金表现的指数是信息比率 (IR). 在计算信息比率时, 将 Sharpe 比率中的 r_f 替代为基准投资组合的收益率 r_B, 且将 σ 替代为 $r-r_B$ 的标准差. 特别的, $IR=(\mu-\mu_B)/\sigma_{r-r_B}$. 有些对冲基金也利用信息比率来计算绩效费用. Jensen 指数是另一种衡量业绩的方法, 在计算 Jensen 指数时, 将 CAPM 模型推广到线性回归模型 $r-r_f=\alpha+\beta(\mu_M-r_f)+\varepsilon$, 其中 ε 是噪声项, 而 α 就是 Jensen 指数 (在 CAPM 模型中 $\alpha=0$). 另一个对 Sharpe 比率的修正是 Sortino 比率 $(\mu-\tau)/\sigma_{\tau-}$, 其中 τ 是目标收益率或最低要求的收益率, $\sigma_{\tau-}=E\{(r-\tau)^2|r\leqslant\tau\}$ 是下半方差. 另一个对于 Sharpe 比率的修正是利用风险调整收益 (RAP) 来对于投资组合进行排序, 该方法由 1985 年诺贝尔奖得主 Franco Modigliani 和他的孙女 Leah Modigliani 一起提出. 由于两个提出者的姓氏都以 M 开头, 因此 RAP 也被称为 M2 测度. RAP 的定义为 $r_f+(\sigma_M/\sigma)(\mu-r_f)$ (参考 Modigliani 和 Modigliani (1997)).

上面提到的业绩评估指标的计算只涉及投资组合、市场组合或基准投资组合的收益率的一阶矩和二阶矩. 如果这些投资组合的收益率服从正态分布, 则前两阶矩就是评估指标的充分统计量. 在收益率不服从正态分布时, 则需要一些其他的风险相关的收益率指标, 2.9 节对于这些指标进行了简要的回顾. 业绩评估指标的时间跨度取决于基金向公众公布的业绩报告的时间跨度或基金评级公司的评级频率. 晨星 (Morningstar) 评级公司评级的频率包括一个月、年初至本日、一年、三年及五年等 (参考 http://www.morningstar.com). 为便于比较不同基金的业绩, 通常将收益率和其标准差进行年化. 在晨星公司的基金评级系统中, 将基金根据三个周期内的收益进行评级. 晨星公司基金评级的依据是公司自己制定的风险调整后的收益评分, 这个评分也包括了基金的销售费用和赎回费用. 根据评分将基金公司分为五类, 评分最低的 10% 为一星, 接下来的 22.5% 为二星, 接下来的 35% 为三星, 接下来的 22.5% 为四星, 评分最高的 10% 为五星.

1.5 数据、分析、模型、优化和算法

在本书副标题列出的这五个主题中, 无论是从套利机会角度, 还是从有效市场的角度来看, “数据” 对于量化交易来说都是最重要的. “分析” 则远远不止对数据的分析. 量化交易中分析的方法很多, 选择哪种分析方法取决于需要如何将分析的结果用于最优决策. 在量化交易中决策是指决定交易策略, 因此 “分析” 既包括对于数据的分析, 也包括对于数据驱动的交易策略的开发. “模型” 是

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

统计决策理论的基本组成成分之一, 在数据和决策之间起到桥梁作用. 除了用于数据和决策之外, 数据、分析和模型在针对不同交易应用程序的各种定量策略中都具有突出的作用, 具体可参见本书第二至七章. 二至七章对于数据分析的 “优化” 也进行了研究. 优化也是量化交易中的重要问题, 它与数据驱动的交易策略有着紧密的联系.

本书的封面对于五个主题的排序与本节的排序相同, 这个顺序也是量化交易中的流程顺序. 在进行量化交易时, 用于分析的机器首先需要导入数据, 然后利用模型和优化程序来开发量化交易策略. “算法” 是指寻找优化方程或其他数学或数据分析问题的解的步骤, 而算法交易 (Algo Trading, 也称为自动交易) 是指利用算法在电子化交易平台上进行交易. 算法交易利用算法来执行包含交易时间、交易价格、交易数量等信息的交易策略. 在算法流程中, 算法的执行结果也是本书副标题列出的量化交易流程的最终输出. 量化交易被广泛应用于公募基金、对冲基金和其他买方机构投资者. 一些卖方投资者, 如做市商、投行和一些为市场提供清算服务的对冲基金, 也倾向于使用电子交易平台来生成和执行交易. 高频交易是算法交易中的一个特例, 本书的第六至七章对于高频交易的算法和策略进行了阐述.

在本书 1.1 节中提到的电子交易平台为算法交易的发展铺平了道路. 根据 Aite Group (http://www.aitegroup.com) 的报告, 在 2006 年, 有三分之一的欧洲和美国的证券交易商开始使用算法交易进行投资.《纽约时报》关于高频交易的报道显示, 截至 2012 年 12 月 20 日, 在美国利用算法交易成交的交易量占总交易量的比例提高到了约 50%. 算法交易也逐渐渗透到外汇交易和期权市场的交易中. 算法交易的发展使得市场上人工交易员的工作逐渐被计算机所取代, 因此用于交易的计算机的运行速度对交易非常重要, 目前计算机的速度大概已经到达毫秒甚至微秒级. 在高频交易中, 买方的交易员需要使用 “交易管理系统” 来面对日新月异的算法交易类型的挑战. 卖方的交易员则需要使用与买方的交易系统相兼容的交易系统 (细节可参考 Shetty 和 Jayaswal (2006)). 第八章对于为高频交易提供信息和计算支持的交易系统进行了详细说明.

1.6 量化交易的跨学科性及本书的使用方法

如前所述, 量化交易具有多面性, 是一个跨学术界和业界的学科, 本书也致力于展现量化交易的这些特点. 本书的作者既包括来自学术界的教授, 也包括业界专家, 他们结合在不同领域的背景和经历, 将量化交易的核心内容呈现在目标读者面前. 本书的目标读者既包括在前言中提到的学习相关课程或培训项

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

目的学生, 也包括交易员、量化策略分析师和监管者. 考虑到目标读者的不同背景和不同的兴趣所在, 本书在结构安排上, 也努力使不同类型的读者能够集中精力在各自感兴趣的章节.

本书首先考虑学生这个目标群体. 如前言所述, 本书的作者们在不同的大学面向硕士生和高年级的本科生开设了相关的课程和培训项目. 其中, 第五至八章的内容是本书作者开设的一门高频交易/算法交易的课程的教材内容, 第七至八章涵盖了课程的重点内容. 在课程中, 学生可以根据他们自身的背景在“补充材料和习题” 章节中选择相关的课题进行深入研究和完成课程报告. 举例来说, 计算机专业和电子专业的学生通常可以从第七章和第八章中选择相关的课题. 我们同时也允许学生组成跨学科的小组来完成课程, 在跨学科的小组中, 学生可以互相学习, 共同分担课程任务. 本书的第一至四章也被用于开设资产管理和金融计量学的课程. 第二、四、五章被用于开设一个以数据分析和量化交易为主题的暑期项目, 而本书作者也正在将该暑期项目扩展为一个学期的课程. 作为这些课程的教材, 本书试图在方法论的发展和实际应用两个方面做平衡, 以实现从不同的角度将主要观点和贡献进行无缝衔接, 从而帮助学生理解前沿的方法、了解业界的工作、学习如何与数据打交道.

接下来, 本书考虑初级交易员和量化分析师这个群体. 他们往往已经获得了前言中提到的量化交易相关的学位, 且希望能够进一步学习量化交易中的尖端方法和现代交易实践. 对于这个群体而言, 本书会是他们的重要参考书, 帮助他们对建模方法的理论基础获得坚实的理解, 启发他们在数据分析、最优策略、限价委托建模等方面的思考, 帮助他们探究交易所的信息协议和匹配引擎相关的信息学知识. 对于散户量化投资者而言, 尽管他们可能精通某个特定的领域, 但对整个交易过程的理解并不完整. 本书涉猎广泛, 提供的方法可以帮助他们巩固对于交易过程的理解. 对于交易执行者而言, 本书可以帮助他们对于交易背后的模型以及如何调整模型的参数有更加深入的理解. 此外, 本书也可以帮助他们对于最新的统计模型的优缺点有一个基本的把握, 从而帮助交易者识别市场基本结构和模型中的市场行为. 另外, 本书对于很多对量化基金交易者非常重要的监管问题也进行了讨论, 如交易前和交易后的风险管理、熔断机制等. 由于新技术的投入应用和监管的实行可能会在不久的将来有重大的改变, 我们建议读者把重心放在本书的关键问题上, 而不是关注细节. 因此, 本书对于交易所的匹配引擎的机制等很多专题都只是做概念性的简略介绍.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

1.7 补充材料和习题

1. **证券交易基础结构的经济理论：双向拍卖市场** 与包含一个卖方和多个买方的经典拍卖市场不同, 在一个双向拍卖市场中, 同时存在多个卖方和多个买方, 卖方同时提供多个物品进行拍卖. 双向拍卖的成交价是买方愿意提供的最高价和卖方能接受的最低价. 在双向拍卖市场中, 最基本的问题是寻找市场的竞争均衡. 在达到均衡时, 存在一个价格体系使得在此价格下供给的商品数量与需求的商品数量相同, 该价格称为均衡价格. 竞争均衡也称为瓦尔拉斯 (Walrasian) 均衡, 以纪念 Leon Walras, 瓦尔拉斯用拍卖的方式阐述了竞争均衡理论. 假设市场中存在一个卖方和多个买方, 买方首先计算出他们在每一个价格下的需求量, 然后将出价提交给拍卖商. 在完全竞争市场中, 假设交易费用为 0, 则在完全信息下拍卖商会将供给和需求进行匹配, 信息完整且没有交易成本. 瓦尔拉斯均衡提出了一种通过迭代来求解均衡的方法, 这种方法称为 “试错法”. 在拍卖中, 卖方和买方的报价反映了他们对商品的供给和需求情况. 在迭代求解时, 当价格高于均衡价格时, 则会导致过量供给, 价格随之下降; 反之当价格低于均衡价格时, 会导致过量需求, 价格随之上升. 随着多次迭代, 最终可得到均衡价格. 很多后续研究基于瓦尔拉斯的基本思想进行了改进或机制设计, 并在不同的情况下进行应用. 后续研究包括 Vickrey (1961), Demange 等 (1986), Roth 和 Sotomayor (1990), Richter 和 Wong (1999), Krishna (2009), Milgrom (2004), Varian (2007), Edelman 等 (2007) 等. 在 20 世纪 80 年代以后, 经济学学者将瓦尔拉斯均衡理论扩展到了双向拍卖市场. 由于问题的复杂性, 很自然的做法是首先考虑市场中只有一个卖方和一个买方的情况, 此时双向拍卖市场简化成了双边交易情况. 在这种情况下, 如果卖方定价为 S (保证定价足以支付生产成本), 而买方愿意支付的价格为 B, 如果 $B \geqslant S$, 则均衡价格可以是任意 $p \in [S, B]$. 在实际拍卖中, 买方和卖方只知道其自身对于物品的估价, 而假定对方出价的先验分布是均匀分布. Myerson 和 Satterthwaite (1983) 证明在贝叶斯博弈中, 在纳什均衡时买方和卖方为了最大化各自的期望收益, 会选择对于物品的实际估价作为最终报价. 在更一般性的双向拍卖市场中, 同时存在多个卖方和多个买方, 此时一个复杂之处是市场的参与者获得的信息是异质的. Glosten 和 Milgrom (1985) 以及 Kyle (1985) 的开创性工作对于由于信息不对称导致的逆向选择对一个存在做市商的双向拍卖市场

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

的影响进行了研究. Back 和 Baruch (2013) 基于市场中存在充分的逆向选择这个假设, 建立了有限委托市场中的纳什均衡, 并对均衡的性质进行了优化. 在双向拍卖市场中, 另一类的研究是在市场机制设计上. McAfee (1992) 对于买方和卖方同时存在占优策略[①]的机制进行了研究. Hollifield 等 (2006) 基于订单提交、执行、取消的历史数据, 开发了一种可以预测在当前市场、完全竞争和卖方垄断这三个市场中可获得的收益的方法, 并将该方法用于现有数据, 从实证的角度证明当前的限价委托市场具有一个完善的机制设计.

2. **机制设计** 在附录 A 中引用的 Vickrey (1961), Myerson 和 Satterthwaite (1983) 以及 McAfee (1992) 这些研究属于经济学研究中一个重要的子课题——机制设计. 机制设计用工程的方法对于市场的经济机制和激励机制进行设计. 机制设计问题通常建立在博弈论的框架下, 假设每个代理人都是理性的, 将代理人的类型映射到收益空间上的机制函数 $y(\cdot)$ 则由委托人来确定. 代理人的类型 θ 反映了代理人关于交易标的的质量的真实信息 (或偏好), 而代理人向委托人汇报时, 可能并不诚实地汇报真实信息, 假设汇报的类型为 $\hat{\theta}$, 且在机制被执行后获取的收益为 $y(\hat{\theta})$. 委托人的目标是优化社会选择方程 f 中的收益函数, 定义在完全信息 θ 下的收益为基准收益, 但由于委托–代理问题的存在, 委托人只能利用 $\hat{\theta}$ 来进行优化. 考虑到 "激励相容" 约束, 委托人应该提供合适的激励来使得代理人能够汇报真实的类型 θ. 考虑到代理人可能并不诚实, 委托人的另一种选择是修正其优化方程. 在附录 A 中提到的 Myerson 和 Satterthwaite (1983) 这篇论文假设市场中存在一个经纪人 (委托人), 市场中的买方和卖方 (代理人) 都会向经纪人汇报其报价, 在此基础上设计了一种同时满足贝叶斯激励相容[②]约束和个体理性约束的机制. 在 2007 年, 诺贝尔经济学奖被授予 Leonid Hurwicz, Eric Maskin 和 Roger Myerson, 以表彰他们在机制设计理论方面奠定的基础. 前面提到的 Vickrey (1961) 这篇论文也是机制设计方面的奠基论文之一, 这篇论文帮助 William Vickrey 在 1996 年获得了诺贝尔经济学奖.

3. **逆向选择** 逆向选择是由市场中买方与卖方之间的信息不对称产生的. 逆向选择的经典例子是 Akerlof 提出的二手车市场的例子. 在二手车市

① 在博弈论中, 如果存在一个与其他竞争对手可能采取的策略无关的最优策略, 该策略就被称为占优策略.

② 如果代理人汇报真实的类型, 则会达到纳什均衡的情况称为贝叶斯激励相容.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

场中, 卖方更清楚车的质量 (究竟是有缺陷的二手车, 还是一直细心保养的二手车, 论文中用 “柠檬” 来比喻状况不佳的二手车), 但买方在购买之前并不知道车的状况的具体信息, 因此只愿意支付二手车市场的平均价格.①

习题 阅读 Glosten 和 Milgrom (1985), 解释为什么当交易平台中存在拥有内部信息的交易者 (内部交易者) 时, 逆向选择会导致正向买卖价差. 内部交易者在公司发布收益报告前便能够了解公司的价值, 而普通交易者根据获取到的公众信息, 使用相应的统计分布来描述他们对证券情况的一致看法; 做市商在交易市场中从维持市场流动性上获得的期望收益为 0, 他们负责在交易发生时随时更新证券的信息.

4. **关于 McAfee 机制的习题** 阅读 McAfee (1992) 这篇论文, 并解释论文摘要中提到的 “McAfee 机制满足以 $1/n$ 的速度收敛至双向拍卖中的买方效率” 和 “McAfee 机制总是能够得到完全信息下的社会最优价格, 而市场缺乏效率是因为该机制阻止了产生最低价值的交易的发生.” 在该论文的第二部分, 对于 McAfee 机制构造进行了说明:

   - 根据买方的报价 $b_1 \geqslant b_2 \geqslant \cdots$ 和卖方的报价 $s_1 \leqslant s_2 \leqslant \cdots$ 进行下单.
   - 找到一个满足 $b_k \leqslant s_k$ 且 $b_{k+1} > s_{k+1}$ 的指标 k.
   - 定义 $p = (b_{k+1} + s_{k+1})/2$.
   - 如果 $s_k \leqslant p \leqslant b_k$, 则定义如下机制: 让报价不低于 b_k 的 k 个买方与报价不高于 s_k 的 k 个卖方以价格 p 进行交易.
   - 如果 $p \notin [s_k, b_k]$, 则定义如下机制: 让报价不低于 b_{k+1} 的 $k-1$ 个买方以价格 b_k 买入, 而让报价不高于 s_{k-1} 的 $k-1$ 个卖方以价格 s_k 向做市商卖出.

   在该机制中, 需要存在一个做市商, 且在 $p \notin [s_k, b_k]$ 的情况下, 保留超额收益 $(k-1)(b_k - s_k)$.

5. 机制设计和双向拍卖理论不仅属于博弈论和微观经济学的研究范畴 (参考 Fudenberg 等 (2007)), 也引起了计算机领域的兴趣. 举例来说, Nisan

① 另一个经典的例子是 Joseph Stiglitz 开创性地提出的医疗保险领域的逆向选择现象. 在 2001 年, Stiglitz、Akerlof 和 Michael Spence 由于他们对于存在不对称信息市场的研究而共同获得了诺贝尔经济学奖.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

和 Ronen (2007) 指出基本上现有的复合拍卖中使用的所有启发式近似算法都符合基于 VCG 的机制[①]. 论文也对基于 VCG 的机制进行了改进来避免出现无法计算得到最优解的情况. 另一个例子是 Dütting 等 (2014), 该论文认为双向拍卖市场由市场双方使用的排序算法和一个复合的规则组成. 论文同时将该框架应用到各种不同的可行性受限的环境中, 并将贪心算法中的近似方法应用到求解双向拍卖市场中的社会福利.

6. **资本市场的有效性和鞅性** 这个标题也是 Le Roy (1989) 论文的题目. 该论文阐述了鞅的重要性, 指出对于包含目标优化的参与者的竞争均衡而言, 经典的随机游走模型中的 "独立增量" 假设 "限制性过强因而难以成立", 而在 Samuelson (1965)[②]中首次提到的弱鞅模型就已经可以刻画随机游走模型在市场中的作用. Lehmann (1990a) 发现很多近期的研究已经证明资本回报率可以以一定的可靠性被预测, 这与有效市场的情况相符合. 而用鞅过程来预测回报率时, 符合回报率服从均值回归过程的特点. 接下来, 论文使用鞅过程对于有效市场假说进行了统计检验, 并通过实证研究说明证券市场 (特别是纽约证券交易所) 的流动性缺失主要出现在市场价格出现大波动性时, 而这些波动是由于投资者的过度反应和认知偏差所导致的.

① VCG 指在 1961 年、1971 年和 1973 年分别提出拍卖市场中的机制设计的 Vickrey, Clarke 和 Groves 三人, 在基于 VCG 的机制的设定中包含无法计算得到的最优结果, 在这种情况下则只能通过次优近似来进行机制设计.

② Le Roy 认为 Samuelson 的论文因为在引起从随机游走模型到鞅模型的转变中的作用而成为最为重要的论文.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

第二章　量化交易中的统计模型与方法

本章首先对于量化交易中的数据和不同类型的量化交易的时间跨度进行介绍. 2.1 节对于证券价格数据的特征进行描述, 并重点对于低频 (以天/星期/日为单位) 的交易和高频 (逐笔) 交易的区别进行分析. 在 1.3 节和 1.7 节中, 本书利用低频交易中的数据对于有效市场假说进行了实证验证. 尽管高频数据的一些特征 (如在 2.1.2 节中提到的反向一阶滞后相关关系) 与有效市场假说不符合, 但是在利用时间序列来分析交易数据时, 考虑市场约束 (即市场微观结构) 的影响, 该特征并不成立. 高频交易数据的模型会在第四章和第五章中单独进行分析, 在第 2.2 和 2.6 节中, 我们对于低频数据分析中用到的统计模型进行回顾, 这些模型符合有效市场假说. 低频交易分析中的模型设定及对应的收益率被广泛用于在有效市场假说下建立投资策略. 2.3 节对于现代投资组合理论进行介绍. Malkiel (2003) 将现代投资组合理论称为 “在华尔街的随机漫步靴” (随机漫步 (游走) 是符合有效市场假说的最简单模型之一). 这些统计模型中都包含未知参数, 一系列的文献对于使用模型的方法进行了研究. 2.4 节对于这些文献进行了总结, 讨论了数据分析中的统计方法. 2.5 节介绍了一种新的模型分析方法. 该模型在投资组合的风险模型中加入了一种反映模型不确定性的未知参数, 并阐述了 “使用统计模型进行数据分析” 和 “根据当前需求开发新模型的应用驱动型数据分析” 之间的区别. 事实上, 用于描述对数化的证券价格的随机游走模型和低频交易收益率的特征并不完全相符, 2.7 节对于讨论

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

这些不符情况的统计模型进行了回顾. 另外, 该节还将 2.5 节中的模型扩展为时间序列模型, 这些时间序列模型符合有效市场假说, 且与低频交易的特征事实相符合. 2.8 节对于数据驱动型的量化交易策略及相关的统计模型进行回顾. 与有效市场假说不同, 这些策略假设市场中存在统计套利的机会. 另外, 本书同时对基于经济学和行为金融学理论的交易策略进行了总结. 2.9 节给出了相关的补充材料和习题.

2.1 股票价格的特征

定义时刻 t 的股票 (或其他资产) 的价格为 P_t, 假设该资产在时刻 $t-1$ 到 t 之间不可分, 则该资产的当期收益率为 $R_t = (P_t - P_{t-1})/P_{t-1}$, 连续收益率或对数收益率为 $r_t = \log(P_t/P_{t-1}) = p_t - p_{t-1}$, 其中 $p_t = \log P_t$. 因此, 该资产的 k 期对数收益率 $r_t(k) = \log(P_t/P_{t-k})$ 是 k 个时期的对数收益率之和 $r_t(k) = r_t + \cdots + r_{t-k+1}$. 具体可参见 Lai 和 Xing (2008, 第 64–65 页).

2.1.1 低频交易数据的时间序列分析

Lai 和 Xing (2008) 一书的 6.1 节, 基于每日和每周的股票价格对数收益率数据的时间序列实证研究, 对于对数收益率的特征进行了归纳总结. 这些特征主要包括 (a) 超额峰度导致的非正态特征; (b) r_t 序列的弱自相关性质; (c) r_t^2 序列的强自相关性质; (d) 股票价格的上下波动幅度的不对称性; (e) 由于外生变量 (如宏观经济变量) 和外部事件 (如定期的收益报告的发布等) 导致的波动率发生显著变化. 由于 r_t^2 序列可以作为潜在的波动率的测度 (参见 Lai 和 Xing (2008, 第 139 和 145–146 页)), 特征 (c) 常常与收益率的剧烈波动的聚集 (即 “波动率聚集”) 特点相联系. 特征 (d) 中提到的不对称性是指收益率的杠杆效应的不确定性, 即大的正向收益的波动率远远低于负向收益的波动率.

2.1.2 高频交易数据的离散化价格变动

交易数据通常是以离散值或逐笔来进行报价. 在 1997 年 6 月 24 日前, NYSE 的最小报价单位为 0.125 美元, 其后一直保持在 0.0625 美元, 直到 2001 年 1 月 19 日, NYSE 和 AMEX 的证券才开始以十进制作为报价单位来进行交易. 在 CME 交易的 S&P 500 期货交易的最小报价单位为 0.05 价位. 由于交易价格是离散化的, 可以观察到的价格变动也只能取值为特定的离散值 (通常包含 0). 这也导致价格变动出现价格聚集的情况, 即交易价格通常会聚集在特定的数据周围.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

逐笔交易价格数据在 19 世纪后期已经可以通过报价机获得, 且在 20 世纪初变得流行. 但尽管如此, 在 20 世纪 50 年代 Osborne (1959) 对于《华尔街日报》的数据进行研究之前, 大部分已发表的实证研究都只对于逐星期或逐月的数据进行分析. Niederhoffer 和 Osborne (1966) 基于 Francis Emory Fitch 公司提供的交易单上的数据的研究是该领域的开创性研究. 该研究使用的数据是 1964 年 10 月道 · 琼斯指数包含的 6 家企业在 22 个交易日的逐笔交易数据, 共计 10536 条. 除了上一节提到的价格的离散特征外, 该研究同时发现了交易强度的日间周期性和季节性. Anderson 和 Bollerslev (1997) 以及 Hasbrouck (1999) 等后续研究对于这些特征进行了更加深入的研究. 由于在每个交易日的开始和结束时, 交易量非常大, 而在中午时交易量很小, 季节性特征也被证券交易所称为 U 形特征. 因此, 交易之间的间隔时间呈现出日周期性特征. 欧元交易市场也有着类似的 U 形特征. 在每个交易日的上午 (此时欧洲市场仍然开放) 的交易明显比交易日的下午 (此时欧洲市场已经关闭) 更活跃. 另外, Niederhoffer 和 Osborne (1966) 的研究发现, 交易之间的价格波动存在负向的一阶滞后自回归特征.

例 2.1 在本例中, 我们利用香港交易所在 2014 年 10 月收集的逐笔交易数据来阐述前面提到的高频交易数据的特征. 为避免将香港证交所中交易的股票价格与证交所自身的股价相混淆, 此处我们使用的是代码为 388 的股票数据. 图 2.1 对于数据的时间序列特征进行了刻画. 图中的两条垂直线之间表

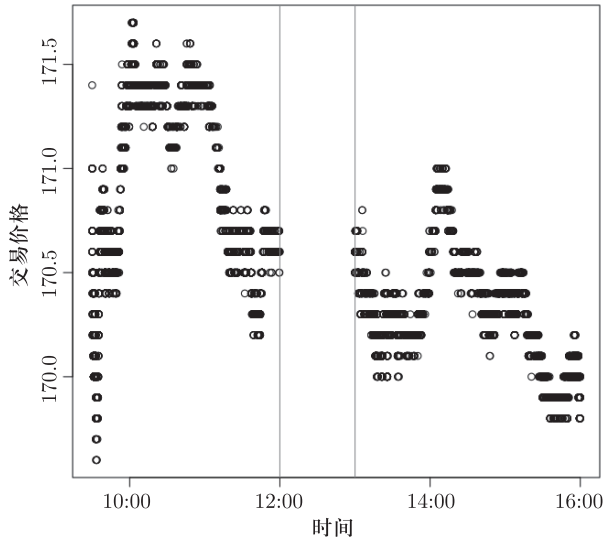


图 2.1 2014 年 10 月 10 日, 代码为 388 的股票的逐笔交易价格 P_t 的时序图

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

示香港时间 12 点到 13 点之间的午休时间. 此处在交易数据之间并没有用线连接, 以更好地显示出交易价格的离散特征. 在该天中共有 10455 笔交易数据, 数据的频率分布直方图在图 2.2 (上图) 中给出. 当天的最高价和最低价分别为 171.7 港元和 169.6 港元. 由于最小报价单位为 0.1, 因此当天的报价范围为 21 个最小报价单位. 交易价格的平均值和中位数分别为 170.4 港元和 170.3 港元. 交易价格的众数为 170.0 港元, 反映出投资者对于整数交易价格的偏好特征.

将第 i 笔交易的价格记为 P_i, 且将对应的价格变动记为 $D_i = P_i - P_{i-1}$. 图 2.2 (下图) 给出了价格变动 D_i (单位为最小报价单位) 的频率分布直方图.

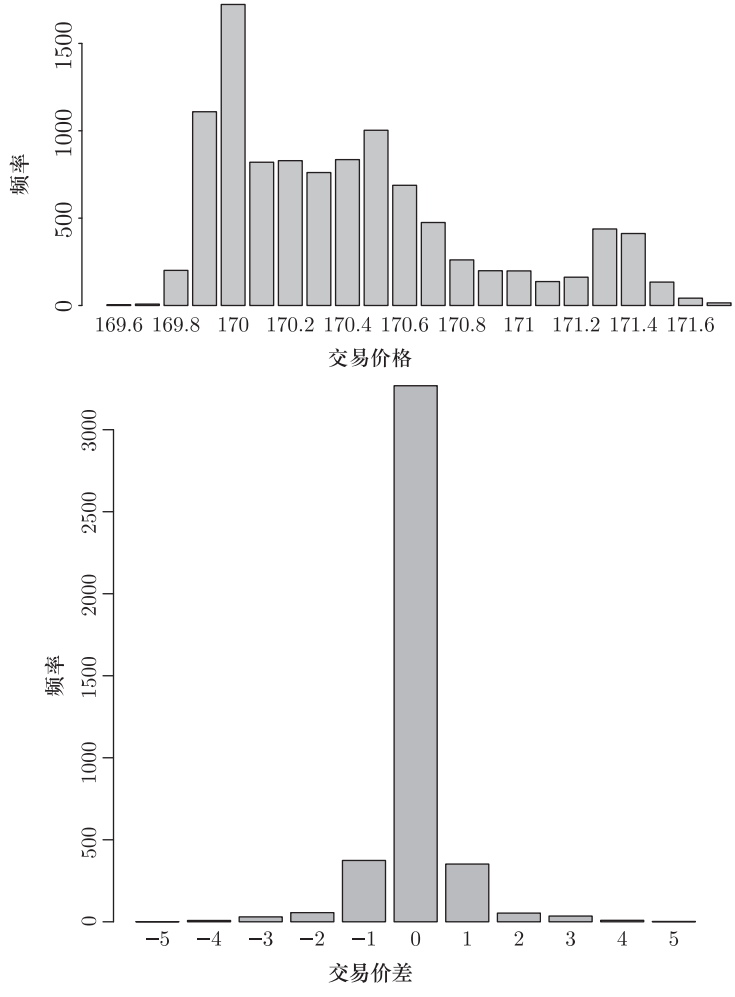


图 2.2 逐笔交易价格分布 (上图); 上午交易时段价差 (下图)

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

直方图显示出交易变动的高度集中在 0 附近 (当天上午的 4189 笔交易中, 有 3268 笔交易的变动为 0). 图 2.3 刻画了每个交易日上午的价格波动的自回归函数 (ACF). ACF 图显示出显著的负向一阶滞后自回归特征, 且其他阶的滞后项显著性为 0 或很低.

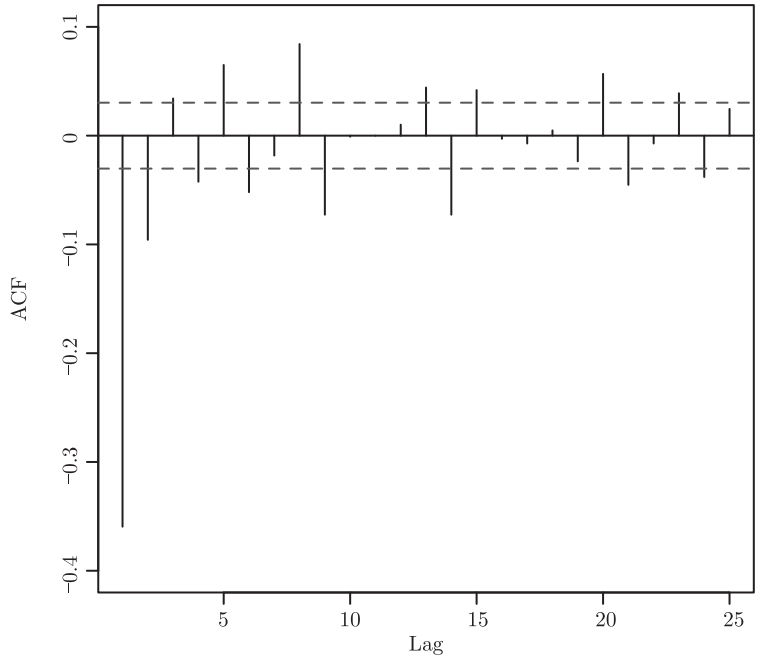


图 2.3 2014 年 10 月 10 日上午交易的代码为 388 的股票的交易价差的 ACF 函数. 虚线表示 5% 水平下的拒绝域临界值

2.2 巴黎证交所的布朗运动和华尔街的随机游走

根据 Jovanovic 和 Le Gall (2001) 以及 Le Gall (2007) 可知, 对于股票价格的研究起源于 Regnault (1863) 和 Bachelier (1900). 这两篇论文使用巴黎证券交易所的报价数据来研究价格变动的平均偏差. 研究表明, 这些偏差与时间间隔的平方根成正比, 即 $E[|P_{t+\Delta} - P_t|] \propto \sqrt{\Delta}$, 其中 P_t 指时刻 t 时的股票价格. Regnault (1863) 通过实证研究验证了平方根定律, 即 Bachelier 在 1990 年博士论文里利用启发式泛函中心极限定理推导出的布朗运动. 对于布朗运动的介绍, 以及适用于更加广义的鞅过程的严格泛函中心极限定理的介绍可以参考 Lai 和 Xing (2008) 的附录 A. 考虑到均衡市场的信息公平, Bachelier (1900,

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

1938) 认为价格变动 D_i 可以当作独立的对称的随机变量, 价格上升和下降 s 个最小报价单位的概率相同. 将连续时间的布朗运动 B_t 限制到均匀间隔的时间, 得到具有正态增量 $N(0,\sigma^2)$ 的随机游走. Samuelson (1973) 对于 Cowles (1933, 1944) 和 Kendall (1953) 研究支持的 Bachelier 的博士论文中提到的 "投机价格" 的随机游走模型进行了回顾, 指出 "18 只英国普通股序列并不存在显著的序列相关性, 因此序列看起来像随机游走." 然而, 注意到 Bachelier 将布朗运动模型应用到巴黎交易所的担保和看涨期权的理性定价中, Samuelson 评论指出 "即使是像 Bachlier 模型这样的开创性研究也会导致出现离奇的结果", 并提出使用几何布朗运动模型来对股票价格 P_t 进行建模, 此时低频数据的对数收益为独立同分布过程, 参见 Lai 和 Xing (2008) 一书 66 页和 328 页.

2.3 现代投资组合理论是在有效市场假说下的华尔街随机漫步靴

在有效市场假说的框架下, Malkiel (2003) 的第三章首先讨论了一个新的 "随机漫步靴", 即 Markowitz 现代投资组合理论下一项新的投资技术. 该技术告诉投资者如何在一定的收益下构建相应的投资组合使得风险最小, 并在资本资产定价模型 (CAPM) 框架下进行建模. 其中 CAPM 模型由 Sharpe, Linter 和 Black 提出, 对于 MPT 理论进行了改进.

Lai 和 Xing (2008) 一书的第 3 章对于 Markowitz 投资组合理论和资产资本定价模型 (CAPM) 进行了详细介绍. 假定有一个包含 m 个资产的投资组合, 每一个资产的期望收益为 μ_i, 投资组合中第 i 个资产所占的比例为 w_i, 且满足 $\sum_{i=1}^m w_i = 1$. 定义 $\boldsymbol{w} = (w_1,\ldots,w_m)^\top$, $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_m)^\top$, $\mathbf{1} = (1,\ldots,1)^\top$, 则投资组合的收益率的均值和方差分别为 $\boldsymbol{w}^\top\boldsymbol{\mu}$ 和 $\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}$, 其中 $\boldsymbol{\Sigma}$ 是资产收益率的方差–协方差矩阵. 在给定一个目标收益率 μ_* 的情况下, Markowitz 指出有效投资组合中各个资产的比例 $\boldsymbol{w}_{\text{eff}}$ 可以通过求解如下优化问题得到:

$$\boldsymbol{w}_{\text{eff}} = \arg\min_{\boldsymbol{w}} \boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w} \quad \text{满足 } \boldsymbol{w}^\top\boldsymbol{\mu} = \mu_*,\ \boldsymbol{w}^\top\mathbf{1} = 1,\ \boldsymbol{w} \geqslant \mathbf{0}. \tag{2.1}$$

在允许卖空的情况下, (2.1) 中的约束条件 $\boldsymbol{w} \geqslant \mathbf{0}$ (即对于任意 i, $w_i \geqslant 0$) 并不成立, 因此该优化问题可以获得如下显式解:

$$\begin{aligned}\boldsymbol{w}_{\text{eff}} &= \arg \min_{\substack{\boldsymbol{w}:\boldsymbol{w}^\top\boldsymbol{\mu}=\mu_*,\\ \boldsymbol{w}^\top\mathbf{1}=1}} \boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w} \\ &= \left\{B\boldsymbol{\Sigma}^{-1}\mathbf{1} - A\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \mu_*\left(C\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - A\boldsymbol{\Sigma}^{-1}\mathbf{1}\right)\right\}/D,\end{aligned}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 $A=\boldsymbol{\mu}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{1}=\mathbf{1}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, $B=\boldsymbol{\mu}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, $C=\mathbf{1}^{\top}\boldsymbol{\Sigma}^{-1}\mathbf{1}$, $D=BC-A^2$. 对于 $m\geqslant 3$ 的情况, 投资组合中各个资产的收益率的标准差 σ 和均值 μ 组成的点集 (σ,μ) 称为可行集. 该可行集是一个左凸的集合, 即该可行域中任意两点的连线并不会与可行集的左边界相交. 可行域左上边界被称为有效前沿.

将 Markowitz 的均值–方差投资组合优化模型拓展到包含 m 个风险资产和一个收益为 r_f 的无风险资产的市场, 则根据 CAPM, 有效前沿是一条与原有的可行域相切的直线, 切点对应唯一的风险资产组合 M, 即为 "单基金定理". "单基金定理" 认为, 有效前沿是由基金 M 和无风险资产的线性组合构成. 假设 M 的收益率的标准差和均值分别为 σ_M 和 μ_M, 则根据 CAPM, 资产 i 的收益率溢价 r_i-r_f 可以通过如下回归模型得到:

$$r_i-r_f=\beta\left(r_M-r_f\right)+\epsilon_i, \tag{2.2}$$

其中 $E(\epsilon_i)=0$, $\mathrm{Cov}(\epsilon_i,r_M)=0$, 且 $\beta_i=\mathrm{Cov}(r_i,r_M)/\sigma_M^2$. 根据 (2.2) 可知,

$$\sigma_i^2=\beta_i^2\sigma_M^2+\mathrm{Var}\left(\epsilon_i\right), \tag{2.3}$$

即资产 i 的收益率的风险可以分解成*系统性风险* $\beta_i^2\sigma_M^2$ 和*个体风险*两个部分. 系统性风险是指与市场相关的风险, 而个体风险是指该资产特有的风险, 个体风险可以通过构建投资组合而分散. 对应于无风险资产收益率 r, 投资组合的 β 值可以类似地定义为 $\beta=\mathrm{Cov}(r,r_M)/\sigma_M^2$. β 值可以看作资产的收益率相对于市场波动的敏感性. 现实中一个常用的市场组合是 S&P 500 指数, 该指数包含一系列加权的股票, 在美国证券市场被广泛应用.

2.4 现代投资组合理论的统计学基础

在 Markowitz 的有效前沿模型中, $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 这两个参数都是未知的, 一个自然的想法是使用历史收益率 $\boldsymbol{r}_t=(r_{1t},\ldots,r_{mt})^{\top}$, $1\leqslant t\leqslant n$ 的样本均值 $\widehat{\boldsymbol{\mu}}$ 和样本方差 $\widehat{\boldsymbol{\Sigma}}$ 来作为参数的估计值. 然而, 由于估计值与真实值实际上是有一定偏差的, 因此嵌入式有效前沿并不是最优的. Frankfurter 等 (1976) 以及 Jobson 和 Korkie (1980) 指出, 利用 $\widehat{\boldsymbol{\mu}}$ 和 $\widehat{\boldsymbol{\Sigma}}$ 得到的嵌入式投资组合的收益率甚至会低于等权重的低效率投资组合. Michaud (1989) 指出, 利用 $\widehat{\boldsymbol{\mu}}$ 和 $\widehat{\boldsymbol{\Sigma}}$ 得到的最小方差权重 $\boldsymbol{w}_{\text{eff}}$ 有着严重的缺陷, 并称使用该方法的投资者为 "错误估值投资者". Best 和 Grauer (1991)、Chopra 等 (1993)、Canner 等 (1997)、Simaan (1997) 以及 Britten-Jones (1999) 等在 Michaud (1989) 的基础上进行了后续深入研究. 在过去的 20 年中, 针对该问题的研究主要有三种方法: 第一

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

种方法使用多因子模型来降低估计 $\boldsymbol{\Sigma}$ 时模型的维度; 第二种方法使用贝叶斯方法或其他缩减技术来估计 $\boldsymbol{\Sigma}$. 这两种方法都对有效前沿中 $\boldsymbol{\Sigma}$ 值的估计方法有一定的改进. 另外, 这两种估计 $\boldsymbol{\mu}$ 的方法也被改进, 如 Black 和 Litterman (1991) 使用拟贝叶斯方法. 第三种方法使用 Bootstrap 方法来减小 $\widehat{\boldsymbol{w}}_{\text{eff}}$ 和 $\boldsymbol{w}_{\text{eff}}$ 之间的偏差.

2.4.1 多因子定价模型

在多因子定价模型中, 将每个资产的收益率 r_i 都使用 k 个因子 $f_1, \ldots, f_k$ 进行回归分析, 模型如下:

$$r_i = \alpha_i + (f_1, \ldots, f_k)^\top \boldsymbol{\beta}_i + \epsilon_i, \tag{2.4}$$

其中 α_i 和 $\boldsymbol{\beta}_i$ 是未知参数, ϵ_i 是均值为 0 且与各个因子 $\boldsymbol{f} := (f_1, \ldots, f_k)^\top$ 不相关的随机残差项. 在 $k = 1$ 时, 该模型称为*单因子模型*. CAPM 模型假设 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 都是已知的, 市场中存在无风险收益 r_f, 市场中的投资者在给定的收益率条件下, 最小化投资组合的方差. 因此, CAPM 模型对应于 $k = 1$, $\alpha_i = r_f$ 和 $f = r_M - r_f$ 的情况, 其中 r_M 是一个假想的市场投资组合的收益率, 在实际中 r_M 通常使用 S&P 500 指数等来进行估计. Ross (1976) 提出了套利定价模型 APT, 与 CAPM 模型不同, 该模型并不涉及市场投资组合或无风险利率. 模型认为在市场中不存在套利, 且风险资产的数量 m 充分大时, 可以利用 (2.4) 来对收益率进行近似估计. 在 APT 模型中, 对于因子的选择和个数并没有进行限制. (2.4) 中进行因子选择时, 通常使用经济学和统计学方法. 统计学方法主要包括主成分分析和因子分析等, 而 Fama-French 模型等常用的方法则是从经济学角度和实证观测出发来进行研究, 具体可参见 Lai 和 Xing (2008) 一书的 3.4 节, 同时该书的 2.9 节对于因子模型近年来的发展进行了回顾.

2.4.2 贝叶斯方法、缩减技术及 Black-Litterman 估计方法

对于均值为 $\boldsymbol{\mu}$, 协方差矩阵为 $\boldsymbol{\Sigma}$ 的独立同分布的随机向量 $\boldsymbol{r}_t$ 而言, 常用的共轭先验分布为逆 Wishart 分布, 即

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N(\boldsymbol{\nu}, \boldsymbol{\Sigma}/\kappa), \quad \boldsymbol{\Sigma} \sim IW_m(\boldsymbol{\Psi}, n_0), \tag{2.5}$$

其中 $IW_m(\boldsymbol{\Psi}, n_0)$ 是自由度为 n_0 的逆 Wishart 分布, 其均值为 $\boldsymbol{\Psi}/(n_0 - m - 1)$. 给定 $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n)$ 时, $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的后验分布具有相同的形式, 如下所示:

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N(\widehat{\boldsymbol{\mu}}, \boldsymbol{\Sigma}/(n + \kappa)), \quad \boldsymbol{\Sigma} \sim IW_m((n + n_0 - m - 1)\widehat{\boldsymbol{\Sigma}}, n + n_0),$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 $\widehat{\boldsymbol{\mu}}$ 和 $\widehat{\boldsymbol{\Sigma}}$ 分别是 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的贝叶斯估计量. 在进行贝叶斯估计时, 其目标是缩减均值 $\bar{\boldsymbol{r}}$ 和样本协方差矩阵 $n^{-1}\sum_{i=1}^{n}(\boldsymbol{r}_i-\bar{\boldsymbol{r}})(\boldsymbol{r}_i-\bar{\boldsymbol{r}})^{\top}$ 相应于其先验值而言的均值. 具体可参见 Lai 和 Xing (2008) 一书的 4.3.3 节.

Ledoit 和 Wolf (2003, 2004) 通过 $\bar{\boldsymbol{r}}$ 来估计 $\boldsymbol{\mu}$, 提出了一种将 $\boldsymbol{\Sigma}$ 的最大似然估计值朝结构化的协方差矩阵压缩的方法. 与直接使用贝叶斯估计不同, 该方法不需要将 $\boldsymbol{\mu}$, κ, n_0 和 $\boldsymbol{\Psi}$ 具体化. 该方法的原理在于, 当 $m(m+1)/2$ 充分大到与 n 可比时, 结构化的协方差矩阵的参数个数会更少, 因此估计值的方差会变得更小. 利用 $\widehat{\boldsymbol{F}}$ 和 $\boldsymbol{S}$, $\boldsymbol{\Sigma}$ 的估计值是:

$$\widehat{\boldsymbol{\Sigma}}=\widehat{\delta}\widehat{\boldsymbol{F}}+(1-\widehat{\delta})\boldsymbol{S}, \tag{2.6}$$

其中 $\widehat{\delta}$ 是将最大似然估计量向结构化的协方差矩阵 $\widehat{\boldsymbol{F}}$ 进行压缩时的最小压缩常数 δ 的估计量. 除了提出单因子模型的协方差矩阵 $\boldsymbol{F}$ 的估计方法外, Ledoit 和 Wolf (2003, 2004) 对于各因子之间相互相关的模型进行研究, 并提出了在因子之间相关性相同时 $\boldsymbol{F}$ 的估计方法. 研究发现该方法在数值模拟和实证研究中都有不错的效果. 同时, 在计算 Markowitz 有效前沿时, 该方法使用的是压缩估计量, 而不是 $\boldsymbol{S}$.

在估计投资组合的收益率均值 $\boldsymbol{\mu}$ 时, 很难得到完美的可靠结果. Black 和 Litterman (1991) 对于困难之处进行了总结, 并提出了一个实用的贝叶斯方法解决这个困难. Jorion (1986) 使用了一个类似于贝叶斯估计量 $\widehat{\boldsymbol{\mu}}$ 的压缩估计量, 而 Black-Litterman 方法则是用“均衡投资组合”的 $\boldsymbol{\mu}$ 估计值来替代投资者的主观估计值. 其中投资者的主观估计通常是基于过去的观察和投资者的个人观点或者专家的观点, 而得到各资产收益率的一个线性组合. 这些观点可以用 $\boldsymbol{P}\boldsymbol{\mu}\sim N(\boldsymbol{q},\boldsymbol{\Omega})$ 来表示, 其中 $\boldsymbol{P}$ 是一个描述投资者对于各个资产选择的 $p\times m$ 矩阵, 而 $\boldsymbol{\Omega}$ 则是用来描述投资者对自己观点不确定性的对角矩阵. “均衡投资组合”是基于均衡市场理论得到的投资组合, 记作 $\widetilde{\boldsymbol{w}}$, 其中 $\widetilde{\boldsymbol{w}}$ 是均值–方差优化问题 $\max_{\boldsymbol{w}}(\boldsymbol{w}^{\top}\boldsymbol{\pi}-\lambda\boldsymbol{w}^{\top}\boldsymbol{\Sigma}\boldsymbol{w})$ 的最优解, 而 λ 是指市场中的平均风险厌恶水平, $\boldsymbol{\pi}$ 则表示市场中对 $\boldsymbol{\mu}$ 的估计. 由上可知, 各变量满足 $\boldsymbol{\pi}=2\lambda\boldsymbol{\Sigma}\widetilde{\boldsymbol{w}}$, 利用市场资本化程度或市场基准投资组合估计得到 $\boldsymbol{\mu}$ 后, 可以对于 $\boldsymbol{\pi}$ 进行估计. 考虑到市场上对于 $\boldsymbol{\mu}$ 的估计值的不确定性, Black-Litterman 假设 $\boldsymbol{\pi}-\boldsymbol{\mu}\sim N(\boldsymbol{0},\tau\boldsymbol{\Sigma})$, 其中 $\tau\in(0,1)$, 且设定外生变量 $\lambda=1.2$. 结合 $\boldsymbol{P}\boldsymbol{\mu}\sim N(\boldsymbol{q},\boldsymbol{\Omega})$, $\boldsymbol{\pi}-\boldsymbol{\mu}\sim N(\boldsymbol{0},\tau\boldsymbol{\Sigma})$ 且假定两个正态过程相互独立, 则可以得到如下估计量, 称为 Black-Litterman 估计量:

$$\widehat{\boldsymbol{\mu}}^{\mathrm{BL}}=\left[(\tau\boldsymbol{\Sigma})^{-1}+\boldsymbol{P}^{\top}\boldsymbol{\Omega}^{-1}\boldsymbol{P}\right]^{-1}\left[(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi}+\boldsymbol{P}^{\top}\boldsymbol{\Omega}^{-1}\boldsymbol{q}\right], \tag{2.7}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中协方差矩阵定义为 $[(\tau\boldsymbol{\Sigma})^{-1}+\boldsymbol{P}^{\top}\boldsymbol{\Omega}^{-1}\boldsymbol{P}]^{-1}$. Black-Litterman 方法的重点在于对投资组合的均值 $\boldsymbol{\mu}$ 进行估计, 而对于 (2.7) 中涉及的 $\boldsymbol{\Sigma}$, 则使用历史数据的样本协方差来进行估计. 当然, 也可以使用贝叶斯方法或压缩估计的方法来估计 $\boldsymbol{\Sigma}$. 一系列论文对于 (2.7) 进行了后续研究, 如 Meucci (2005, 2010) 和 Bertsimas 等 (2012).

2.4.3 Bootstrap 方法和重抽样方法前沿

为了调整估计值 $\widehat{\boldsymbol{w}}_{\text{eff}}$ 与 $\boldsymbol{w}_{\text{eff}}$ 之间的偏误, Michaud (1989) 使用 Bootstrap 估计法的均值来作为 $\boldsymbol{w}_{\text{eff}}$ 的估计值, 如下:

$$\bar{\boldsymbol{w}}=B^{-1}\sum_{b=1}^{B}\widehat{\boldsymbol{w}}_b^*, \tag{2.8}$$

其中 $\widehat{\boldsymbol{w}}_b^*$ 是对于观测样本 $\{\boldsymbol{r}_1,\ldots,\boldsymbol{r}_n\}$ 进行有放回抽样时, 第 b 次 Bootstrap 样本 $\{\boldsymbol{r}_{b1}^*,\ldots,\boldsymbol{r}_{bn}^*\}$ 的最优投资组合的权重向量. 定义第 b 次 Bootstrap 抽样的样本均值和协方差矩阵分别为 $\widehat{\boldsymbol{\mu}}_b^*$ 和 $\widehat{\boldsymbol{\Sigma}}_b^*$, 则用 $\widehat{\boldsymbol{\mu}}_b^*$ 和 $\widehat{\boldsymbol{\Sigma}}_b^*$ 来替代 (2.1) 中的 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 即可得到相应的权重向量 $\widehat{\boldsymbol{w}}_b^*$. 因此, 重抽样的有效前沿可以利用 $\bar{\boldsymbol{w}}^{\top}\widehat{\boldsymbol{\mu}}$ 和 $\sqrt{\bar{\boldsymbol{w}}^{\top}\widehat{\boldsymbol{\Sigma}}\bar{\boldsymbol{w}}}$ 得到, 其中 $\bar{\boldsymbol{w}}$ 根据 (2.8) 得到, 而 $\widehat{\boldsymbol{w}}_b^*$ 则是由目标收益率 $\boldsymbol{\mu}_*$ 决定.

2.5 一种考虑了参数不确定性的新方法

嵌入式 (Plug-in) 的有效前沿是指使用 $\boldsymbol{S}$ 来估计 $\boldsymbol{\Sigma}$, 使用 $\bar{\boldsymbol{r}}$ 来估计 $\boldsymbol{\mu}$ 所得的有效前沿. 它包括两种变形情况: (1) 使用 (2.6) 来估计 $\boldsymbol{\Sigma}$、利用 Black-Litterman 方法来估计 $\boldsymbol{\mu}$; (2) 使用重抽样方法来估计有效前沿. 嵌入式方法的困难之处在于, Markowitz 的方法使用 $\boldsymbol{w}^{\top}\boldsymbol{r}_{n+1}$ 的方差来作为投资组合的风险度量, 但是投资组合的风险并不能简单地利用嵌入式方法的用 $\boldsymbol{w}^{\top}\widehat{\boldsymbol{\Sigma}}\boldsymbol{w}$ 来估计 $\text{Var}(\boldsymbol{w}^{\top}\boldsymbol{r}_{n+1})$ 和用 $\boldsymbol{w}^{\top}\widehat{\boldsymbol{\mu}}$ 来估计 $E(\boldsymbol{w}^{\top}\boldsymbol{r}_{n+1})$ 的结果来衡量. Broadie (1993) 中提到了这一问题, 并且使用 "真实有效前沿" 和 "估计有效前沿" 分别来指代 Markowitz 的有效前沿 (已知 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的情况) 和嵌入式有效前沿. 论文同时建议使用估计有效前沿投资组合的 "真实" 均值和方差. 然而当给定平均收益率 $E(\boldsymbol{w}^{\top}\boldsymbol{r}_{n+1})$ 的值 μ_* 时, 最小化 $\text{Var}(\boldsymbol{w}^{\top}\boldsymbol{r}_{n+1})$ 的问题在 Markowitz 的框架下才有意义, 此时 $E(\boldsymbol{r}_{n+1})$ 和 $\text{Cov}(\boldsymbol{r}_{n+1})$ 已知, 忽略 $\widehat{\boldsymbol{\mu}}$ 和 $\widehat{\boldsymbol{\Sigma}}$ 自身误差的情况下, 替代问题是在给定约束条件 $\boldsymbol{w}^{\top}\widehat{\boldsymbol{\mu}}=\mu_*$ 时最小化 $\boldsymbol{w}^{\top}\widehat{\boldsymbol{\Sigma}}\boldsymbol{w}$. Lai, Xing 和 Chen (2011b)(在本章后面部分简称 LXC) 研究了如下问题:

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$$\max\left\{E(\boldsymbol{w}^\top \boldsymbol{r}_{n+1}) - \lambda \mathrm{Var}(\boldsymbol{w}^\top \boldsymbol{r}_{n+1})\right\}, \tag{2.9}$$

其中 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 未知, 且作为状态变量, 其不确定性可以用给定贝叶斯后验分布来计算. (2.9) 中的权重 $\boldsymbol{w}$ 是由 $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n$ 所决定的随机向量. 注意到, 如果先验分布退化为 $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ 处的质点, 则 (2.9) 中的优化问题可以退化成已知 $\boldsymbol{\mu}_0$ 和 $\boldsymbol{\Sigma}_0$ 的 Markowitz 问题. (2.9) 中的拉格朗日乘子项 λ 可以被理解为在利用方差来作为风险量度时投资者的风险厌恶系数.

2.5.1 优化问题的求解

由于方差项 $\mathrm{Var}(\boldsymbol{w}^\top \boldsymbol{r}_{n+1}) = E[(\boldsymbol{w}^\top \boldsymbol{r}_{n+1})^2] - [E(\boldsymbol{w}^\top \boldsymbol{r}_{n+1})]^2$ 中包含 $[E(\boldsymbol{w}^\top \boldsymbol{r}_{n+1})]^2$ 这一项, 问题 (2.9) 并不是一个标准的随机优化问题. 在贝叶斯设定下, 标准的优化问题的形式为 $\max\limits_{a\in\mathcal{A}} Eg(\boldsymbol{X}, \boldsymbol{\theta}, a)$, 其中 $g(\boldsymbol{X}, \boldsymbol{\theta}, a)$ 是在采取行动 a 时的收益值, $\boldsymbol{X}$ 是分布 $F_{\boldsymbol{\theta}}$ 上的随机变量, $\boldsymbol{\theta}$ 服从特定的先验分布, 且该优化问题是在某个行动集 $\mathcal{A}$ 下进行优化. 在求解该问题时, 其关键是使用全期望公式 $Eg(\boldsymbol{X}, \boldsymbol{\theta}, a) = E\{E[g(\boldsymbol{X}, \boldsymbol{\theta}, a)|\boldsymbol{X}]\}$ 将该问题转化为选择相应的行动 a 来最大化后验收益值 $E[g(\boldsymbol{X}, \boldsymbol{\theta}, a)|\boldsymbol{X}]\}$. LXC 通过加入一个新的变量将 (2.9) 转化为一个标准的随机优化问题. 定义 $W = \boldsymbol{w}^\top \boldsymbol{r}_{n+1}$, 且注意到 $E(W) - \lambda \mathrm{Var}(W) = h(EW, EW^2)$, 其中 $h(x, y) = x + \lambda x^2 - \lambda y$. 定义 $W_B = \boldsymbol{w}_B^\top \boldsymbol{r}_{n+1}$, $\eta = 1 + 2\lambda E(W_B)$, 其中 $\boldsymbol{w}_B$ 是贝叶斯权重向量, 则

$$\begin{aligned}
0 &\geqslant h(EW, EW^2) - h(EW_B, EW_B^2) \\
&= E(W) - E(W_B) - \lambda\{E(W^2) - E(W_B^2)\} + \lambda\{(EW)^2 - (EW_B)^2\} \\
&= \eta\{E(W) - E(W_B)\} + \lambda\{E(W_B^2) - E(W^2)\} + \lambda\{E(W) - E(W_B)\}^2 \\
&\geqslant \{\lambda E(W_B^2) - \eta E(W_B)\} - \{\lambda E(W^2) - \eta E(W)\}.
\end{aligned}$$

在 $EW = EW_B$ 时, 此时第一个不等式满足 $EW^2 = EW_B^2$, 此时最后一行的不等式取等号. 在其余情况下, 最后一行的不等式都严格取不等号. 因此, 最后一项的取值始终小于等于 0.

$$\lambda E(W^2) - \eta E(W) \geqslant \lambda E(W_B^2) - \eta E(W_B), \tag{2.10}$$

此处等号当且仅当 W 与 W_B 的均值与方差都相同时才成立. 因此, (2.9) 中的随机优化问题等价于根据 $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n$, 寻找最优的权重向量 $\boldsymbol{w}$ 以最小化 $\lambda E[(\boldsymbol{w}^\top \boldsymbol{r}_{n+1})^2] - \eta E(\boldsymbol{w}^\top \boldsymbol{r}_{n+1})$. 由于 $\eta = 1 + 2\lambda E(W_B)$ 是 (2.9) 的解的一个线性函数, 因此并不能等价地用于 η. 在 LXC 中, 对于一族关于 η 的标准优化问题进行求解, 然后选择能使得 (2.9) 的收益最大的 η 来作为最优解.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

综上所述, LXC 通过将 (2.9) 的问题转化为求解如下关于 η 的优化问题:

$$\max_{\eta} \left\{ E\left[\boldsymbol{w}^{\top}(\eta)\boldsymbol{r}_{n+1}\right] - \lambda \mathrm{Var}\left[\boldsymbol{w}^{\top}(\eta)\boldsymbol{r}_{n+1}\right] \right\}, \tag{2.11}$$

其中 $\boldsymbol{w}(\eta)$ 是如下随机优化问题的解:

$$\boldsymbol{w}(\eta) = \arg\min_{\boldsymbol{w}} \left\{ \lambda E\left[(\boldsymbol{w}^{\top}\boldsymbol{r}_{n+1})^2\right] - \eta E(\boldsymbol{w}^{\top}\boldsymbol{r}_{n+1}) \right\}. \tag{2.12}$$

2.5.2 最优权重向量的计算

过去和现在的收益率 $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n$ 的集合 $\mathcal{R}_n$ 的后验分布的一阶矩和二阶矩分别为 $\boldsymbol{\mu}_n$ 和 $\boldsymbol{V}_n$. 由于 $\boldsymbol{w}$ 取决于 $\mathcal{R}_n$, 根据 $E(\boldsymbol{r}_{n+1}|\mathcal{R}_n) = \boldsymbol{\mu}_n$ 和 $E(\boldsymbol{r}_{n+1}\boldsymbol{r}_{n+1}^{\top}|\mathcal{R}_n) = \boldsymbol{V}_n$ 可知:

$$E(\boldsymbol{w}^{\top}\boldsymbol{r}_{n+1}) = E(\boldsymbol{w}^{\top}\boldsymbol{\mu}_n), \quad E\left[(\boldsymbol{w}^{\top}\boldsymbol{r}_{n+1})^2\right] = E(\boldsymbol{w}^{\top}\boldsymbol{V}_n\boldsymbol{w}). \tag{2.13}$$

在不允许卖空时, (2.11) 中的权重向量 $\boldsymbol{w}(\eta)$ 可以由 (2.1) 类推得到:

$$\boldsymbol{w}(\eta) = \arg \min_{\substack{\boldsymbol{w}:\boldsymbol{w}^{\top}\mathbf{1}=1,\\ \boldsymbol{w}\geqslant\mathbf{0}}} \left\{ \lambda \boldsymbol{w}^{\top}\boldsymbol{V}_n\boldsymbol{w} - \eta \boldsymbol{w}^{\top}\boldsymbol{\mu}_n \right\}, \tag{2.14}$$

该问题可以使用二次规划方法进行求解 (如使用 MATLAB 软件的 quadprog 命令). 而当允许卖空, 但是对于卖空的上限有限制时, 则约束条件 $\boldsymbol{w} \geqslant \mathbf{0}$ 可替换为 $\boldsymbol{w} \geqslant \boldsymbol{w}_0$, 其中 $\boldsymbol{w}_0$ 是一个负向量. 而当允许完全卖空时, (2.14) 中的卖空条件可以被忽略, (2.12) 式中的 $\boldsymbol{w}(\eta)$ 可以直接由下式得到:

$$\begin{aligned} \boldsymbol{w}(\eta) &= \arg \min_{\boldsymbol{w}:\boldsymbol{w}^{\top}\mathbf{1}=1} \left\{ \lambda \boldsymbol{w}^{\top}\boldsymbol{V}_n\boldsymbol{w} - \eta \boldsymbol{w}^{\top}\boldsymbol{\mu}_n \right\} \\ &= \frac{1}{C_n}\boldsymbol{V}_n^{-1}\mathbf{1} + \frac{\eta}{2\lambda}\boldsymbol{V}_n^{-1}\left(\boldsymbol{\mu}_n - \frac{A_n}{C_n}\mathbf{1}\right), \end{aligned} \tag{2.15}$$

其中第二个等式可以由拉格朗日乘子法得到, 且在上式中

$$A_n = \boldsymbol{\mu}_n^{\top}\boldsymbol{V}_n^{-1}\mathbf{1} = \mathbf{1}^{\top}\boldsymbol{V}_n^{-1}\boldsymbol{\mu}_n, \quad B_n = \boldsymbol{\mu}_n^{\top}\boldsymbol{V}_n^{-1}\boldsymbol{\mu}_n, \quad C_n = \mathbf{1}^{\top}\boldsymbol{V}_n^{-1}\mathbf{1}.$$

对于比 (2.14) 中的约束更加一般的线性或二次约束, $\boldsymbol{w}(\eta)$ 可以由二次规划问题求解.

定义给定 $\mathcal{R}_n$ 时的后验协方差矩阵为 $\boldsymbol{\Sigma}_n$, 根据全期望公式, 可知 $\mathrm{Var}(W)$ 满足:

$$\begin{aligned} \mathrm{Var}(W) &= E\left[\mathrm{Var}(W|\mathcal{R}_n)\right] + \mathrm{Var}\left[E(W|\mathcal{R}_n)\right] \\ &= E(\boldsymbol{w}^{\top}\boldsymbol{\Sigma}_n\boldsymbol{w}) + \mathrm{Var}(\boldsymbol{w}^{\top}\boldsymbol{\mu}_n). \end{aligned} \tag{2.16}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

此处使用 $\boldsymbol{\Sigma}_n$ 来取代 $\boldsymbol{\Sigma}$, 因为后者假定 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 是已知的, 因而忽略了 (2.16) 中 $\boldsymbol{w}^\top\boldsymbol{\mu}_n$ 的方差项, 这也是与嵌入式有效前沿相关的 Markowitz 方法与传统的有效前沿的区别所在. 为了使建模方法更加通用, 可以放宽前面提到的贝叶斯方法中对于先验分布的要求, 允许加入一些未指定的超参数. 这些超参数可以根据样本用最大似然方法或矩估计等方法得到. 举例来说, 在 (2.5) 的对偶问题中, 可以假设 $\boldsymbol{\nu}$ 和 $\boldsymbol{\Psi}$ 是与 (2.4) 类似的多因子模型相关的超参数的函数. 在此情况下, 使用经验贝叶斯方法来估计随机优化模型 (2.4) 中的参数 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. 在问题 (2.4) 中, 除了 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的先验分布外, 还需要使用独立同分布的收益率的分布信息来估计 $E_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{w}^\top\boldsymbol{r}_{n+1})$ 和 $\mathrm{Var}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{w}^\top\boldsymbol{r}_{n+1})$. 下一节中的 Bootstrap 方法提出了一个非参数方法来估计这些值.

2.5.3 收益率的 Bootstrap 估计和 NPEB 法则

首先, 注意到 Bootstrap 方法可以用于评估资产配置准则的大概率表现. 根据观测样本 $\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n\}$, $1 \leqslant b \leqslant B$ 进行 Bootstrap 抽样, 得到样本 $\{\boldsymbol{r}^*_{b1}, \ldots, \boldsymbol{r}^*_{bn}\}$. 对于一个权重向量 $\boldsymbol{w}_n$ 取决于 $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n$ 和无信息先验分布的 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的投资组合 Π 而言, 其收益率的均值和方差可以根据 Bootstrap 样本计算, 分别为 $E_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{w}_n^\top\boldsymbol{r}_{n+1}) = E_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{w}_n^\top\boldsymbol{\mu})$ 和 $\mathrm{Var}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{w}_n^\top\boldsymbol{r}_{n+1}) = E_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{w}_n^\top\boldsymbol{\Sigma}\boldsymbol{w}_n) + \mathrm{Var}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{w}_n^\top\boldsymbol{\mu})$. 另外, 可以使用贝叶斯估计等方法来估计 (2.14) 或 (2.15) 中的 $\boldsymbol{\mu}_n$ 和 $\boldsymbol{V}_n$, 然后选择 η 来最大化 Bootstrap 估计量 $E_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{w}_n^\top\boldsymbol{r}_{n+1}) - \lambda\mathrm{Var}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{w}_n^\top\boldsymbol{r}_{n+1})$. LXC 论文中, 使用 $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n$ 的经验分布来作为收益率的分布, 并使用无信息先验分布的 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 然后在 (2.14) 中使用 $\bar{\boldsymbol{r}}$ 来替代 $\boldsymbol{\mu}_n$, 在 (2.15) 中使用二阶矩 $n^{-1}\sum_{t=1}^n \boldsymbol{r}_t\boldsymbol{r}_t^\top$ 来替代 $\boldsymbol{V}_n$, 从而得到 2.5.1 节中的最优准则的一个 "非参数经验贝叶斯" (Non-parametric empirical Bayes) 变形形式, LXC 称之为 NPEB 准则. LXC 同时基于上述理论进行了数值模拟. 假设有 $m = 4$ 个资产, 其年收益率独立同分布, 收益率的均值和方差分别基于正态分布和逆 Wishart 分布 (参数为 $\kappa = 5$, $n_0 = 10$, $\boldsymbol{\nu} = (2.48, 2.17, 1.61, 3.42)^\top$), 超参数 $\boldsymbol{\Psi}$ 定义为 $\boldsymbol{\Psi}_{11} = 3.37$, $\boldsymbol{\Psi}_{22} = 4.22$, $\boldsymbol{\Psi}_{33} = 2.75$, $\boldsymbol{\Psi}_{44} = 8.43$, $\boldsymbol{\Psi}_{12} = 2.04$, $\boldsymbol{\Psi}_{13} = 0.32$, $\boldsymbol{\Psi}_{14} = 1.59$, $\boldsymbol{\Psi}_{23} = -0.05$, $\boldsymbol{\Psi}_{24} = 3.02$, $\boldsymbol{\Psi}_{34} = 1.08$. 在进行数值模拟时, 考虑了在 $n = 6$, 且不存在卖空限制时的四种不同的情形. 令 $\lambda = 1, 5, 10$, 在第一种情形中假设了相应的先验分布, 并研究了贝叶斯收益. 其他三种情形分别考虑了 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的三种取值下的大概率收益, 分别记为 Freq 1, Freq 2, Freq 3. 表 2.1 比较了 (2.9) 最大化问题的贝叶斯准则 (后面记为 "Bayes") 和如下三种准则: (a) "甲骨文" (Oracle) 准则, 即假设 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 都是已知的; (b) 嵌入式

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

(Plug-in) 准则, 即将 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 替代为其样本估计量; (c) NPEB 准则. 注意到在 (b) 和 (c) 中都使用了样本的均值向量和方差矩阵, 但在 (b) 中是使用样本估计量来替代 Oracle 准则下的参数, 而在 (c) 中则是使用经验分布来替代贝叶斯法则下的一般分布. 在嵌入式准则下, 在样本协方差矩阵接近奇异矩阵时, 二次规划问题不易进行数值模拟, 通常的处理办法是在协方差矩阵中加上 $0.005\boldsymbol{I}$. 表 2.1 对于 LXC 的数值模拟结果进行了总结, 每一个结果都是基于 500 次模拟得到. 在每种情形下, NPEB 准则的收益都与贝叶斯准则的结果相近, 且小于甲骨文准则下的结果. 嵌入式准则下的收益相对最小, 而在 λ 更大时, 区别更大.

表 2.1 四种准则下由 $m = 4$ 个资产组成的投资组合的收益

		Bayes	Plug-in	Oracle	NPEB
λ=1	Bayes	0.0324	0.0317	0.0328	0.0324
	Freq 1	0.0332	0.0324	0.0332	0.0332
	Freq 2	0.0293	0.0282	0.0298	0.0293
	Freq 3	0.0267	0.0257	0.0268	0.0267
λ=5	Bayes	0.0262	0.0189	0.0267	0.0262
	Freq 1	0.0272	0.0182	0.0273	0.0272
	Freq 2	0.0233	0.0183	0.0240	0.0234
	Freq 3	0.0235	0.0159	0.0237	0.0235
λ=10	Bayes	0.0184	0.0067	0.0190	0.0183
	Freq 1	0.0197	0.0063	0.0199	0.0198
	Freq 2	0.0157	0.0072	0.0168	0.0159
	Freq 3	0.0195	0.0083	0.0198	0.0196

2.6 从随机游走模型到更符合事实的鞅模型

在随机优化问题 (2.9) 和其解中, 即 2.5.2 节和 2.5.3 节中的 (2.14)–(2.15) 中, 我们都假设收益率向量 $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n+1}$ 独立同分布, 以便与文献中的其他方法进行比较. 此外, 我们假设在 NPEB 准则中, $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 服从无信息先验分布, 从

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

而使其表现可以与其他并未涉及先验分布的方法进行比较. 然而, 正如 LXC 中指出的那样, 我们应该使用一个改进的模型, 从而使得在对于现有的独立同分布问题进行检验时, 能够显示其不足之处. 在本节中, 我们继续对 2.2 节中的资产定价模型进行回顾, 然后介绍 LXC 论文中是如何将 (2.14) 和 (2.15) 应用到更符合 2.1.2 节中提到的资产收益率的时间序列模型中的. 在本节中, 按照时间顺序的简要概述也更多地侧重于 “数据” 和 “模型” (本书的副标题中的两个词), 而在介绍 LXC 如何在有效市场假说下使用第二代模型时, 则更多地侧重于 “分析” 与 “优化” (本书副标题中的另两个词).

2.6.1 从高斯形式到帕累托形式的随机游走

在 2.2 节中, 我们提到了 Kendall (1953) 及其他论文中对于对数化的股票收益率的随机游走模型的实证验证. Fama (1963) 总结了后续对于对数收益率的分布的研究, 这些研究通常认为对数收益率的增量是独立同分布的随机游走, 且根据 Mandelbrot (1963) 可知, 研究的重点为 “稳定帕累托假设”. “帕累托形式” 是指由帕累托首先提出的帕累托分布, 满足随着 $x \to \infty$, 随机变量大于 x 的概率为 $cx^{-\alpha}$. 帕累托分布在 $\alpha \leqslant 1$ 时均值无穷大, 而在 $1 < \alpha \leqslant 2$ 时方差无穷大. Mandelbrot (1963) 对于 1900—1905 年间每天的对数棉花价格的差值和 1880—1940 年间每月的对数棉花价格的差值进行了研究. 这些从美国农业部获取的数据显示出比正态分布更加厚尾的特征. 价格的分布是稳定分布 $\text{Stable}(\alpha, \beta, \gamma, \delta)$ 的一种特例. 特别地, $\alpha = 2$ 且 $\beta = 0$ 的情况对应于正态分布 $Y \sim N(\delta, \gamma^2)$, 而 $\alpha = 1$ 且 $\beta = \delta = 0$ 的情况对应于 $Y \sim \text{Cauchy}(\gamma)$, 其概率密度函数为 $f(y) = \{\pi\gamma(1 + (y/\gamma)^2)\}^{-1}$, $-\infty < y < \infty$. 稳定分布的一个重要的特征是其包含了所有的中心化并按比例调整的独立同分布随机变量 (也称为随机游走) 之和. 另一个重要的特征是其自身可分解: 定义独立同分布变量 $Y_1, \ldots, Y_n$ 服从 $\text{Stable}(\alpha, \beta, \gamma, \delta)$, 且 $Y \sim \text{Stable}(\alpha, \beta, \gamma, \delta)$, 则存在一系列的 $c_n > 0$ 和 d_n, 使得 $Y_1 + \cdots + Y_n$ 与 $c_n Y + d_n$ 具有相同的分布. 特别地, 对于 $\beta = \delta = 0$, 即对称稳定分布的情形, 可以选择常数 $c_n = n^{1/\alpha}$ 和 $d_n = 0$.

Mandelbrot 将非正态形式的分布称为 “稳定帕累托分布”, 因为这些分布是正态化的帕累托形式 (即满足幂法则). 由于稳定分布的概率密度函数 $\text{Stable}(\alpha, \beta, \gamma, \delta)$ 对于一些特殊的分布, 如正态分布、柯西分布、逆高斯分布 (即布朗运动首达时间的 Lévy 分布), 并没有显式的表达式. 因此对于稳定分布的参数的估计并没有通用的方法, 在 Mandelbrot 的奠基性论文发表后的五十年间, 对于参数估计方法的研究逐步深入.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

2.6.2 包含最优抽样时间的随机游走

除了将布朗运动换成更加一般的稳定过程 (即增量服从独立分布的过程), Mandelbrot 还用从属过程提出了另一个更加广义的过程, 从属过程是原过程的一个“随机时钟”. Mandelbrot 和 Taylor (1967) 中的从属过程模型使用 $B_{T(t)}$ 来对于价格 P_t 进行建模, 其中 $\{B_s, s \geqslant 0\}$ 是一个布朗过程, 而从属过程 $T(t)$ 则是一个增量为正的稳定过程, 且 $T(t)$ 与 $\{B_s, s \geqslant 0\}$ 相互独立. 另一个在金融中广泛使用的从属过程是 Madan 和 Seneta (1990) 提出的方差伽马过程, 其形式为 $\sigma B_{\tau(t)} + \theta\tau(t)$, 其中 $\tau(t)$ 是一个独立增量过程且服从均值为 $\gamma t/\lambda$、方差为 $\gamma t/\lambda^2$ 的伽马分布.

Mandelbrot 和 Taylor 提出的随机过程 $B_{T(t)}$ 具有平稳的增量, 且由于 $T(t)$ 是一个增量为正的稳定过程, $B_{T(t)}$ 的增量的方差为无穷大. Clark (1973) 去掉了 $T(t)$ 服从稳定过程这个假设, 而是假设 $v_t = T(t) - T(t-1)$ 是均值为 μ、方差为 $\sigma^2 > 0$ 的独立同分布的随机变量. 假设给定 v_t 时, $\Delta_t = P_t - P_{t-1}$ 服从 $N(0, \gamma^2 v_t)$ 分布. 因此, Δ_t 的均值为 0, 方差为 $\mathrm{Var}(\Delta_t) = E[\mathrm{Var}(\Delta_t|v_t)] = \gamma^2 E(V_t) = \gamma^2\mu$. 由于 $E(\Delta_t^4) = E[E(\Delta_t^4|v_t)] = 3\gamma^4 E(v_t^2) = 3\gamma^4(\mu^2 + \sigma^2)$, Δ_t 的峰度为

$$\frac{E(\Delta_t^4)}{[\mathrm{Var}(\Delta_t)]^2} = 3 \times \frac{\mu^2 + \sigma^2}{\mu^2} > 3,$$

因此, Δ_t 的分布为尖峰分布, 且具有有限的方差, 这也与对价格的限制条件相符合, 即价格在有限的时间区间内应该是有限的. Clark (1973) 放宽了独立增量 $T(t)$ 的平稳性这个假设, 他提出一个重要的观点, 认为 $v_t = T(t) - T(t-1)$ 可能会随着第 t 天的交易行为变化. 特别地, Clark 提出了一个第 t 天的交易量 V_t 的参数模型来估计 v_t. 具体而言, 他提出使用 AV_t^a 或 $\exp(B + bV_t)$ 来近似 $\gamma^2 v_t$.

例 2.2 Clark 的模型中, 在将 Δ_t 对 V_t 回归时, 代理变量包含非线性参数, 因此一个更方便的方式是对 $\log\Delta_t^2$ 进行分析. $\log\Delta_t^2$ 的分布为 $\gamma^2 v_t \chi_1^2$ 的对数的分布, 使用下面这两个模型之一:

$$\log\Delta_t^2 = \log A + a\log V_t + \varepsilon_t \quad \text{或} \quad \log\Delta_t^2 = B + bV_t + \varepsilon_t. \tag{2.17}$$

此处, 我们使用 2005 年 1 月 1 日—2014 年 12 月 31 日 Pfizer 公司的日收盘价和相应的交易量的数据, 在去掉 43 个 $\Delta_t = 0$ 的数据后, 对两组数据分别进行回归分析. 图 2.4 的上图描述了价格 P_t 的时间序列, 下图描述了价格变化 $\Delta_t = P_t - P_{t-1}$ 的时间序列特征. V_t 的均值大概是 10^7 量级, 参数估计值为

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

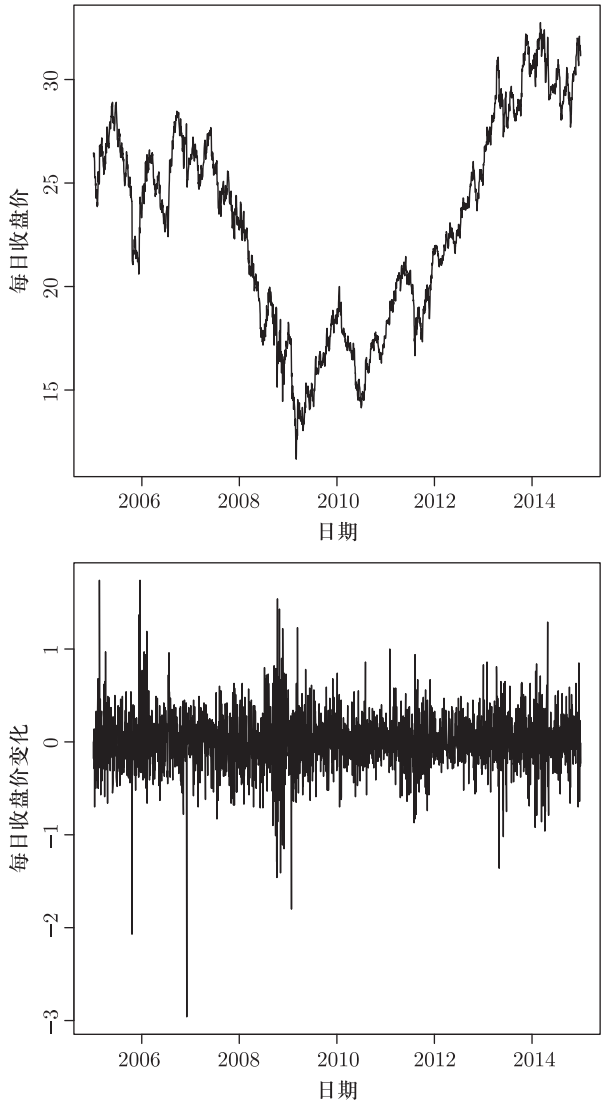


图 2.4 2005 年 1 月 1 日—2014 年 12 月 31 日 Pfizer 公司收盘价 P_t 的时间序列图 (上图) 和 Δ_t 的时间序列图 (下图)

$\log(\hat{A}) = -25.2$, $\hat{a} = 1.23$; $\hat{B} = -4.83$, $\hat{b} = 2.47 \times 10^{-8}$. 由于在 Clark 模型下, 满足 $\Delta_t/\sqrt{v_t} \sim N(0, \gamma^2)$, 我们此处使用下面两个量的正态 QQ 图:

$$z_t^{(a)} = \frac{\Delta_t}{\sqrt{\hat{A} \times V_t^{\hat{a}}}}, \qquad z_t^{(b)} = \frac{\Delta_t}{\sqrt{\exp\left\{\hat{B} + \hat{b} V_t\right\}}} \tag{2.18}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

来分析 (2.17) 中的回归模型是否合理.① 相比于图 2.6 的上图而言, 图 2.5 对于 Δ_t 的拟合性能更好. 另外, $z_t^{(a)}$ 和 $z_t^{(b)}$ 的峰度分别是 3.50 和 3.56, 相比于 Δ_t 的峰度值 9.90, 与 3 更接近.

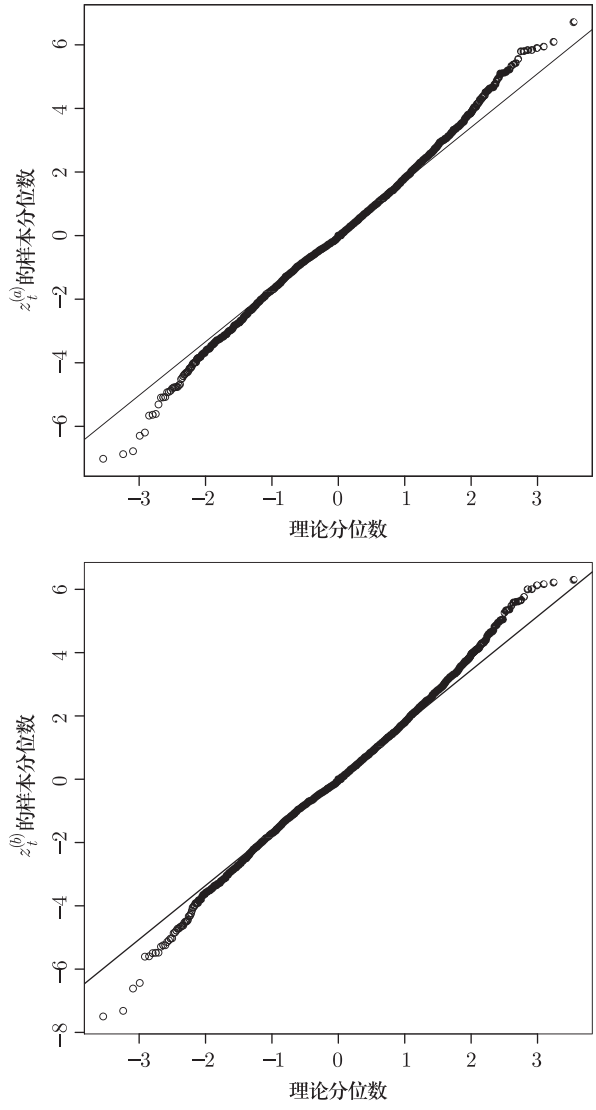


图 2.5 正态 QQ 图: $z_t^{(a)}$ (上图)、$z_t^{(b)}$ (下图), 其中 $z_t^{(a)}$ 和 $z_t^{(b)}$ 由 (2.18) 定义

① 尽管 $\hat{b}$ 大概是在 10^{-8} 量级, 但在分析 $z_t^{(b)}$ 时并不能直接忽略, 因为 V_t 的均值大概是 10^7 量级的.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

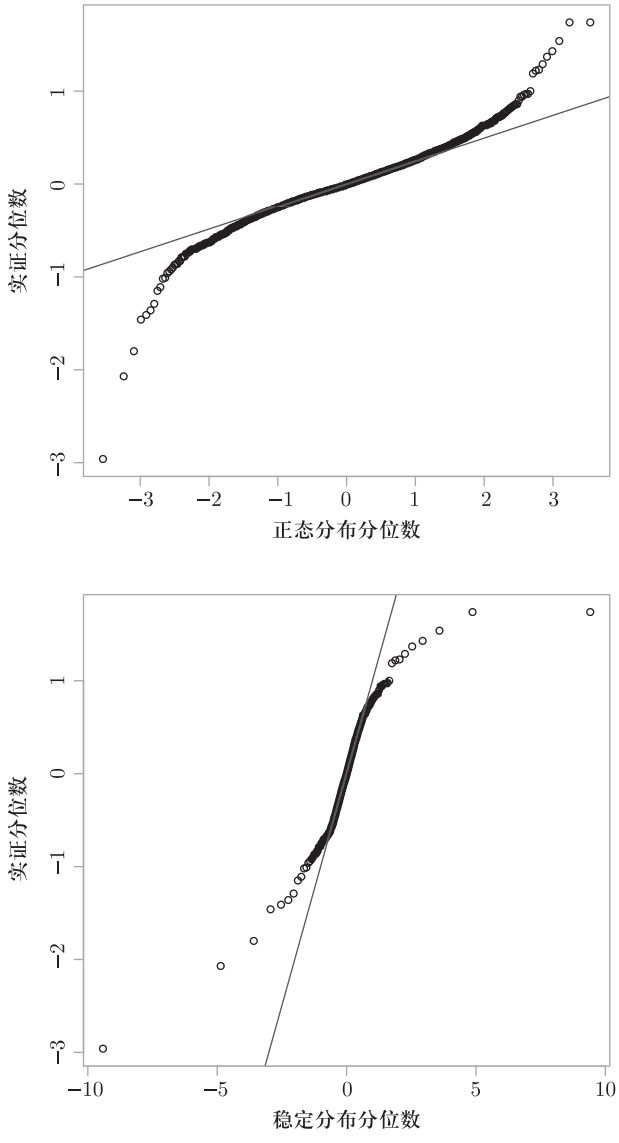


图 2.6　Δ_t 的正态分布拟合结果的 QQ 图 (上图) 和对称稳定分布拟合结果的 QQ 图 (下图)

2.6.3　从随机游走模型到 ARIMA、GARCH 和鞅回归等时间序列模型

Box 和 Jenkins (1970) 将 IBM 公司的日收盘价进行了变形 (取对数), 使得一阶差分后 (因此 ARIMA 模型中的 "I" 表示差分整合) 的数据可以用 ARMA (Autoregressive Moving Average) 模型来进行分析, 并分别利用自相关

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

函数和偏自相关函数来判断 ARMA 模型的阶数. 随后, Box 和 Pierce (1970) 以及 Ljung 和 Box (1978) 引入了基于样本自相关系数的统计检验量来检验时间序列数据的独立性. Lai 和 Xing (2008) 一书的第 5 章对于 Box-Jenkins 方法、对宽平稳模型的 Wold 分解以及相关的统计软件工具进行了总结. 在 Box-Jenkins 方法在计量经济学中占据主导地位后, 另一个主要的研究方向是研究波动聚类现象, 该现象最初是在研究股票收益率和其他时间序列模型中发现, Engle (1982) 对于该现象进行了研究, 并提出了 ARCH (自回归条件异方差) 模型. Bollerslev 将 ARCH 模型推广到 GARCH (广义自回归条件异方差) 模型, 该模型将 ARCH 模型中的独立同分布残差拓展为收益率平方 r_t^2 的鞅差分残差. Lai 和 Xing (2008) 一书的第 6 章对于 GARCH 模型及其变形 EGARCH 模型进行了介绍, EGARCH 模型 (指数广义自回归条件异方差模型) 可以解释 2.1.1 节中提到的杠杆效应. 第六章同时也将 ARMA 模型与 GARCH 模型或 EGARCH 模型相结合以分别反映 r_t 的水平和波动性特征, 即使用 ARMA-GARCH 或 ARMA-EGARCH 模型来分析 r_t 的特征, 书中对于相关的模型拟合和预测软件包进行了介绍. 一个最近开发的 R 语言包是 `rugarch`, 其中 "`u`" 指单变量 (Univariate), 而 "`r`" 指已实现的 (Realized).

因此, 在 1982 年 Engles 发表其奠基性论文时[①], 时间序列模型已经开始取代随机游走模型被用于对价格或对数价格进行建模, 因为前者更符合低频交易证券的对数收益率特征. 相对于 Mandelbrot, Taylor 和 Clark 在之前二十年的研究而言, 这样的变化是对于随机游走模型进行了更决定性的改进. 1982 年也是更加广义的基于时间序列数据的 "鞅回归" 方法开始应用的一年, 而 ARMA 模型和 GARCH 模型都是该方法的特殊情形. 该方法最初由 Lai 和 Wei (1982) 提出, 当时称为 "随机回归". Lai 和 Wei (1982) 使用该方法来研究随机输入–输出系统的行为, 其中输入项由一个反馈控制系统生成, 当前的输入 $\boldsymbol{x}_t$ 由过去的输出和输入项决定, 且当前的输出项为 $y_t = \boldsymbol{\beta}^\top \boldsymbol{x}_t + \varepsilon_t$, 其中 ε_t 是随机误差项, 且 ε_t 为一个鞅差分序列. 具体而言, 定义 $\mathcal{F}_t$ 为 $(\boldsymbol{x}_s, y_s), s \leqslant t$ 生成的 σ 域 (信息集), 假设在随机回归中满足 $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, 且 $\boldsymbol{x}_t$ 是 $\mathcal{F}_{t-1}$ 可测的. AR(p) 是在满足 $\boldsymbol{x}_t = (y_{t-1}, \ldots, y_{t-p})^\top$ 时的一个特例. ARMA(p, q) 模型相对而言更复杂, 因为其随机波动项来源于 $\varepsilon_t + c_1\varepsilon_{t-1} + \cdots + c_q\varepsilon_{t-q}$. 另一方面, 如果 $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ 是可观测的, 则该模型可以简化成一个满足 $\boldsymbol{x}_t = (y_{t-1}, \ldots, y_{t-p}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q})^\top$ 的随机回归过程. 这也是 Lai 和 Wei (1986) 使用广义的最小二乘法估计 ARMA 模型参数的基本思想. 当参数已知时, 可以

① Robert Engle 在 2003 年因为该论文在时间序列建模中的开拓性贡献, 被授予诺贝尔经济学奖.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

令 $\boldsymbol{x}_s = \mathbf{0}$ 以及在 $s \leqslant 0$ 时, $\varepsilon_s = 0$, 则可以从 $(\boldsymbol{x}_s, y_s), 1 \leqslant s \leqslant t$ 中得到 $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$, 这也是计算似然方程的过程. 类似于广义的最小二乘法, Lai 和 Ying (1991b) 将鞅理论应用到分析递归的最大似然估计中.

此处, 我们假设随机误差项 ε_t 服从鞅差分过程, 而不是独立同分布零均值过程. 该假设也使得在对 ε_t 进行时间序列分析时, 可以在波动率建模中加入动态变化项. 事实上, GARCH(h,k) 模型在满足 $\varepsilon_t = \sigma_t\zeta_t$, $\sigma_t^2 = \omega + \sum_{i=1}^{h}\beta_i\sigma_{t-i}^2 + \sum_{j=1}^{k}\alpha_j\varepsilon_{t-j}^2$ 且 ζ_t 独立同分布、均值为 0、方差为 1 时, 可以将 ε_t^2 写成 ARMA 模型形式:

$$\varepsilon_t^2 = \omega + \sum_{j=1}^{\max(h,k)}(\alpha_j + \beta_j)\varepsilon_{t-j}^2 + \eta_t - \sum_{i=1}^{h}\beta_i\eta_{t-i},$$

其中 $\eta_t = \varepsilon_t^2 - \sigma_t^2$ 是一个鞅差分项, 当 $j > k$ 时满足 $\alpha_j = 0$, 当 $i > h$ 时满足 $\beta_i = 0$, 具体参见 Lai 和 Xing (2008) 一书第 147 页. 因此, 在此处我们使用广义的 "鞅回归" 方法, 以便在随机回归模型 $y_t = \boldsymbol{\beta}^\top\boldsymbol{x}_t + \varepsilon_t$ 中加入鞅差分噪声项 ε_t 的动态波动率模型. 同时, 动态波动率模型中也包含鞅结构, 即鞅差分新息项 η_t. 这些鞅回归模型使用更一般的时间序列模型代替随机游走模型来对于 2.4 和 2.5 节中现代投资组合理论下的对数化股票价格建模. 同时, 这些鞅回归模型也更符合 2.1.1 节中提到的资产价格的事实. Lai 等 (即 2.5 节中 LXC 论文) 提出, 在对于收益率 r_{it} 进行建模时, 使用如下的鞅回归模型来替代 "现有的独立同分布模型":

$$r_{it} = \boldsymbol{\beta}_i^\top\boldsymbol{x}_{i,t-1} + \epsilon_{it}, \tag{2.19}$$

其中 $\boldsymbol{x}_{i,t-1}$ 中包含常数 1、$t-1$ 期的 S&P 500 等市场投资组合的收益率等因子和 $r_{i,t-1}, r_{i,t-2}, \ldots$ 变量的滞后项. (2.19) 的基本思想是引入协变量 (包括加入滞后项以考虑时间效应), 从而使得误差项 ϵ_{it} 可以被认为是独立同分布的, 就像在 "现有的独立同分布" 模型中一样. 回归模型的参数 $\boldsymbol{\beta}_i$ 可以使用矩估计得到, 结果等价于使用最小二乘法进行估计. 我们同时在模型中加入了异方差项, 假设 $\epsilon_{it} = s_{i,t-1}(\boldsymbol{\gamma}_i)z_{it}$, 其中 z_{it} 独立同分布、均值为 0、方差为 1, $\boldsymbol{\gamma}_i$ 是一个参数向量, 可以通过最大似然估计或者广义矩估计得到, $s_{i,t-1}$ 是一个关于 $r_{i,t-1}, r_{i,t-2}, \ldots$ 的给定的函数. 一个典型的例子是 GARCH$(1,1)$ 模型:

$$\epsilon_{it} = s_{i,t-1}z_{it}, \quad s_{i,t-1}^2 = \omega_i + a_i s_{i,t-2}^2 + b_i r_{i,t-1}^2, \tag{2.20}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 $\boldsymbol{\gamma}_i = (\omega_i, a_i, b_i)$. 这些鞅回归模型表明, 在半强形式的有效市场假说下, 鞅过程远远比随机游走过程更适用于获取有效信息, 这也验证了 1.7 节的补充材料 6 中提到的 LeRoy 和 Lehmann 的观点.

2.7 新现代投资组合理论下的鞅回归模型

在本节中, 我们继续讨论 2.5 节提到的 LXC 的新方法, 讨论现代投资组合理论中当期资产收益率 $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n$ 已知, 但下一期的收益率 $\boldsymbol{r}_{n+1}$ 未知的情况. 在 2.5.1 和 2.5.2 节中, 最优权重向量 $\boldsymbol{w}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n)$ 的求解和计算方法也适用于其他对数据拟合的效果更好的模型, 这些模型相比于 2.5.3 节提到的独立同分布而言, 在相同的假设下更符合数据的特征. LXC 提出使用符合有效市场假说的鞅回归模型 (2.20) 来替代不符合资产收益率特征的独立同分布模型. 在将 LXC 的新方法应用到现代投资组合理论, 或者说 "新现代投资组合理论" (Neomodern Portfolio Theory, Neo-MPT) 时的另一个问题是如何选择 (2.9) 中的 λ. 在讨论 LXC 论文如何将 NPEB 准则应用到鞅回归模型 (2.20) 之后, 我们在 2.7.2 节中讨论了该问题. 最后, 2.7.3 节介绍了 LXC 论文中对于新现代投资组合理论的实证研究.

2.7.1 在 NPEB 中加入时间序列效应

正如 2.5.2 节中提到的, 在求解优化问题 (2.9) 中的最优权重向量时, 一个关键的步骤是求解 $(\boldsymbol{\mu}_n, \boldsymbol{V}_n)$, 其中 $\boldsymbol{\mu}_n = E(\boldsymbol{r}_{n+1}|\mathcal{R}_n)$, $\boldsymbol{V}_n = E(\boldsymbol{r}_{n+1}\boldsymbol{r}_{n+1}^{\top}|\mathcal{R}_n)$. 与经典的独立同分布收益率模型不同, 我们可以将 m 个资产的信息和时间序列建模相结合, 从而利用 $\boldsymbol{\mu}_n$ 和 $\boldsymbol{V}_n$ 对于未来收益率进行更好的估计. 我们选择 (2.19)–(2.20) 中的回归变量 $\boldsymbol{x}_{i,t-1}$ 来建立一个综合的有效–实证预测模型, 具体可参考 Lai 和 Xing (2008) 一书 7.5 节. 由于模型 (2.19)–(2.20) 的目的是生成独立同分布的 $\boldsymbol{z}_t = (z_{1t}, \ldots, z_{mt})^{\top}$, 我们仍然可以使用 NPEB 方法来确定最优权重向量, 根据 $\boldsymbol{z}_t$ 的分布来进行 Bootstrap 抽样. 注意到 (2.19)–(2.20) 是将不同资产的收益率分别进行建模, 而不是建立一个多元回归或多元 GARCH 模型这样的联立模型, 这样也避免了模型中需要估计的参数过多的问题. 尽管 $\boldsymbol{z}_t$ 是独立同分布的, 但在 (2.20) 中并没有假定各个部分是相互不相关的, 而是将各个部分分别建模, 并没有建立联立模型. 在模型 (2.19)–(2.20) 中, 给定 $\mathcal{R}_n$ 时, 资产 i 和资产 j 的收益率的截面数据的条件协方差是:

$$\mathrm{Cov}(r_{i,n+1}, r_{j,n+1}|\mathcal{R}_n) = s_{i,n}(\boldsymbol{\gamma}_i)s_{j,n}(\boldsymbol{\gamma}_j)\mathrm{Cov}(z_{i,n+1}, z_{j,n+1}|\mathcal{R}_n), \qquad (2.21)$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

注意到在 (2.20) 中, $s_{i,n}^2$ 是通过 $\mathcal{R}_n$ 递归求得, 而 $\boldsymbol{z}_{n+1}$ 与 $\mathcal{R}_n$ 相互独立, 因此协方差矩阵可以利用残差项 $\widehat{\boldsymbol{z}}_t$ 进行一致估计. NPEB 方法利用如下的方法估计 (2.14) 中的 $\boldsymbol{\mu}_n$ 和 $\boldsymbol{V}_n$:

$$\boldsymbol{\mu}_n = (\widehat{\boldsymbol{\beta}}_1^\top \boldsymbol{x}_{1,n}, \ldots, \widehat{\boldsymbol{\beta}}_m^\top \boldsymbol{x}_{m,n})^\top, \quad \boldsymbol{V}_n = \boldsymbol{\mu}_n \boldsymbol{\mu}_n^\top + (\widehat{s}_{i,n}\widehat{s}_{j,n}\widehat{\sigma}_{ij})_{1\leqslant i,j\leqslant n},$$

其中 $\widehat{\boldsymbol{\beta}}_i$ 是 $\boldsymbol{\beta}_i$ 的最小二乘估计量, $\widehat{s}_{l,n}$ 和 $\widehat{\sigma}_{ij}$ 是基于 $\mathcal{R}_n$ 对 $s_{l,n}$ 和 $\mathrm{Cov}(z_{i,1}, z_{j,1})$ 的估计量.

2.7.2 有效前沿上的最优信息比率

2.3 节提到, 2.5 节中的 λ 与投资者在最大化投资组合的期望效用时的风险厌恶程度有关. 该参数体现了对风险的惩罚值, 此处风险是以投资组合收益率的方差来衡量. 但是在实际中, 很难对于 2.5.1 节中提到的风险厌恶系数 λ 进行具体量化. 在研究投资组合的业绩, 特别是比较基金业绩时, 一个常用的衡量标准是信息比率 $(\mu - \mu_0)/\sigma_e$, 即单位风险的超额收益率. 超额风险定义为 $\mu - \mu_0$, 其中 $\mu_0 = E(r_0)$, 而 r_0 是基准投资收益率, σ_e^2 是收益率的方差. 此处可以将 λ 看成调整参数, 用来选择合适的 λ 来最大化信息比率. 在计算信息比率时, 可以对于 2.5.2 节中的 NPEB 方法进行修改, 使用 $E_{\boldsymbol{\mu},\boldsymbol{\Sigma}}[\boldsymbol{w}^\top(\eta)\boldsymbol{r}] - \lambda \mathrm{Var}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}[\boldsymbol{w}^\top(\eta)\boldsymbol{r}]$ 的 Bootstrap 抽样估计值来寻找优化问题 (2.9) 中的最优权重向量 $\boldsymbol{w}_\lambda$. 具体而言, 我们可以使用如下的 Bootstrap 估计量来估计 $\boldsymbol{w}_\lambda$, 相应的信息比率为

$$E_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{w}_\lambda \boldsymbol{r} - r_0) \Big/ \sqrt{\mathrm{Var}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{w}_\lambda^\top \boldsymbol{r} - r_0)} \tag{2.22}$$

然后基于格点来寻找使得信息比率最大的 λ 值. 具体方法会在 2.7.3 节中介绍.

2.7.3 新现代投资组合理论的实证检验

在本节中, 我们对于 LXC 方法的样本外表现的实证研究进行介绍, 并对于资产收益率的均值和方差未知时的均值–方差资产组合方法进行介绍. 在研究时, 使用的数据为 1985 年 1 月至 2009 年 12 月间的月交易数据, 数据来源为 CRSP 数据库①. 在进行实证研究时, 论文对于存在时间在 10 年 (120 个月) 以上的投资组合每个月的样本外表现都进行了评估, 从而得到了训练数据集. 利

① CRSP 数据库是指芝加哥大学布斯商学院的 Center for Research in Security Prices 数据库, 该数据库可以通过宾夕法尼亚大学的 Wharton Research Data Services 获得 (网址为 `http://wrds.wharton.upenn.edu`).

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

用 Ledoit 和 Wolf (2004) 的方法, LXC 在 1995 年 1 月至 2009 年 12 月这个时间段的每个月初, 从在前面 120 个月没有缺失数据的股票中选择市值最大的 $m = 50$ 只股票, 在 t 期的投资组合由这 m 只股票组成.

注意到在这个时间段中包含一些股市波动率很大的时期, 如 1987 年的“黑色星期一”、互联网泡沫破灭、2001 年 911 恐怖袭击和 2007 年开始的次贷危机. 为了减小这些时期的影响, LXC 选择使用 120 期的训练数据来构建下一期的投资组合. 由于 LXC 论文的实证研究的目的是阐述论文提出的方法的基本统计性质, 并与其他利用 Markowitz 均值–方差投资组合优化原理的方法相比较. 因此与 2.4.2 节提到的 Black-Litterman 方法不同, LXC 论文中投资组合的构建只与这些数据有关, 并没有使用股票对应的公司信息等其他信息. 另外, 为了进行公平的比较, LXC 在贝叶斯方法中并没有对于先验分布作任何假定, 而是只使用了 NPEB 方法.

投资组合的业绩用相对于基准投资组合的超额收益率 e_t 来衡量. 在 1995 年 1 月至 2009 年 12 月的时间段中, 随着时间 t 的变化, LXC(1) 将直到 t 期的已实现超额收益率相加, 得到累积超额收益率 $\sum_{l=1}^{t} e_l$; (2) 使用平均已实现超额收益率的均值和标准差计算得到已实现的信息比率为 $\sqrt{12}\bar{e}/s_e$, 其中 $\bar{e}$ 为月超额收益率的样本均值, s_e 为相应的标准差. 类似于 Ledoit 和 Wolf (2004), 此处乘上 $\sqrt{12}$ 以将信息比率进行年化. 注意到已实现的信息比率是 180 个周期的月超额收益率的统计量, LXC 将该投资业绩测度与累积已实现的超额收益率的时间序列图相结合, 而 e_t 可以通过将累积已实现的超额收益率进行差分化获得.

LXC 将 S&P 500 指数作为基准投资组合 u_t, 从而得到超额收益率 e_{it}, 并使用模型 (2.19)–(2.20) 对于训练数据集进行拟合, 从而对于 e_{it} 的均值和方差进行预测. 在此处允许进行卖空, 对于任意 i, 约束条件为 $w_i \geqslant -0.05$. 由于 $\sum_{i=1}^{m} w_i = 1$, 则有

$$\sum_{i=1}^{m} w_i r_{it} - u_t = \sum_{i=1}^{m} w_i (r_{it} - u_t) = \sum_{i=1}^{m} w_i e_{it}, \tag{2.23}$$

在积极的资产组合优化中, 研究的是满足 $\sum_{i=1}^{m} \widetilde{w}_i = 0$ 情况下的权重向量 $\widetilde{w}_i = w_i - w_{i,B}$. 在 (2.23) 中, 目标函数是在一定的超额收益方差的条件下, 最大化超额收益率的均值 (这与获得一定的信息比率相关), 因此我们可以在投资组合优化问题 (2.1) 或 (2.9) 中, 将收益率 r_{it} 替代为超额收益率 e_{it}, 并使用比 r_{it} 更简单的平稳时间序列过程来对 e_{it} 进行建模. 在模型 (2.19)–(2.20) 中, 当满足 $e_{i,t} = (1, e_{i,t-1}, u_{t-1})\boldsymbol{\beta}_i + \epsilon_{i,t}$ 时, 可以得到投资组合 NPEB_{SRG} (其中 SRG

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

代表误差项服从 GARCH(1, 1) 分布的随机回归模型). 除此之外, LXC 也使用了 AR(1) 模型 $e_{it} = \alpha_i + \gamma_i e_{i,t-1} + \epsilon_{it}$ 来与嵌入式投资组合、协方差压缩的投资组合和 Bootstrap 重抽样投资组合进行比较, 在这些模型中都假定了独立同分布模型, 并使用了年化收益率 $\mu_* = 0.015$.

对于 NPEB_{AR} 和 NPEB_{SRG} 这两个投资组合, LXC 基于 2.7.2 节中的训练样本选择使得信息比率最大化的 $\lambda \in \{2^i : i = -3, -2, \ldots, 6\}$. LXC 论文的图 4 比较了在 180 个月的周期内 NPEB_{SRG}、NPEB_{AR}、2.4 节第一段提到的嵌入式投资组合、Ledoit-Wolf 投资组合 (2.6) 和 Michaud 的 Bootstrap 投资组合 (2.8) 这几个投资组合相对于 S&P 500 的已实现超额收益率, 在计算时目标收益率设置为 $\mu_* = 0.015$. 由图可知, NPEB_{SRG} 显著地优于 NPEB_{AR}, 而后者远远优于其他三个相互接近的投资组合. 表 2.2 给出了在 S&P 500 作为基准投资组合时, 上述五个投资组合的信息比率和平均已实现超额收益率. 除了 $\mu_* = 0.015$ 的情形外, 表中同时列出了 $\mu_* = 0.02, 0.03$ 的情形. 在每种情形下, 分别计算了嵌入式投资组合 (P)、协方差压缩投资组合 (S) 和 Bootstrap 重抽样投资组合 (B) 在 180 个周期内给定权重向量约束时的相关变量.

表 2.2 S&P 500 作为基准投资组合时, 五种投资组合的信息比率和平均已实现超额收益率

$\mu_* = 0.015$			$\mu_* = 0.02$			$\mu_* = 0.03$			NPEB	
P	S	B	P	S	B	P	S	B	AR	SRG
0.527	0.352	0.618	0.532	0.353	0.629	0.538	0.354	0.625	0.370	1.169
[0.078]	[0.052]	[0.077]	[0.078]	[0.051]	[0.078]	[0.076]	[0.050]	[0.077]	[0.283]	[0.915]

2.8 有效市场假说之外的统计套利和策略

1.3 节和 1.7 节对于有效市场假说的实证检验和相关争论进行了回顾. 在本节中, 我们先将有效市场假说放在一边, 而是讨论一系列的量化交易策略, 这些策略通过寻找统计套利机会 (Statistical Arbitrage Opportunities, SAOs) 并进行资本化, 来获得超过一定水平的期望收益. 在本节, 我们详细介绍了在 1.3 节中简要提到的策略和统计套利机会, 并讨论了相关的统计学和经济学原理.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

2.8.1 非参数回归和变点模型中的技术准则和统计背景

技术准则基于如下前提: 过去的市场价格中包含收益的特征和趋势信息, 而这些特征和趋势在未来还会重现. 人们使用 “技术分析” 来从历史数据中识别相关的特征, 从而建立相应的买卖证券的策略. 由于技术分析通常包含数据可视化 “图表”, 因此技术分析通常被称为 “图表分析”. 有效市场假说的支持者一直对技术分析持负面意见, 比如说 Malkiel (2003) 认为 “很显然, 我对于图表分析者持负面看法 …… 技术分析是对于学术世界的诅咒 …… 技术分析的方法明显是错误的.” 他对于技术分析的强烈反对很可能是因为技术分析是严格的数据驱动的分析, 并不具备经济学意义. 另一方面, Brock 等 (1992) 的 1732 页中指出, 尽管技术分析受到了 Malkiel 等学者的负面评价, “但它却在华尔街享受着复兴”, 并且 “所有的大型券商都在基于技术分析 …… 对于市场和单个证券发表技术评论”. Brock 等 (1992)、Lo 等 (2000) 以及 Lai 和 Xing (2008) 一书的 11.1.1 节对于非参数回归和变点检测模型进行了一定的介绍, 这两个模型为很多统计准则提供了统计背景. 在此处, 我们对于这两个模型进行更加详细的介绍, 并介绍了更多的参考资料.

筛选规则和等价的 CUSUM 变点检测规则 筛选规则是一个非常流行的交易规则, 在该规则中包含筛选时间 t 的跌幅 $D_t = \max\limits_{0\leqslant k\leqslant t} P_k - P_t$ 和涨幅 $U_t = P_t - \min\limits_{0\leqslant k\leqslant t} P_k$. 当满足 $D_t/(\max\limits_{0\leqslant k\leqslant t} P_k) \geqslant c$ 时 (表明相对于现行的最大价格有一个 $100c\%$ 的跌幅), 筛选规则发出买入的信号, 而相应地当满足 $U_t/(\min\limits_{0\leqslant k\leqslant t} P_k) \geqslant c$ 时, 发出卖出的信号. 定义 $r_k = \log(P_k/P_{k-1})$ 和 $S_t = \log P_0 + \sum_{k=1}^{t} r_k$, 注意到

$$U_t \Big/ \left(\min_{0\leqslant k\leqslant t} P_k\right) \geqslant c \Longleftrightarrow S_t - \min_{0\leqslant k\leqslant t} S_k \geqslant \log(1+c). \tag{2.24}$$

因此卖出信号对应于质量控制中的 CUSUM 规则: 当满足 $S_t - \min\limits_{0\leqslant k\leqslant t} S_k \geqslant b$ 时, 立即进行改进, 具体可参考 Lai 和 Xing (2008) 一书的 279 页以及 Page (1954). Alexander (1961) 以及 Fama 和 Blume (1966) 通过实证研究表明当考虑交易成本时, 筛选规则并不比简单的买入持有策略的收益更高, 其中买入持有策略是指买入股票并一直持有. Lam 和 Yam (1997) 分析了更加一般化的 CUSUM 规则, 并将它们转化成了包含涨幅和跌幅的筛选规则的形式, 并通过实证研究表明这些策略相比于传统的筛选策略而言有一定的改进, 具体可参见 2.9 节.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

移动平均规则和有限窗宽下的 GLR 故障检测方案 Lai 和 Xing (2008) 一书的 277 页、Brock 等 (1992) 和 Sullivan 等 (1999) 对于技术分析中常用的移动平均规则进行了介绍, 这些规则非常简洁且符合直觉. 除了固定时长的移动平均规则外, Brock 等 (1992) 对于时长可变的移动平均规则也进行了介绍, 这些规则与 Willsky 和 Jones (1976) 以及 Lai (1995) 提出的有限窗宽下的故障检测方案密切相关, 具体可参见 2.9 节.

基于神经网络和广义线性模型的方向性投资 Gencay (1998) 介绍了一系列的方向性投资策略. 在这些策略中, 周期为 n 的投资的累积对数收益率为 $S_n = \sum_{t=1}^{n} \hat{y}_t r_t$, 其中 $\hat{y}_t = 1$(或 -1) 代表在 t 时期对股票/证券/货币做多 (做空). $\hat{y}_t$ 是基于直到 $t-1$ 期的伯努利变量 $y_t = 2 \times (I_{\{r_t>0\}} - 1)$ 进行估计. Gencay (1998) 使用了单层神经网络模型和交叉验证方法来确定神经网络中的隐单元的数量. 对于神经网络模型和交叉验证方法的介绍可以参考 Lai 和 Xing (2008) 的 7.3.2 节和 7.4.4 节, 该书对于交叉验证进行了介绍, 并将神经网络作为一种非参数回归方法进行了介绍. Gencay (1998) 通过实证研究说明使用论文中的方法的收益率远高于简单的买入持有策略的收益率. 注意到论文中在建立和评估交易策略时, 交易成本和预测偏误 $y_t - \hat{y}_t$ 被忽略. 当考虑交易成本和预测偏误时, 可以得到如下改进的方向性交易策略:

$$\hat{y}_t = \begin{cases} 1, & \text{当 } \hat{\pi}_t^1 \geqslant q \text{ 时}, \\ -1, & \text{当 } \hat{\pi}_t^2 \geqslant q \text{ 时}, \\ 0, & \text{其他}, \end{cases}$$

其中 $\hat{\pi}_t^1$ 或 $\hat{\pi}_t^2$ 分别是在 $r_t \geqslant c_1$ 或 $r_t \leqslant c_2$ 两种情况下, 给定 $t-1$ 时间的信息集 $\mathcal{F}_{t-1}$ 和 $2q < 1$ 时的条件概率的估计量. 条件概率可以使用 Logistic 回归来进行估计. 见 Lai 和 Xing (2008) 的 4.1 节.

2.8.2 时间序列、动量策略和配对交易策略

Lai 和 Xing (2008) 一书的 278 页指出, 时间序列建模和预测在技术准则的发展中有着直接或间接的应用: "(非参数) 回归模型主要侧重于对于数据的特征进行建模, 其基本思想是过去的特征也适用于未来, 而通过使用合适的模型可以对于未来的特征进行更加精确的估计." 书中以动量策略和配对交易为例, 说明时间序列框架在技术分析和研究技术准则中的重要作用, 具体可参见书中第 11.1.3 节. 在本节中, 我们接下来会对于 Lai 和 Xing (2008) 没有提到的截面动量策略和时间序列动量策略进行阐述.

截面动量策略认为过去的业绩和未来的业绩之间有连续性, 因此通过买入

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

在过去相对表现较好的投资组合、卖出在过去表现较差的投资组合可以获得较高的收益. Jegadeesh 和 Titman (1993) 通过实证研究说明在过去 3–12 个月表现较好的公司在接下来的时间内的表现依然很好. 在大概 10 年之后, Lewellen (2002) 对于动量问题从时间序列角度进行了研究, 研究发现 “市场中的价格信息甚至不是弱有效的”. 论文使用 CRSP 数据库, 分析了 NYSE、AMEX 和 NASDAQ 的股票回报率的自回归性和市场投资组合的自回归性. Moskowitz 等 (2012) 引入了时间序列动量来作为动量策略的补充, 并通过实证研究说明时间序列动量策略 (基于证券在过去的绝对业绩, 而不是基于相对业绩) 的表现比截面动量策略更好. 他们发现基于时间序列动量而建立的分散化投资组合在市场出现极端情况下表现是最优的. 另外, 通过分析投机者和对冲者的交易行为, 他们发现基于时间序列动量策略进行投资的投机者从对冲者处获得了收益.

2.8.3 逆向投资策略、行为金融及投资者的认识偏差

逆向投资者通过对于市场的投资者情绪进行监测来发现市场中投资者的认识偏差. David Dreman 是逆向投资研究领域的领军者, 他以出版逆向投资方面的书籍和成立逆向投资公司 Dreman Value Management 而闻名. Dreman (1979、2012) 认为从心理学角度解释的投资决策在股票价格中起到重要作用, 投资者的心理状况常常会使得股票价格偏离股票的内在价值. 心理情况, 特别是从众心理, 是人的天性的一部分. 另一个投资策略中的缺陷在于很难从公司的年报、新闻或评论中对于公司未来价值进行合理估计. 虽然股票的价格最终会回归到内在价值附近, 但由于估值错误或心理偏差而导致的对内在价值的错误估计, 投资者仍然会因为在一定的时期内错误决策而产生损失. 逆向投资策略则是逆市场而行, 通过买入市场并不看好的股票、卖出市场看好的股票, 然后等待市场发现之前对于股票定价的错误, 使得估价过高的股票价格回落, 而估价过低的股票价格上升, 从而获取差价.

因此, 逆向投资策略与行为金融理论密切相关. 行为金融是近年来较新的金融和经济学领域, 该领域将心理学理论和行为科学引入到金融建模、预测和决策、经济学分析和政策研究中. 正如 1.3 节指出, 相对于认为价格应该反映其内在价值的有效市场假说而言, 行为金融提供了另一种对市场的新的观点. 实际上, 行为金融理论使得对于股票或其他证券的风险建模发生了根本性的改变. 行为金融理论将 Kahneman 和 Tversky (1979) 提出的前景理论引入到不确定性决策中. 进一步的细节和讨论参见 2.9 节.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

2.8.4 从价值投资到全球宏观策略

公司的 “内含价值” 在价值投资中起主导作用. 内含价值的理论要追溯到 Graham 和 Dodd (1934) 的经典教材[①], Lai 和 Xing (2008) 一书的 11.1.5 节中对于内含价值理论进行了讨论, 在这一节中也对于内含价值和逆向投资、从众行为的联系进行了讨论. 对于一个公司或者其业务部门业绩的基本分析在分析公司的内含价值方面起重要的作用. 特别是在内含价值分析涉及对未来的现金流进行预测时, 如进行宏观经济趋势预测时, 作用尤其明显.

对于全球宏观发展情况的分析和预测在宏观全球策略中起主导作用. 该策略基于对于政府政策、政府间关系等地区政治经济学的分析和预测, 在全球范围内进行投资. 对冲基金公司在建立系统性全球宏观策略时, 其分析模块主要包括对于利率趋势、经济周期、全球基金流网络、全球经济不平衡特征和新兴经济中不断变化的增长模型等进行监测. 正如 1.4 节中指出, 对冲基金公司的策略可以被广义地归类到 “方向性策略” 中, 策略的制定主要依赖于基金经理的技能和经验; 也可以被归类为 “系统性策略”, 策略的制定也依赖于对数据的数量分析和对于交易头寸的分析的计算机模型和算法. 全球宏观策略在商品和期货市场中得到广泛应用. 建立传奇性的索罗斯基金公司的乔治 · 索罗斯在他的全球宏观基金中获得了巨大的收益, 最为人所知的就是他在 1992 年预测到英国政府在欧洲汇率体系崩溃时英镑将贬值, 他通过做空英镑获得了超过 10 亿美元的收益.

2.8.5 投资策略的样本内评估和样本外评估以及多重检验方法

Lai 和 Xing (2008) 一书的 11.1.4 节讨论了对于交易策略的赢利能力评估中的数据挖掘问题, 并对于 White (2000) 和 Hansen (2005) 中的 Bootstrap 方法进行了回顾, 同时对于 Romano 和 Wolf (2005) 中利用步降式 (Step-down) Bootstrap 测试法来控制多元测试中的总体错误率的方法进行了分析. 这些方法可以用于对交易策略进行有效的样本内测试. 当然, 样本外测试 (后验测试) 的方法更具有决定性, 但在测试时该策略已经被使用, 且已经产生了损失, 因此样本外评估的方法具有一定的滞后性. 在使用样本内测试方法来替代样本外测试方法时, 可以使用 k 层交叉验证的方法, 具体参见 Hastie 等 (2003) 以及 Lai 和 Tsang (2016). 尽管前面提到的 Bootstrap 方法可以控制第一类错误率 (FWER), 且可以避免在对于交易策略的收益性进行实证检验时出现数据探测

① Graham 和 Dodd 都是哥伦比亚大学商学院的教授, 他们最著名的学生是沃伦 · 巴菲特, 巴菲特在接管 Berkshire Hathaway 公司后在基金投资方面取得了重大成功.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

法陷阱, 但对于 Bootstrap 方法在多重检验中的势的研究, 目前并没有系统的仿真研究或理论研究. Lai 和 Tsang (2016) 弥补了这方面的空白, 论文针对有效多重检验问题, 提出了一个综合的理论模型和方法.

正如 Henriksson 和 Merton (1981) 指出, 如何对于投资经理或投资策略的表现进行评估一直是金融中的重要问题. 针对该问题, 该论文提出了一个统计模型来检验做市商的预测能力. Pesaran 和 Timmermann (1992) 提出了一个非参数检验方法来对经济变量的变化方向的预测能力进行评估, 并将该方法应用到对于英国制造业的两项实证研究中. 需要指出的是, 上述研究假设每一对预测值和实际值都是独立同分布的. 但正如我们在 2.6.3 节看到的那样, 时间序列效应存在于资产回报率中, 因此这些统计量应该考虑时间序列效应. 另外, 一些更加复杂的估计方法自身就包含参数不确定性. 举例来说, 一些方法在分析价格是否会上涨时, 会引入上涨概率这个变量, 而在 2.8.1 节末尾中提到方向性投资策略等策略中, 并不直接包含上涨概率这个变量, 因此带来了新的不确定性. 而当时间序列效应不能忽略时, 对于上涨概率的估计更加困难, 但这些问题在 Lai, Gross 和 Shen (2011a) 中通过利用预测模型中的鞅特性得到了解决.

2.9 补充材料和习题

1. **稳定 (stable) $(\alpha,\beta,\gamma,\delta)$ 分布** 该分布可以由如下的特征方程来表示:

$$E\{\exp(iuY)\}=\begin{cases}\exp\left\{-\gamma^{\alpha}|u|^{\alpha}\left[1-i\beta\left(\tan\dfrac{\pi\alpha}{2}\right)\operatorname{sign}(u)\right]+i\delta u\right\}, & \alpha\neq 1,\\ \exp\left\{-\gamma|u|\left[1+i\beta\dfrac{2}{\pi}\operatorname{sign}(u)\log|u|\right]+i\delta u\right\}, & \alpha=1,\end{cases}$$

其中 $\gamma>0, i=\sqrt{-1}$. Famma 和 Roll (1971) 提出的用于对称稳定分布的近似分位数方法是比较早的对稳定分布的参数估计研究的论文. 当 Y 对称 (即 Y 与 $-Y$ 具有相同的分布) 时, $E[\exp(iuY)]$ 为实数. McCulloch (1986) 随后提出了一个一般的分位数方法, 该方法可以利用 R 语言程序包 `StableEstim` 中的 `McCullochParametersEstim` 命令实现. 该程序包同时包含 `MLParametersEstim` 和 `GMMParametersEstim` 命令, 分别对应于最大似然估计方法和广义矩估计方法. `StableEstim` 中的另一个估计方法是 `KoutParametersEstim`, 该方法使用迭代 Koutrouvelis 回归方法, 且回归时使用不同的格点, 实证特征方程即基于这些格点进行计算. 具体而言, 我们使用 `McCullochParametersEstim` 来对于图 2.4 的下图中

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

Pfizer 公司的 2517 天的收盘价的差分值 Δ_t 来拟合对称稳定分布. 参数估计值为 $\hat{\alpha} = 1.652, \hat{\gamma} = 0.171, \hat{\beta} = \hat{\delta} = 0$. 图 2.6 对于 Δ_t 的拟合对称分布的 QQ 值与拟合正态分布的 QQ 值进行了比较. 由图可知, 数据相比于正态分布更厚尾, 而比稳定分布模型尾部更薄.

2. **投资组合风险调整后的表现** 在 1.4 节中, 我们引入了信息比率、Sortino 比率和 Modigliani 的风险调整后表现 (risk-adjusted performance, RAP) 等指标来对于投资组合的风险调整后收益率进行衡量. 这些指标通常用于对投资基金进行排序, 在计算这些指标时, 只涉及收益率一阶矩和二阶矩, 或者说只涉及收益率的均值和方差, 因此这些指标适用于收益率服从高斯分布的情况. 而当收益率并不服从高斯分布时, Gatfaoui (2010) 利用数值模拟, 阐述了样本与正态分布的偏差对于这些指标本身的计算以及在利用 Sharpe 比率对于收益率进行排序时的影响. 在收益率独立同分布时, 利用 Bootstrap 方法可以得到对于这些指标的无偏差估计, 具体参见 Lai 和 Xing (2008) 一书 3.5.2 节.

   在 Mandelbrot (1963) 之后, Rachev 和 Mittnik (2000) 以及 Martin 等 (2005) 提出, 应该使用 2.6.1 节中的均值有限的 (即满足 $1 < \alpha < 2$ 情况下) 稳定分布来替代高斯分布. 而在 σ 无穷大的情况下, 这两篇论文提出使用超额收益率 $r - r_f$ 在大于特定的分数值情况下的条件均值来替代收益率的均值. 这个条件均值也被称为期望损失 (Expected Shortfall), 具体可参见 Lai 和 Xing(2008) 一书 12.1.3 节. 注意到在实际估计时, 通常会使用样本均值 $\bar{r} - r_f$ 来替代 $E(r - r_f)$, 使用样本标准差 s 来替代 σ (σ 可能无穷大). 研究表明, 当收益率服从稳定分布时, 自正则变量 $(\bar{r} - r_f)/s$ 在原点附近的矩母函数是有限的, 具体参见 de la Peña 等 (2009). 尽管通常使用的 Sharpe 比率 $E(r - r_f)/\sigma$ 在 $\sigma = \infty$ 时, 如 Rachev 等人研究的稳定分布中, 取值为 0, 没有直观含义. 但是此时重要的是未来投资区间的样本 Sharpe 比率, 根据前面提到的自正则和的特性, 其均值不为 0, 且此时 Bootstrap 方法依然适用.

3. **Black-Litterman 资产配置模型** Black 和 Litterman (1991) 首先假设时间 t 的资产收益率 $\boldsymbol{r}_t$ 服从均值为 $\boldsymbol{\mu}$ 的正态分布, 即 $\boldsymbol{r}_t|(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. 为简化问题, 论文假设 $\boldsymbol{\Sigma}$ 已知 (且使用历史数据的样本协方差矩阵来替代 $\boldsymbol{\Sigma}$). 假设先验分布为 $\boldsymbol{\mu} \sim N(\boldsymbol{\pi}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}})$, 等价地

$$\boldsymbol{\mu} = \boldsymbol{\pi} + \boldsymbol{\epsilon}^{(e)}, \quad \boldsymbol{\epsilon}^{(e)} \sim N(\mathbf{0}, \tau\boldsymbol{\Sigma}), \tag{2.25}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 $\boldsymbol{\pi}$ 对应于风险资产的 “均衡风险溢价”. 与贝叶斯方法中 “引出” 先验分布的方法类似, Black-Litterman 方法使用 “反向优化” 方法来得到先验均值, 先验均值服从如下形式:

$$\boldsymbol{\pi} = \gamma \boldsymbol{\Sigma} \boldsymbol{w}^*, \tag{2.26}$$

其中 γ 是市场的风险厌恶系数, 而 $\boldsymbol{w}^*$ 则是指市场资本化下各资产的权重. 为进一步简化, 论文假设 $\boldsymbol{\Sigma}_{\boldsymbol{\mu}}$ 与协方差矩阵 $\boldsymbol{\Sigma}$ 成正比, 即

$$\boldsymbol{\Sigma}_{\boldsymbol{\mu}} = \tau \boldsymbol{\Sigma}, \tag{2.27}$$

其中 τ 表示 $\boldsymbol{\mu}$ 的先验分布的不确定性, 对于该变量的取值, 不同的学者持有不同意见. 一个常用的经验取值是 $\tau = \dfrac{1}{T}$, 其中 T 是样本时长. 如在分析五年期的月数据时, 设置 $\tau = 0.017$. 与 2.4.2 节第一段中考虑给定历史资产收益率 $\mathcal{F}_t$, $1 \leqslant t \leqslant n$ 下, $\boldsymbol{\mu}$ 的后验分布不同, Black-Litterman 在研究时, 假设投资者对于市场信息有着自己的观点, 即假设

$$\boldsymbol{P}\boldsymbol{\mu} = \boldsymbol{q} + \boldsymbol{\epsilon}^{(v)}, \quad \boldsymbol{\epsilon}^{(v)} \sim N(\boldsymbol{0}, \boldsymbol{\Omega}), \tag{2.28}$$

其中 $\boldsymbol{P}$ 是一个 $p \times m$ 矩阵, 表示投资者选取的资产, $\boldsymbol{q}$ 则表示投资者的观点, $\boldsymbol{\Omega}$ 是一个对角协方差矩阵, 表示投资者观点的不确定性. 假定 $\boldsymbol{\epsilon}^{(e)}$ 和 $\boldsymbol{\epsilon}^{(v)}$ 相互独立, 作者将投资者的观点 (2.28) 与先验分布 (2.25)–(2.27) 相结合, 从而得到了在给定投资者的观点的情况下 $\boldsymbol{\mu}$ 的后验分布. 这也是 Black-Litterman 方法的本质所在.

习题 使用贝叶斯定理证明, $\boldsymbol{\mu}$ 的后验分布服从正态分布, 其均值为 (2.7)、协方差矩阵则为 (2.7) 下面一行中的矩阵.

4. 当投资者将其专业知识和统计建模相结合时, 2.7.3 节中的新现代投资组合方法表现会更好. 事实上, 如果仅仅是通过对过去的信息进行实证分析来得到 (2.19) 式的预测模型, 模型的预测能力大大下降, 并大大降低了 2.5 节中提到的模型的通用性. 该模型在广义贝叶斯框架下建立, 投资者可以利用该模型来预测未来收益率 $\boldsymbol{r}_{n+1}$, 在进行预测时, 先验分布可以既包含投资者对于未来股票价格的观点, 也包含市场对于未来股票价格的观点, 其中市场的 “观点” 可以通过市场均衡投资组合来反映, 该方法与 2.4.2 节中提到的 Black-Litterman 方法类似. 注意到 Black-Litterman 方法假设这些 “观点” 的先验分布都是正态分布, 分布的方差项反映这些

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

观点自身的不确定性. 而我们提出的贝叶斯方法在此方法的基础上更进一步, 在求解优化问题 (2.9) 时加入投资组合收益率的实际均值和方差, 考虑这些不确定性的因素, 而不是直接在嵌入式方法中使用已估计的均值和方差.

5. **多目标优化和有效前沿** 定义 $\boldsymbol{f}:\mathbb{R}^n\to\mathbb{R}^m$ 为 n 个决策变量的向量价值函数, $f_i(\boldsymbol{x})$ 为第 i 个目标函数. 决策变量可能会受到不等式或等式的约束, 因此优化问题可以看作在 $\boldsymbol{x}\in S$ 上的最大化问题, 其中 S 为约束集. 当满足 $f_i(\boldsymbol{x})\geqslant f_i(\boldsymbol{x}')$, $i=1,\ldots,m$, 且至少有一个 i 取严格不等式时, 即称 $\boldsymbol{x}\in S$ 相对于 $\boldsymbol{x}'\in S$ 占优. 当集合中不存在任何 $\boldsymbol{x}\in S$ 相对现有的决策变量占优时, 称该决策变量为帕累托最优. 包含所有 S 的帕累托最优元素的集合称为帕累托边界. 若 $\boldsymbol{x}$ 是一个随机变量, 则 $f_i(\boldsymbol{x})$ 为 $\boldsymbol{x}$ 的期望函数. 特别地, Markowitz 均值–方差投资组合优化问题是一个二元随机优化问题, 满足 $f_1(\boldsymbol{w})=E(\boldsymbol{w}^\top\boldsymbol{r})$, $f_2(\boldsymbol{w})=-\mathrm{Var}(\boldsymbol{w}^\top\boldsymbol{r})$. 在做优化时, 人们希望投资组合的均值增大, 而方差减少, 因此帕累托边界对应于有效前沿. 在 Markowitz 有效边界上的投资组合可以被看作在给定的收益率下方差最小的投资组合, 或者说在给定的波动性 (即收益率的标准差) 下, 收益率均值最大的投资组合. 由于投资者往往更在意收益率, 因此投资经理往往更倾向于使用前一种表达方式, 但这会给嵌入式方法带来一定的问题, 具体问题会在 3.1.2 节中讨论.

6. **Fama-French 三因子模型和五因子模型** Fama 和 French (1993) 提出了资产定价三因子模型, 其中三个因子分别指 (a) CAPM 中市场投资组合 M 上的超额收益率 r_M-r_f; (b) 市值小的公司组成的投资组合 S 的收益率与市值大的公司组成的投资组合 L 的收益率之差 r_S-r_L; (c) 账面价值高的公司组成的投资组合 h 与账面价值低的公司组成的投资组合 l 的收益率之差 r_h-r_l. Fama 和 French (2015) 在三因子模型的基础上添加了两个新的因子, 构建了五因子模型. 新添加的因子为: (d) 盈利能力高的投资组合的收益率与盈利能力低的投资组合的收益率之差; (e) 投资比例高的公司组成的投资组合的收益率与投资比例低的公司组成的投资组合的收益率之差. Kenneth French 在 Dartmouth 大学商学院的网站上指出, 这些因子模型被用于开发和改进投资策略, 从而为 “金融顾问和机构客户提供投资建议”. 该网站同时给出了一个数据库, 包含这些因子和 CRSP 数据库中的一些相关的股票的月收益率.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

7. **统计模型的相关评述** 在 2.6.3 节中, 我们在介绍证券收益率的时间序列模型时提到的 George Box 对统计模型和模型的构建提出了如下看法: "本质上来说, 所有模型都是错误的, 但有些模型具有实用价值." 这句话不仅与本节的内容相关, 也适用于后续章节提出的模型. 这一观点在 Box 和 Draper (1987) 一书第 424 页中提出, 它与 Box (1976) 之后的一系列其他类似的观点一样, 都阐述了 Box 的观点: 简化的模型虽然并不完全正确, 但它们能够给出有效的近似. 基于 Box 的理论和 Cox (1990)、Lehmann (1990b) 等其他后续的理论, Lai 和 Wong (2006) 提出了一个时间序列建模的 "专业–实证" 模型, 该模型结合领域知识 (Cox 称之为专业知识) 和实证时间序列模型 (2.6.3 节中的模型). 联系到 5.3 节中的限价委托模型, Breiman (2001) 引入了*算法模型*以适用于最近兴起的统计分析大数据建模.

8. **因子模型和高维协方差矩阵** 与 APT 相关联的多因子模型 (2.4) 实际上包含一个存在多只股票的经济. Chamberlain 和 Rothschild (1983) 指出, Ross (1976) 只对于 APT 给出了一个启发式的阐述: 在一个无套利且存在多个资产的市场中, "资产价格可以近似表达为因子的线性函数". 论文利用均值–方差内积的 Hilbert 空间定义了: (a) 各资产及其投资组合的价格; (b) 无套利的条件. 在此框架下, 论文定义了一个严格 r 因子结构方程来刻画前 N 个资产的收益率: $\boldsymbol{\Sigma}_N = \boldsymbol{B}_N \boldsymbol{B}_N^{\top} + \boldsymbol{D}_N$, 其中 $\boldsymbol{\Sigma}_N$ 为收益的协方差矩阵, $\boldsymbol{B}_N$ 是一个 $N \times r$ 矩阵, $\boldsymbol{D}_N$ 是一个对角矩阵, 对角元素有着统一的界限. 论文同时定义了一个近似 r 因子结构方程, 此时 $\boldsymbol{D}_N$ 被一个非负定矩阵 $\boldsymbol{R}_N$ 取代, 且满足 $\sup_N \lambda_{\max}(\boldsymbol{R}_N) < \infty$. 论文证明, 近似 r 因子方程是 Ross 理论的充分估计. 更重要的是, 论文指出, 当 $\boldsymbol{\Sigma}_N$ 只包含 r 个有界的特征值时, 近似 r 因子结构存在且唯一, 且对应的 r 个单位特征向量收敛, 此时即为因子载荷. Lai 和 Xing (2008) 的 2.2 节和 3.4.3 节对于主成分分析 (PCA) 方法和它在估计 r 和近似 r 因子模型中的因子的应用进行了阐述. 该书 3.4.1 节则介绍了因子分析, 和其在估计 r 和 r 因子模型中的因子的应用进行了阐述. 在经典的独立同分布收益率假设下, Chamberlain 和 Rothschild 提出的因子模型在协方法矩阵 $\boldsymbol{\Sigma}_N$ 的维度缩减和估计中起到重要作用.

   与大数据中的统计/机器学习中的高维协方差矩阵估计和随机矩阵理论的发展相关, 目前在包含大量资产时的资产收益因子建模方面也有

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

很多进展. 随机矩阵理论 (RMT) 的基础是 $p \times p$ 实对称或者复 Hermite 矩阵的特征值的近似分布. 特别地, Wigner (1995) 使用了对称 $p \times p$ 矩阵的实证测度来对重原子建模, 此时矩阵的上对角部分都是独立同分布, 且与独立同分布的对角项独立, 且上对角项为零均值随机变量. 论文同时证明了 $p \to \infty$ 时实证测度向半圆分布收敛的特性. 在 1967 年, Marčenko 和 Pastur (1967) 发表了 RMT 的奠基性论文. 该论文将随机 Hermite 矩阵拓展到随机自轭线性因子中, 并引入了一个包含 Stieltjes 变换的方法, 从而求出特征值的极限分布. Bai (1999), Bai 和 Silverstein (1998) 以及其他论文对于样本协方差矩阵的特例进行了进一步研究. El Karoui (2008) 以及 Ledoit 和 Wolf (2012) 利用这些结果, 以及 Marčenko-Pastur 方程, 求得大样本协方差矩阵的估计值, 并得出其近似最优特征. Marčenko-Pastur 方程将实证谱分布 (即对于 $p \times p$ 矩阵的每一个协方差矩阵赋予同样权重的概率分布) 与 $t \to \infty$ 时的极限谱分布的积分相联系. 而在因子模型方面, Johnstone 和 Lu (2009) 以及 Onatski (2010, 2012) 提出了“弱影响因子” 的近似分布.

Johnstone 和 Lu 假设如下因子模型中的 ε_t 独立同分布, 均值为 0:

$$\boldsymbol{x}_t = \boldsymbol{C}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T, \tag{2.29}$$

其中 $\boldsymbol{f}_t = (f_{t1}, \ldots, f_{tr})^\top$ 是因子向量. Onatski 的模型中允许 $\boldsymbol{\varepsilon}_t$ 中存在截面短期相关性如下:

$$\boldsymbol{\varepsilon}_t = \boldsymbol{A}(b_1(L)u_{1t}, \ldots, b_p(L)u_{pt})^\top, \tag{2.30}$$

其中 u_{ij} 独立同分布, 均值为 0, 方差为 σ^2. $\boldsymbol{A}$ 是一个 $p \times p$ 相关系数矩阵, 表示截面相关性, $b_i(L)$ 是滞后因子 L 的多项式, 满足 $b_i(L)u_{it} = u_{it} + b_{i,1}u_{i,t-1} + \cdots + b_{i,\nu_i}u_{i,t-\nu_i}$. 定义 $\boldsymbol{F} = (\boldsymbol{f}_1, \ldots, f_T)$, 并将因子及其载荷标准化, 从而使得 $T^{-1}\boldsymbol{F}^\top\boldsymbol{F} = \boldsymbol{I}$ 且 $\boldsymbol{C}^\top\boldsymbol{C}$ 是对角矩阵, 则弱影响因子的假设为:

$$\boldsymbol{C}^\top\boldsymbol{C} \to \operatorname{diag}(d_1, \ldots, d_r), \quad \text{其中 } \min_{1 \leqslant i \leqslant r} d_i \geqslant 0, \quad T \to \infty, \frac{p}{T} \to \rho, \tag{2.31}$$

其中 ρ 为特定的常数. Johnstone 和 Lu (2009) 以及 Onatski (2010) 指出, 当 $\rho > 0$ 时, PCA 方法得到的因子并不是一致的. 他们同时提出了

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

PCA 估计量的修正方法, 从而得到 "可识别" 因子和特征向量的一致估计量.

与 (2.31) 中对弱影响因子的假设相反, Fan 等 (2013) 假设 "广泛条件":

$$\lim_{p\to\infty} \inf p^{-1}\lambda_{\min}(\boldsymbol{C}^\top \boldsymbol{C}) > 0 \tag{2.32}$$

且假设 "近似稀疏条件", 即对于特定的 $0 \leqslant \delta \leqslant 1$,

$$\max_{i\leqslant p} \sum_{j=1}^{p} |E(\varepsilon_i \varepsilon_j)|^\delta = O(1). \tag{2.33}$$

在此情况下, 存在 $1 \leqslant \nu \leqslant r$, 使得向量 $\boldsymbol{x}_t$ 的 $p \times p$ 阶协方差矩阵 $\boldsymbol{\Sigma}$ 的最大的 ν 个特征向量是 p 阶的. Fan 等, (2013) 说明了这些特征值和它们相应的单位特征向量可以一致估计得到, 估计时使用 $\boldsymbol{\Sigma}$ 的 Chamberlain-Rothschild 近似因子分解的残差估计 $\boldsymbol{R}_p$ 的阈值, 估计值等于利用样本协方差矩阵来替代 $\boldsymbol{\Sigma}$ 时的主成分正交分解阈值 (POET) 估计量. 论文同时提出了 ν 的一致估计量.

9. **高维回归: 稀疏和路径依赖变量选择** 线性回归模型:

$$y_t = \alpha + \beta_1 x_{t1} + \cdots + \beta_p x_{tp} + \epsilon_t, \quad t = 1, \ldots, n, \tag{2.34}$$

其中 ϵ_t 表示随机且无法观测的误差项, 满足 $E(\epsilon_t) = 0$. 该模型是广泛研究、广泛应用的统计模型. 当选取 x_{ti} 为观测输入值的基本函数时, (2.34) 可以使用非参数方法求得. 另外, 当某个 x_{ti} 选取为 y_{t-i}, 而其他的 x_{ti} 选取为外生的协变量和它们的滞后项时, 可得到一系列的时间序列模型; Lai 和 Xing (2008) 一书的 7.1–7.4、9.4 和 10.3 节对此进行了阐述. 然而, 尽管目前对于时间序列的理论和应用研究很多, 但是 (2.34) 在上一个十年再次成为重要的研究领域, 原因在于大数据使得 p (输出项的数量) 大于 n (样本数量). $p \gg n$ 最初看起来是无解的, 因为其变量数远大于样本数, 因此参数无法被很好地估计, 导致估计量的方差过大. 从另一方面来说, 回归函数 $f(x_1, \ldots, x_p) = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p$ 在回归模型 "稀疏" 时, 仍然是可估计的. 同时, 该模型的很多应用也包含稀疏回归模型. 在 $p \gg n$ 时, 估计量 $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ 主要有两个问题. 第一个问题是 $\boldsymbol{X}^\top \boldsymbol{X}$ 的奇异性, 注意到在 $p \gg n$ 时, $\boldsymbol{\beta}^\top \boldsymbol{x}_1, \ldots, \boldsymbol{\beta}^\top \boldsymbol{x}_n$ 的值并不能唯一

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

确定 $\boldsymbol{\beta}$, 其中 $\boldsymbol{X}=(x_{tj})_{1\leqslant t\leqslant n,1\leqslant j\leqslant p}$, 而 $\boldsymbol{x}_i^\top$ 是 $\boldsymbol{X}$ 的第 i 个行向量. 当假设 ϵ_t 独立同分布, 服从正态分布, 且 $\boldsymbol{\beta}$ 的先验分布为 $\boldsymbol{\beta}\sim N(0,\lambda\boldsymbol{I})$ 时, $\boldsymbol{\beta}$ 的后验分布为岭回归估计量 $\hat{\boldsymbol{\beta}}^{\text{ridge}}=(\boldsymbol{X}^\top\boldsymbol{X}+\lambda\boldsymbol{I})^{-1}\boldsymbol{X}^\top\boldsymbol{Y}$, 其中 $\boldsymbol{Y}=(y_1,\ldots,y_n)^\top$ (参见 Lai 和 Xing (2008, 第 107 页)), 因此可以消除奇异性. 后验均值最小化加惩罚函数时的残差平方和 $\|\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B}\|^2+\lambda\|\boldsymbol{\beta}\|^2$, 其惩罚函数为 L_2 形式 $\|\boldsymbol{\beta}\|^2=\sum_{j=1}^p\beta_j^2$, 正则参数为 λ. 尽管参数的个数远远大于样本数, 此时惩罚函数的值大多都很小, 且收敛于 0. 尽管 L_2 惩罚函数并不能得到稀疏估计量 $\hat{\boldsymbol{\beta}}^{\text{ridge}}$, 但是 L_1 形式的惩罚函数则可以. Tibshirani (1996) 提出的惩罚函数 $\|\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\|^2+\lambda\sum_{j=1}^p|\beta_j|$ 最小化估计量 $\hat{\boldsymbol{\beta}}^{\text{lasso}}$ 称为 LASSO (最小化绝对压缩和选择因子) 估计量, 该惩罚函数会使得一些特定的系数为 0, 且将其他系数压缩至 0, 因此可以同时进行子集选择和压缩. 与岭回归不同, $\hat{\boldsymbol{\beta}}^{\text{lasso}}$ 并没有显式解, 除非 $\boldsymbol{X}$ 的列向量是正交的, 但是 $\hat{\boldsymbol{\beta}}^{\text{lasso}}$ 可以利用凸优化方法求出. Zou 和 Hastie (2005) 引入了弹性净估计量 $\hat{\boldsymbol{\beta}}^{\text{enet}}$, 对于 L_1,L_2 形式的惩罚函数的如下线性组合进行最小化:

$$\hat{\boldsymbol{\beta}}^{\text{enet}}=(1+\lambda_2)\arg\min_{\beta}\{\|\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\|^2+\lambda_1\|\boldsymbol{\beta}\|_1+\lambda_2\|\boldsymbol{\beta}\|^2\},\qquad(2.35)$$

其中 $\|\boldsymbol{\beta}\|_1=\sum_{j=1}^p|\beta_j|$. (2.35) 中的 $(1+\lambda_2)$ 用于纠正弹性净估计量的 "双重压缩" 效应, 该统计量是贝叶斯统计量, 其先验密度为 $\exp\{-\lambda_2\|\boldsymbol{\beta}\|^2-\lambda_1\|\boldsymbol{\beta}\|_1\}$, 这也是高斯先验分布 (用于岭回归) 和双指数分布 (用于 LASSO) 之间的折中. 注意到 (2.35) 式仍然是一个凸优化问题. (2.35) 中的正则参数 λ_1,λ_2, 以及岭回归或 LASSO 回归中的 λ, 利用 k 重交叉验证方法计算.① R 语言程序包 `glmnet` 也可用于计算 $\hat{\boldsymbol{\beta}}^{\text{lasso}}$, $\hat{\boldsymbol{\beta}}^{\text{enet}}$, 以及进行 k 重交叉验证.

由于对应于 L_0 惩罚函数的如下非凸优化问题, 即最小化

$$\left\{\|\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\|^2+\lambda\sum_{j=1}^p I_{\{\beta_j\neq0\}}\right\}\qquad(2.36)$$

的问题, 在 p 很大时是无法求解的, LASSO 有时作为 (2.36) 式的凸优化近似. Ing 和 Lai (2011) 近年来在 Temlyakov (2000) 的基础上, 提出了一

① Lai 和 Xing (2008) 一书的 7.3.2 节对于去掉一个观测值的交叉验证方法进行了说明, 该方法将去掉第 i ($i=1,\ldots,n$) 次观测值的数据作为测试集, 而剩下的数据作为训练集. 一个更灵活的方法是将数据分为 k 个大致相等的数据集 (当 k 可以被 n 整除时, 每个数据集的大小为 n/k), 将其中 1 个部分作为测试集, 其他的 $k-1$ 个部分作为训练集. 交叉验证估计得到的平方预测误差是 k 个训练集的平方预测误差的均值.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

个快速步进回归方法, 称为正交贪婪算法 (OGA), 并将其与高维信息准则 (HDIC) 相结合, 用于 OGA 路径的变量选择. 该方法提供了 L_0 惩罚函数情况下的近似解, 求解包含三部分. 第一部分是利用贪婪算法来进行前向选择, 从而使得第一步选取的变量都能最小化现有变量和选择变量的 OLS 的残差平方和. OGA 方法对于选取的变量进行正交化处理, 利用逐部分的线性回归计算 OLS, 从而避开了矩阵求逆. OGA 算法在信号处理中也称为正交匹配算法 (Tropp 和 Gilbert (2007)). 第二个部分是一个停止准则, 在包含 K_n 个变量之后停止算法. $K_n = O\left(\left(\frac{n}{\log p_n}\right)^{\frac{1}{2}}\right)$ 的选取是基于收敛速度, 以反映 OGA 迭代中的偏差–方差权衡. 第三部分是在 OGA 路径上的变量选择, 选取时基于:

$$HDIC(J) = n \log \hat{\sigma}_J^2 + \#(J) w_n \log\ p,$$

其中 $J \subset \{1, \ldots, p\}$ 反映了选择变量集, $\hat{\sigma}_J^2 = n^{-1} \sum_{t=1}^{n} (y_t - \hat{y}_{t:J})^2$, 其中 $\hat{y}_{t:J}$ 表示 $\boldsymbol{Y}$ 映射到 $\boldsymbol{X}_j, j \in J$ 生成的线性空间时的拟合值, $w_n = c$ 表示信息准则, 如在 HDBIC 时, 选择 $w_n = \log n$, 而在 HDAIC 时, 选择 $w_n = c$ 为不随时间变化的常数. 定义 $\hat{J}_k = \{\hat{J}_1, \ldots, \hat{J}_k\}$ 为在 OGA 迭代中到第 k 步时选取的变量, Ing 和 Lai (2011) 选取 $\hat{k}_n = \arg \min_{1 \leqslant k \leqslant K_n} HDIC(\hat{J}_k)$, 并在 $HDIC(\hat{J}_{\hat{k}_n} - \{\hat{J}_l\}) > HDIC(\hat{J}_{\hat{k}_n})$ 时去除 OGA 路径上无关变量 $\hat{J}_l$. 该方法通常用 OGA+HDIC+Trim 表示, 论文中指出在强稀疏假设下, 在一个最小相关集下, 该方法的结果等价于 OLS.

10. **动态因子模型和多步预测** 我们继续讨论补充材料 8 中的因子模型, 并阐述因子的时间序列模型. Barigozzi 和 Hallin (2016) 在 Fan 等 (2013) 的基础上, 利用近似稀疏假设 (2.33) 和只有 ν 个 "尖" 的特征值对高维协方差矩阵进行时间序列分析, 基于 p ($T \ll p$) 个股票的收益的时间序列数据求解. 论文指出, 该方法并没有充分考虑时间序列特性, 尽管 ε_t 是严平稳序列 (而不是独立同分布序列), 且具有充分快的混合速率. 基于如下的宏观建模中的动态因子模型, 论文提出了一个一般动态因子模型 (DFM), 该模型包含 "一般水平" 和 "特质水平" 两个部分, 其中 "水平" 指通过向量自回归模型中的历史收益的资产收益水平. 另外, 论文同时提出了一个相应的特质水平部分的波动率的非线性转换分解. 该模型与 2.6.3 节中 Lai 等人 (LXC) 提出的鞅回归模型有关, 在该模型中也包含水平相关的部分 (2.19) 和波动率相关的部分 (2.20).

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

Barigozzi 和 Hallin (2016) 中的 DFM 模型起源于宏观时间序列, 特别是用于估计财政和其他经济政策的发展和评估的主要宏观变量的多步前瞻预测. 由于多步预测是包含在经济的政策创新影响的分析和预测中, 在 Bernanke 和 Blinder (1992) 以及 Sims (1992) 之后, 向量自回归模型 (VAR)[①]被广泛用于相关研究. 然而, Bernanke 等 (2005) 指出, "为了保证一定的自由度, 标准的 VAR 模型通常不会包含多于 6–8 个变量", 此时的预测至少会产生三个问题. 一是 "价格问题", 用于预测在冲突的财政策略冲击后价格水平的增长 (而不是预期的下降). 第二个问题是在 VAR 分析的信息集中, "需要基于特定的观测值的测度" (如使用工业产出或实际 GDP 来作为经济行为的代表性测度). 第三点是 "脉冲冲击只能基于包含的变量进行, 因此通常只由研究者和策略制定者关心的一小部分变量构成", 但是并不能考虑策略冲击对其他变量的可能影响. 因此, 论文提出了一个因子增强向量自回归模型 (FAVAR). 此处的因子分析或主成分分析通常指宏观指示值 $\boldsymbol{x}_t$ 的协方差矩阵, 它们在 FAVAR 中用于确定 VAR 分析中基本经济变量 $\boldsymbol{y}_t$ 的增加预测值, 参见 Forni 等 (2005) 以及 Stock 和 Watson (2002, 2011). 定义 $\boldsymbol{y}_t = (y_{t1}, \dots, y_{tq})^\top$ 为步进预测中的基本变量的向量. FAVAR 模型是对于因子增加时间序列 $\boldsymbol{A}(L)^\top \boldsymbol{y}_t = \boldsymbol{G}(L)^\top \boldsymbol{f}_{t-1} + \boldsymbol{\eta}_t$, $\boldsymbol{\Phi}(L)^\top \boldsymbol{f}_t = \boldsymbol{\Psi}(L)^\top \boldsymbol{y}_{t-1} + \boldsymbol{w}_t$ 的 VAR 模型, 其中 $\boldsymbol{f}_t$ 为因子模型 (2.29) 中的因子, $\boldsymbol{A}(L) = \boldsymbol{I} - \sum_{s=1}^{d_1} \boldsymbol{A}_s L^s$, $\boldsymbol{G}(L) = \sum_{s=0}^{d_2} \boldsymbol{G}_s L^s$, $\boldsymbol{\Phi}(L) = \boldsymbol{I} - \sum_{s=1}^{d_3} \boldsymbol{\Phi}_s L^s$, $\boldsymbol{\Psi}(L) = \sum_{s=0}^{d_4} \boldsymbol{\Psi}_s L^s$ 为滞后因子 L 的矩阵多项式.

11. **断点检测和相关的技术准则** 2.8.1 节中提到的筛选规则与断点检测中的 CUSUM 准则之间的联系可以扩展到其他断点检测的技术准则中. Lai (1995) 提出了一种可以将假设检验转化成断点检测准则的通用方法. 论文中的引理 1 指出, 如果 $X_1, X_2, \dots$ 是独立同分布的随机向量, 且 τ 是满足 $P(\tau < \infty) \leqslant \alpha$ 的关于 $X_1, X_2, \dots$ 的停时. 将 τ 应用于 $X_k, X_{k+1}, \dots$, 所获得的停时定义为 N_k, 则 $N = \min\limits_{k \geqslant 1}(N_k + k - 1)$ 是一个停时, 且满足 $EN \geqslant 1/\alpha$. 在单边序贯概率比检验 (SPRT) 下, 假设在满足如下条件时停止抽样:

$$\tau = \inf\left\{ t : \sum_{i=1}^{t} \log\left(\frac{f_1(X_i)}{f_0(X_i)}\right) \geqslant \log\gamma \right\}$$

且拒绝原假设 $H_0 : f = f_0$, 而接受备择假设 $H_1 : f = f_1$. 在该检验中, 满

① Lai 和 Xing (2008) 的第 9.4 节给出了 VAR 模型的介绍.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

足 $P_0(\tau<\infty)\leqslant\exp\{-\log\gamma\}=\gamma^{-1}$, 则根据引理可知 $E_0(N)\geqslant\gamma$, 其中 P_0 和 E_0 分别指 $X_1,X_2,\ldots$ 独立同分布, 概率密度函数为 f_0 时的概率测度和期望. 另外, 根据单边 CUSUM 准则, N 满足

$$N=\inf\left\{n\geqslant 1:\max_{1\leqslant k\leqslant n}\sum_{i=k}^{n}\log\left(\frac{f_1(X_i)}{f_0(X_i)}\right)\geqslant\log\gamma\right\}.$$

习题 证明筛选准则 (2.24) 的双边 CUSUM 准则对应于如下的双边 SPRT 准则: 原假设 $H_0:\mu=-\theta$, 备择假设 $H_1:\mu=\theta$, 停止准则为:

$$\tau^*=\inf\left\{n:\prod_{i=1}^{n}\frac{f_\theta(X_i)}{f_{-\theta}(X_i)}\notin\left(\frac{2\theta}{1+c},2\theta(1+c)\right)\right\},$$

其中 f_μ 为 $N(\mu,1)$ 的概率密度函数. 对于假设检验 $H_0:\mu=0$, $H_1:\mu=\theta$ 而言, 对应的 CUSUM 准则与包含斜率参数的筛选规则是等价的. Lam 和 Yam(1997) 将该斜率变量称为筛选速度.

习题 阅读 Lam 和 Yam (1997), 并讨论添加了筛选速度变量之后的筛选规则与通常的筛选规则之间的区别.

12. **移动平均和序贯 GLR 检测准则** Lai (1995) 将引理 1 应用到 Neyman-Pearson 检验中 (样本量固定为 m), 研究发现窗宽 m 取值合适时, 移动平均准则的检测效率与 CUSUM 准则的有效性大致相当. 阅读 Lai (1995) 论文中引理 2 的证明, 并阅读论文 3.3 节中对于同时使用不同窗宽的方法的讨论, 该方法首先由 Willsky 和 Jones (1976) 提出, 用于在变化后参数未知时计算广义似然比率 (GLR).

13. **无处不在的价值和动量** 这是 Asness 等 (2013) 论文的题目. 在该论文中, 对于四个资本市场 (美国、英国、欧洲大陆和日本) 中的个股和其他四种资产 (全球股票指数、全球政府债券、外汇和商品远期) 的价值和动量进行了实证分析. 研究提出了一个三因子模型, 三个因子分别为全球市场指数、包含各种资产的零成本价值策略和零成本动量策略. 该模型利用这些 "与价值和动量相关的全球变量", 可以 "捕捉到任意一种资产在全球市场与局部市场之间的联动关系". 该模型 "不仅可以捕捉 Fama-French 美股资产组合的特征, 也可以捕捉对冲基金指数的特征". 随着金融市场 "越来越趋于全球化", 建立 "一个适用于所有资产的模型, 而不是只适用于某个特定市场的模型" 显得更有必要.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

14. **基于前景理论的行为金融和投资组合选择** 2.8.3 节提到了前景理论的概念, 该理论也是 Daniel Kahnman 获得 2002 年诺贝尔经济学奖的原因. 与经济学经典的期望效用理论不同, 行为经济学中的前景理论对于人们在面对不确定收益时的行为进行了如下解释:

> 人们在面对不确定收益时的选择在很多方面都与效用理论不符合. 特别是相对于不确定性收益而言, 人们更偏好确定性收益, 这种现象被称为 "确定效应". 确定效应解释了人们在面对确定性收益时的风险规避行为和在面对确定性损失时的风险偏好行为. 另外, 人们通常会忽略在不同情况下都存在的部分, 这种效应称为孤立效应. 孤立效应说明, 当同样的选择以不同形式来呈现时, 人们往往会做出不一致的选择. 在另一个关于选择的替代性理论中, 直接对于收益和损失定义价值, 而不是对于金融资产定义, 且使用决策权重来替代概率. 前景理论中的价值函数在正收益部分是凹函数, 而在负收益部分是凸函数, 且在负收益部分曲线往往更陡. 前景理论中使用的决策权重通常低于事件自身的概率, 而在低概率部分高于自身概率, 这一特性也解释了保险和赌博受欢迎的原因.

Benartzi 和 Thaler (1995) 对于价值函数符合前景理论特征的投资者在单期内的投资决策进行了研究. Barberis 等 (2001) 将该理论应用到资产定价领域中, 作为衡量资产平均收益率的模型. Jin 和 Zhou (2008) 在前景理论的框架下建立了一个广义连续时间行为经济学资产选择模型, 在模型中假设投资者的效用函数符合 S 形曲线, 且使用扭曲概率来替代事件自身概率. 扭曲概率是对于事件本身概率的一个非线性变换, 增大了小概率事件的权重, 而减小了大概率事件的权重. 在使用扭曲概率时, 相应的应该使用 Choquet 积分[①]来替代传统的积分.

① 令 $\nu : \mathcal{F} \to [0, \infty)$ 单调, 即在 $A \subset B$ 时, $\nu(A) \leqslant \nu(B)$, 其中 $\mathcal{F}$ 为集合 S 的子集的集合. $f : S \to \mathbb{R}$ 对应于 ν 的 Choquet 积分定义为:

$$\int_{-\infty}^{0} (\nu\{s : f(s) \geqslant x\} - \nu(S))dx = \int_{0}^{\infty} \nu\{s : f(s) \geqslant x\}dx,$$

假设 f 对 ν 可积 (即对于所有 $x \in \mathbb{R}$, $\{s : f(s) \geqslant x\} \in \mathcal{F}$).

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

第三章　积极型投资组合管理和投资策略

量化投资组合管理的基础在于对于大量的投资机会进行期望收益预测、风险评估以及投资组合优化. 量化投资组合管理策略致力于将收益率与风险不同的各个资产组合进行优化, 通过合理的风险分散使得风险最小化. 在本章中, 投资组合管理是指构建与管理资产组合. 通常而言, 投资组合管理中的量化基金管理需要以下几个方面的专家:

- 量化研究员: 负责市场预测、风险评估、投资组合优化;
- 交易员: 负责日常的交易和再平衡, 并负责与券商商讨相关的交易事宜;
- 代销商: 负责产品推广, 以增加资产管理规模 (Asset under Management, AUM), 并分散客户群;
- 运营经理: 负责监督基金的一般事务管理, 包括客户财产的重新配置和赎回;
- 基金经理: 负责监督现有基金的所有事宜.

基金管理分为积极型管理和消极型管理. 消极型管理策略通过构建一个收益率等于某个特定指数的投资组合来进行管理, 该投资组合被称为 “追踪指数的投资组合”. 消极型管理策略认为任何偏离于该投资组合, 不管是收益率高于还是低于该指数的投资组合, 都不是最优的组合. 消极型管理策略的拥护者

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

认为提高投资净收益的关键在于降低投资支出, 而追踪指数的方法的优势在于可以直接利用特定的指数, 并不需要对于个股进行信息搜集和研究, 因此其成本较低. 另外, 消极型投资策略要求的交易频率较低, 使得资本收益也较低, 相应需要缴纳的税也更低, 进一步降低了成本. 积极型投资组合管理策略则致力于建立一个收益率高于 S&P 500、三个月国债收益率等现有的指数的投资组合 (参见 Grinold 和 Kahn (2000)). 当投资收益率高于基准利率时, 投资经理会被嘉奖; 反之, 当投资组合的收益率低于基准指数或者波动性高于基准指数时, 投资经理则会被处罚. 高于基准利率部分的收益率通常会被称为投资组合的 α. 通常而言, 在分析 α 时, 会计算增加一单位的 α 时, 所增加的风险. 而在评估一个投资组合的表现时, 则会使用信息比率等指标. 其中信息比率是指单位超额收益值与其标准差之间的比率.

如果一个投资经理所使用的投资组合仅仅是复制基准投资组合, 并使用杠杆来增加收益率进行投资, 则该投资组合相对于基准投资组合而言, 会产生额外的波动率, 其风险也更高. 收益率增加的效应会被风险的增加所抵消, 导致信息比率并无变化. 因此, 单纯地增加杠杆并不会为投资经理带来额外的价值. 要想增加价值, 需要利用更加有效的信息收集、数据分析及模型优化方法, 从而在提高 α 的同时也能够使得额外的风险最小化. 3.1 节和 3.2 节引入了积极型策略的 α 和 β, 讨论了积极型策略的相关问题和局限性, 并提出了解决这些问题的方法. 3.3 节讨论了积极型投资组合管理中包含的动态投资策略, 以及动态投资策略中的多种优化和滤波方法. 特别地, 当存在交易成本和其他限制时, 多期投资组合优化问题可以看成是 (1) 消极型策略和 (2) 连续时间投资组合重组问题的结合, 其中 (2) 在不存在交易成本的积极型策略中也会出现. 3.4 节和 3.5 节对于这些问题和相关的研究进行了相关阐述.

3.1 积极型投资组合管理中的 α 和 β

与第二章一致, 本章定义包含 m 个资产的投资组合的收益率向量为 $\boldsymbol{r} \in \mathbb{R}^m$, 投资组合的权重向量为 $\boldsymbol{w}_P$, 基准投资组合的权重向量为 $\boldsymbol{w}_B$, 则投资组合 P 的收益率与基准投资组合 B 的收益率分别为

$$r_P = \boldsymbol{w}_P^\top \boldsymbol{r}, \quad r_B = \boldsymbol{w}_B^\top \boldsymbol{r}.$$

在单因子模型下, 可得 $r_P - r_f = \alpha + \beta(r_B - r_f) + \epsilon$, 其中 r_f 是无风险利率. 等式两边同时减去 $r_B - r_f$, 则可得到积极型投资组合管理中的收益率如下:

$$r_P - r_B = \alpha + \beta_P (r_B - r_f) + \epsilon, \tag{3.1}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 α 为相对于基准收益率的超额收益率, $\beta_P = \beta - 1$ 则是积极型投资策略的 β, 而 ϵ 是一个零均值的随机变量, 一般被称为残差项, 其与 $r_B - r_f$ 不相关.

3.1.1 超额收益率 α 的来源

积极型投资策略的关键是产生超额收益率 α, 而 α 值也是评价基金经理能力的指标. α 值的两个主要来源包括 (a) 信息优势; (b) 对于信息的有效处理. 如今上市公司公布的信息会受到多方面的监管审查, 因此基于内幕消息的投资行为并不可行. 为了在投资市场上占据一定的优势, 投资者不仅需要对于公开的财务报告和信息披露进行研究, 更需要从相关的信息中发掘出有价值的观点, 从而对资产未来的价值进行预测. 举例来说, 当投资者或者股票分析师实地访问一个公司时, 可以对于公司的财务报告的可信性有更好的了解, 从而为对于公司股票未来价值进行预测提供基础.

对于信息的有效利用包括对于现有的模型进行研究, 并开发出能够对于未来的资产价值进行更精确的估计的模型. 新模型的预测结果应该能够合理利用市场的基本信息, 并且能够根据现有信息预测市场的行为. 举例来说, 通过对于公司的供应链进行更好的了解, 新的模型可以对于供应链之间的关系进行更详细的描述, 从而在供应链关系中断时, 能够帮助投资者更好地理解和量化该事件对于公司的影响. 在大数据时代, 投资市场广阔, 对于市场中丰富的信息进行处理和应用以获得超额收益, 机会与挑战并存.

3.1.2 积极的 α 之外的奇异 β

Litterman (2004, 2008) 指出, 超常收益不仅可以来源于积极的 α, 也可以来源于 "不太积极的" β. Litterman 指出, Black, Jensen 和 Scholes (1969) 提到 "一个波动率 β 较低的股票与一个 β 值较高的股票相比, 其收益率显得过高, 而与 CAPM 理论显得不够一致". 根据 (3.1) 式可知, $\mathrm{Var}(r_P - r_B) = \beta_P^2 \mathrm{Var}(r_B) + \mathrm{Var}(\varepsilon)$, 因此当 β_P^2 下降时, 意味着 $r_P - r_B$ 的方差更小, 因此与相同收益率但 β 值更大的投资组合相比, 其信息比率更高. 特别地, 由于基准投资组合的 β 值为 1, 因此一个 α 值为正, 且 $|\beta| < 1$ 的投资组合的信息比率高于基准投资组合. 这样的投资组合被称为含有 "奇异 β (exotic beta)". Litterman (2004) 指出传统的 CAPM 框架只适用于单个 β 的情况, 而目前市场上的投资者 "会同时投资于不同的资产, 对于不同的资产会有不同的 β", 从而 "既获得超额收益 α, 又能够比市场投资组合表现更好". Litterman 提出, "将使用积极型投资组合管理以产生 α 与使用消极型策略来获得奇异 β 相结合",

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

可以获得一个更好的资产管理的方法. Litterman (2008) 进一步指出, “市场的 β” 是指 “经典的 CAPM 模型下市场投资组合” 的 β. 该投资组合的奇异 β “使投资者可以用频谱的方式来分析收益率的来源, 频谱从成本低且市场供应量为无穷大但 Sharpe 比率很低的市场 β 开始”. 另一方面, 奇异的 β 的来源 “从消极投资组合管理中的资产到一些不太常见的资产类型, 如高收益率的债券、环球小型企业股票、新兴市场股票、新兴市场债务、商品和环球房地产市场等”, 这些资产的优势在于其市场供给大, 收益潜力高, 且与环球市场投资组合的关联较低. Litterman 认为, 市场上最终会达到均衡, 且奇异 β 作为一种收益率的来源并不会永远存在. 市场均衡可能会在很多年后才会发生, 且 “同时, 投资者会在所投资的资产价值被抬高之后获益”.

3.1.3 一种积极型投资组合优化的新方法

积极型投资组合管理策略旨在获得比基准投资组合更高的收益率, 投资经理的酬劳也与超额收益率有关. 定义基准投资组合的权重向量为 $\boldsymbol{w}_B$, 新的投资组合 P 的权重向量为 $\boldsymbol{w}_P$. 两个权重向量之差 $\widetilde{\boldsymbol{w}} = \boldsymbol{w}_P - \boldsymbol{w}_B$ 满足 $\widetilde{\boldsymbol{w}}^\top \mathbf{1} = 0$. 对于 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 未知, 且权重向量 $\widetilde{\boldsymbol{w}}$ 的积极型投资组合问题, 定义 Markowitz 优化问题 (2.9) 如下:

$$\begin{aligned} &\max \left\{ E(\widetilde{\boldsymbol{w}}^\top \boldsymbol{r}_{n+1}) - \lambda \mathrm{Var}(\widetilde{\boldsymbol{w}}^\top \boldsymbol{r}_{n+1}) \right\}, \\ &\text{满足 } \widetilde{\boldsymbol{w}}^\top \mathbf{1} = 0, \widetilde{\boldsymbol{w}} \in \mathcal{C}, \end{aligned} \tag{3.2}$$

其中 $\mathcal{C}$ 代表做多–做空限制等约束条件. 通过调整 λ 可得到有效前沿. 此处, 我们选择使用 2.7.2 节中的方法, 根据数据来求得能使信息比率最大的参数值. 因此, 该优化问题的解可以利用类似 2.7 节中的方法来求得. 相比于 2.7 节中的方法, 该方法的创新之处在于其是通过基准投资组合权重的方法来求解, 而权重值每天都会更新, 因此已经包含了 “免费的” 市场信息, 参见 3.4 节的补充材料 1. 该方法使我们可以求得 (3.2) 的近似解, 而且正如我们在 2.9 节的补充材料 4 中指出, 该解中包含专业知识和对未来股票及其相关联的公司和金融部门的变动情况的先验估计. 如 3.1.1 节中指出, 利用附加信息和更先进的分析, 即可获得 α.

上述求解方法与传统的优化方法不同. 在传统的方法中, 投资组合经理通常会先制定一个目标收益率 $\widetilde{\mu}_*$, 该收益率也是收取管理费的依据. 投资经理在制定积极型投资策略时, 会考虑如下优化问题:

$$\min E(\widetilde{\boldsymbol{w}}^\top \boldsymbol{\Sigma} \widetilde{\boldsymbol{w}}), \quad \text{满足 } \widetilde{\boldsymbol{w}}^\top \boldsymbol{\mu} = \widetilde{\mu}_*, \ \widetilde{\boldsymbol{w}}^\top \mathbf{1} = 0, \ \widetilde{\boldsymbol{w}} \in \mathcal{C}; \tag{3.3}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

参见 (Ledoit 和 Wolf (2004), 在本章中, 后续简称为 LW). Lai 等 (2011b, 简称为 LXC) 在 LW 的基础上进行了后续研究. LXC 定义基准投资组合为由 50 只股票组成的加权投资组合, 约束条件为 $\mathcal{C} = \{\widetilde{\boldsymbol{w}} : -\boldsymbol{w}_B \leqslant \widetilde{\boldsymbol{w}} \leqslant c\mathbf{1} - \boldsymbol{w}_B\}$, 其中 $c = 0.1$, 即该投资组合只能做多, 不能做空, 且投资者持有的任意股票的头寸数不得高于 c. 论文使用的样本取自 CRSP 数据库中在过去 120 个月没有缺失数据, 且月初市值排名在前 50 的股票. 此处需要求解的优化问题类似于(3.3), 但其中 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 替代为利用上述样本对于嵌入式积极型投资组合的估计值. 而在分析方差压缩投资组合时, 则使用相对应的 $\boldsymbol{\Sigma}$ 估计值. 在进行压缩估计时, 压缩指向一个相关系数都相等的矩阵 (参见 Ledoit 和 Wolf (2003)). LXC 将 2.5 节中的 NPEB 方法进行拓展, 构建了相应的 NPEB 积极型投资组合, 但并未包含与表 3.1 中其他方法可比的专业知识.

表 3.1 价值加权的投资组合的年平均已实现收益率的均值和标准差

$\widetilde{\mu}_*$	0.01	0.015	0.02	0.03
λ	2^2	2	2^{-1}	2^{-2}
(a) 通过在某些区间重新定义投资组合的所有测试区间结果				
Plug-in	0.001 (4.7e-3)	0.002 (7.3e-3)	0.003 (9.6e-3)	0.007 (1.4e-2)
Shrink	0.003 (4.3e-3)	0.004 (6.6e-3)	0.006 (8.8e-3)	0.011 (1.3e-2)
Boot	0.001 (2.5e-3)	0.001 (3.8e-3)	0.001 (5.1e-3)	0.003 (7.3e-3)
NPEB	0.029 (1.2e-1)	0.046 (1.3e-1)	0.053 (1.5e-1)	0.056 (1.6e-1)
(b) 所有投资组合都定义准确的测试区间结果				
Plug-in	0.002 (6.6e-3)	0.004 (1.0e-2)	0.006 (1.4e-2)	0.014 (1.9e-2)
Shrink	0.005 (5.9e-3)	0.008 (9.0e-3)	0.012 (1.2e-2)	0.021 (1.8e-2)
Boot	0.001 (3.5e-3)	0.003 (5.3e-3)	0.003 (7.1e-3)	0.006 (1.0e-2)
NPEB	0.282 (9.3e-2)	0.367 (1.1e-1)	0.438 (1.1e-1)	0.460 (1.1e-2)

在表 3.1 中, 分析了四种不同类型的投资组合, 包括嵌入式 (表中的 “Plug-in”)、方差压缩 (表中的 “Shrink”)、重抽样 (表中的 “Boot”) 和 NPEB(表中的 “NPEB”). 同样的 $\widetilde{\mu}_*$ 对于有的投资组合可能是没意义的. 定义目标收益率取值分别为 $\widetilde{\mu}_* = 0.01, 0.015, 0.02, 0.03$, $\boldsymbol{\Sigma}$ 取值为样本方差–标准差矩阵或 Ledoit-Wolf 压缩估计值时, (3.3) 式分别在 92, 91, 91, 80 个测试区间上有解. 目标收益率越高, 在这 180 个测试区间中, (3.3) 式有解的区间就越少. 另一方面, 当 $\widetilde{\mu}_*$ 取值低于 1% 时, 并不会吸引投资经理注意. 当 (3.3) 式无解时, 则在测试区间中使用价值加权的基准投资组合来代替. 在表 3.1(a) 中, 分别给出了在 180

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

个测试区间内嵌入式、方差压缩和重抽样这三类投资组合的实际 (年化) 平均收益率 $12\bar{e}$ 以与目标收益率相比较, 并给出了相应的标准差 $\sqrt{12}s_e$. 如表所示, 这些数字非常小, 说明尽管其信息比率的取值在 0.24 到 0.83 之间变化, 但是这三类投资组合与基准投资组合并无很大差别. 因此, 在忽略实际的平均超额收益率均值的情况下, 仅考虑信息比率可能会有误导作用. 表 3.1(a) 也给出了 NPEB 准则下投资组合的已实现年化收益率的均值和标准差, 其中在选取 λ 的值时, 选择在 $\lambda = 2^j$ $(-2 \leqslant j \leqslant 2)$ 范围内能够使得平均收益率与 $\widetilde{\mu}_*$ 大致相同的 λ 值. 注意到 NPEB 准则下投资组合的平均收益率显著高于其他三类投资组合.

在 (3.3) 式无解时的测试区间中, LXC 定义了一个违约期权. 期权基于一个包含收益率排名前 10 的股票, 且其权重相同的投资组合. 研究发现, 在使用违约期权的这些测试区间内, 已实际超额收益率的均值 $12\bar{e}$ 均为负数 (取值在 -17.4% 与 -16.3% 之间), 而 $\widetilde{\mu}_*$ 的取值则在 1% 到 3% 之间. 表 3.1(b) 仅仅对于 80–92 测试区间进行了研究, 在这些区间中嵌入式、方差压缩和重抽样类的投资组合在 $\widetilde{\mu}_* = 0.01, 0.015, 0.02$ 和 0.03 时都可以由 (3.3) 式进行定义. 注意到这三类投资组合的平均超额收益率都非常小, 而 NPEB 类的投资组合的收益率比这三类大很多. NPEB 类投资组合的已实现信息比率取值在 3.015 到 3.954 之间, 而其他三类投资组合的信息比率则在 0.335 到 1.214 之间. LXC 中也将 S&P 500 指数与价值加权的投资组合进行了比较. S&P 500 指数在该区间内的均值为 0.006, 标准差为 0.046, 而价值加权的投资组合的均值和标准差分别为 0.0137 和 0.045, 可见价值加权的投资组合相对于此处的基准投资组合, 即 S&P 500 指数的优势所在.

3.2 交易成本和买卖限制

3.2.1 交易成本的组成

在不存在交易成本时, 积极型投资组合管理显然比消极型策略更好, 但是在买卖投资组合的各资产过程中存在交易成本时, 该优势往往会被抵消.

交易成本可以分为三类: 佣金、买卖价差和市场影响. 这三类成本共同组成了投资组合的总*执行落差*, 即实际的投资组合收益率与不存在任何交易成本的理想情况下投资组合收益之差.

交易成本的第一部分, 即佣金, 包括经纪人佣金、交易所佣金、清算所费用等组成成分. 这些成本通常是交易量的一个固定比例, 不同的标的资产的佣

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

金也会有所不同. 股票的佣金通常只有每股几分钱, 而期货和期权的佣金每单会接近一美元. 由于佣金是一个固定的数额, 因此在建立投资组合时很容易对其建模.

交易成本的第二部分, 即买卖价差, 随着市场情形的变化而变化, 对于不同的标的资产买卖价差也不同. 在正常的市场情形下, 美国交易所上市的股票的买卖价差通常会在股票价格的 0.02% 左右, 价差的单位是 \$0.01, 这个单位也被称为最小报价单位. 最小报价单位由交易所制定, 且适用于所有的股票. 期货的买卖价差随着标的资产的不同而不同, 通常会在每个合约 \$10 到 \$25 之间. 买卖价差是资产价格波动性和流动性的函数, 因此是一个随机变量. 在近期, 由于高频做市商在提供市场流动性方面的激烈竞争, 买卖价差大幅下降. 对于大部分资产而言, 买卖价差保持在最小报价单位左右摇摆.

交易成本的第三部分, 即市场影响, 是最难进行量化的. 市场影响无法预测, 只能利用统计学和经济学模型进行推测. 大量的论文对此进行了研究, 而近期的研究主要侧重在市场微观结构和行为金融学方面, Hasbrouck (1996) 和 Bouchaud 等 (2002) 对于相关的研究进行了综述. 在对于交易成本进行建模时, 一个被广泛使用的基本形式是将交易成本作为期望收益的一个比例. 在 Markowitz 投资组合优化模型 (2.1) 中加入交易成本, 则可得到如下形式:

$$\begin{aligned} \boldsymbol{w}_{\text{eff}} &= \arg\min_{\boldsymbol{w}} \boldsymbol{w}^{\top}\boldsymbol{\Sigma}\boldsymbol{w} \\ &\text{满足 } \boldsymbol{w}^{\top}\boldsymbol{\mu} + \boldsymbol{\phi}^{\top}(\boldsymbol{w}_{+} + \boldsymbol{w}_{-}) = \mu_{*}, \\ &\quad \boldsymbol{w}^{\top}\mathbf{1} = 1,\ \boldsymbol{w}_{+} \geqslant \mathbf{0}, \boldsymbol{w}_{-} \geqslant \mathbf{0}, \end{aligned}$$

其中 $\boldsymbol{w}_{+}$ (或 $\boldsymbol{w}_{-}$) 分别是正权重和负权重, $\boldsymbol{w} = \boldsymbol{w}_{+} - \boldsymbol{w}_{-}$. 向量 $\boldsymbol{\phi}$ 指投资组合中各资产的单位交易成本.

3.2.2 买卖限制和其他约束条件

在求解投资组合优化问题时, 交易成本约束可以看成是良性约束, 因为其并不会显著增加优化问题的复杂程度, 优化问题依旧可以用二次规划的方法进行求解. 除交易成本外, 买卖限制也是一个良性约束. 当存在买卖约束时, 优化问题可写成如下形式:

$$\begin{aligned} \boldsymbol{w}_{\text{eff}} &= \arg\min_{\boldsymbol{w}} \boldsymbol{w}^{\top}\boldsymbol{\Sigma}\boldsymbol{w} \\ &\text{满足 } \boldsymbol{w}^{\top}\boldsymbol{\mu} + \boldsymbol{\phi}^{\top}(\boldsymbol{w}_{+} + \boldsymbol{w}_{-}) = \mu_{*}, \\ &\quad \boldsymbol{w}^{\top}\mathbf{1} = 1,\ 0 \leqslant \boldsymbol{w}_{+}^{\top}\mathbf{1} \leqslant \psi_{L},\ 0 \leqslant \boldsymbol{w}_{-}^{\top}\mathbf{1} \leqslant \psi_{S}, \end{aligned}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 ψ_L 和 ψ_S 分别指代投资组合买入和卖出的权重上限. 然而, 一些其他的约束条件可能会大幅增加优化问题的复杂性, 特别是在添加了约束条件后, 优化问题不再是一个凸优化问题时, 此时非凸优化问题可能在多项式时间上无解. 在此情况下, 优化问题只会在约束条件较少或者使用启发式算法时才能求解. 基数约束是一个常用的对交易量、交易合同数或者持有的资产头寸数的约束条件. 举例来说, 一个指数追踪类的投资组合可能会需要对于很多个资产进行追踪, 考虑到交易成本和可行性问题, 对于需要追踪的资产的数量进行限制显得非常有必要. 存在基数约束的指数追踪问题可以写成基数约束的二次规划问题 (Cardinality-Constrained Quadratic Program (CCQP)) 的形式, 如下所示:

$$\begin{aligned} \boldsymbol{w}_{\text{eff}} &= \arg\max_{\boldsymbol{w}} \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{1}{2}\lambda \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} \\ &\text{满足 } \boldsymbol{A}\boldsymbol{w} \geqslant \boldsymbol{b},\ \boldsymbol{w}^\top \mathbf{1} = 1,\ \sum_i \mathbf{1}_{\{w_i \neq 0\}} = K, \end{aligned}$$

其中 K 是最优投资组合的权重向量中非零项的个数. 如果投资者对于优化的要求不高的话, 可以选用连续截断的方法来得到可行解. 在求解时, 首先得到一个可行解, 然后在每次迭代过程中, 将权重较小或权重为 0 的项去掉, 最终得到合适解. 一些更精确的优化方法则需要利用能够对于目标函数在可行集上进行全局优化的算法, 或者能够在短时间内对于全局最优解进行充分逼近的算法. 3.4 节的补充材料 1 和 2 对于 CCQP 问题的求解和一些启发式算法进行了介绍.

3.3 多期投资组合管理

diBartolomeo (2012) 在讨论 “更聪明的投资组合再平衡” 问题时, 敏锐地指出, Markowitz 提出的 MPT 理论 “讨论了一个单期投资模型, 模型中的收益率服从一个固定的概率分布, 并假设分布的相关参数已知且不随时间变化”, 然而 “这两个假设并不符合现实情况”. 尽管目前在学术界对于 “多周期全局优化问题” 进行了大量讨论, 但这些方法并没有被 “投资专家”(如基金经理或投资顾问等) 实际使用. 目前投资专家依然是将精力集中于降低成本, 并使用单期均值–方差优化方法来进行投资组合的权重再平衡. diBartolomeo 注意到, 尽管 “全押” 是目前常用的方法, 但至少有两种 “消极的观点” 是不利于投资专家的. 第一种观点是: 当一些积极的投资经理由于再平衡的收益太小而变得不太活跃时, 客户会认为投资经理是在失职, 而不是为了降低交易成本而做的理性决定.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

第二种观点是: 当市场经历一段时间的不活跃期之后, 最终的再平衡策略会集中在某一个时间交易, 产生市场影响和高额交易成本. 在本节中, 我们对于 (a) 多期投资组合理论的学术研究; (b) 积极型投资组合经理和其他投资专家对于存在约束条件情况下的投资组合再平衡策略的讨论, 这两方面的内容进行了综述. 综述也引出了一些基本的建模和优化技术, 具体细节在 3.4 节中进行讨论, 在 3.4 节中, 也对于积极型投资组合管理方面近期一些跨学科方法和动态再平衡方法进行了讨论. 3.4 节中补充材料 4-7 介绍了动态优化准则和方法, 这些方法对于第六章和第七章中的量化交易的发展提供了背景.

3.3.1 Samuelson-Merton 基于随机控制理论的 "终身投资组合选择理论"

该理论最初在 Samuelson (1969) 的第一节提出, 论文脚注中提到其学生 Merton 在杂志同一期中发表的论文 "讨论了一个更复杂的相关问题, 即存在连续时间的随机变动情况下的最优控制问题" (即现在的随机控制问题). Samuelson 指出 "大多数的投资组合选择问题, 不管是 Markowitz-Tobin 均值优化问题, 还是更一般的类型, 都是对于单期的投资收益进行最大化," 但是, "终身的消费和投资决策问题" 涉及多期的拓展. Samuelson 在一个有限期的离散时间框架下, 利用标准的反向归纳法, 将多期优化问题转化成了 "经典的单期投资组合优化问题". 在每一期中, 将投资者的期望效用相加, 这也是广泛使用的理论决策方法. Samuelson 指出, 虽然该方法只适用于有限期的情况, 但在 "合理的条件下", 该方法可以推广到无限期 (即终身投资组合选择) 的情形.

Merton (1969) 在连续时间框架下对该问题进行了讨论, 在此情况下市场上包含两种投资产品: 一个收益率 $r > 0$ 恒定的债券和一个收益率服从平均回报率为 $\alpha > 0$、波动率为 $\sigma > 0$ 的几何布朗运动的股票. 因此, 在时间 $t \geqslant 0$, 股票价格为

$$dS_t = S_t(\alpha dt + \sigma dB_t), \tag{3.4}$$

其中 $\{B_t : t \geqslant 0\}$ 是一个标准的布朗运动过程. 定义投资者持有的头寸为 (X_t, Y_t), 其中

$$\begin{aligned} X_t &= \text{所投资的债券的价值}, \\ Y_t &= \text{所投资的股票的价值}, \\ y_t &= \text{持有的股票的数目}. \end{aligned} \tag{3.5}$$

其中 $Y_t = y_t S_t$. 3.4 节的补充材料 4 和 5 对于离散情况和连续情况的动态规划问题的基础进行了介绍.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

投资者的投资和消费决策包含三个非负随机过程 C, L 和 M, 其中 C 在任意有限时间区间内都可积, 而 L 和 M 则是非减序列, 满足右连续且有左极限. 具体来说, 投资者在时间 t 的消费率为 C_t, 在 $[0,t]$, $0 \leqslant t \leqslant T$ 时间累积买入 (或卖出) 的股票价值为 L_t (或 M_t). 因此, 投资者持有的头寸满足:

$$dX_t = (rX_t - C_t)dt - dL_t + dM_t, \tag{3.6}$$

$$dY_t = \alpha Y_t dt + \sigma Y_t dB_t + dL_t - dM_t. \tag{3.7}$$

在投资者在任意时间都具备偿付能力 (即净资产非负) 的前提下, 投资者持有的头寸应该始终处于偿付能力集 $\mathcal{D}$ 中, $\mathcal{D}$ 是一个闭合的凸集合, 并满足

$$\begin{aligned} \partial_\lambda \mathcal{D} &= \{(x,y) : x > 0,\ y < 0,\ x + y = 0\}, \\ \partial_\mu \mathcal{D} &= \{(x,y) : x \leqslant 0,\ y \geqslant 0,\ x + y = 0\}. \end{aligned} \tag{3.8}$$

定义可行策略集为 $\mathcal{A}(t,x,y)$. 对于投资者持有的头寸 $(X_t, Y_t) = (x,y)$, 在任意时间 $t \leqslant s \leqslant T$, 满足 $(X_s, Y_s) \in \mathcal{D}$. 或者等价地说, 定义 $Z_s = X_s + Y_s$, 则对于任意 $t \leqslant s \leqslant T$ 满足 $Z_s \geqslant 0$. 在时间 t, 投资者的目标是在可行集 $\mathcal{A}(t,x,y)$ 内, 最大化如下的期望效用函数:

$$J(t,x,y) = E\left[\int_t^\top e^{-\beta(s-t)} U_1(C_s)ds + e^{-\beta(T-t)} U_2(Z_T) \middle| X_t = x, Y_t = y\right], \tag{3.9}$$

其中 $\beta > 0$ 是折现系数, U_1 和 U_2 是两个凹函数, 代表消费和最终财产的效用. 假设 U_1 和 U_2 会从如下的 CRRA[①]函数中选取:

假设 U_1 可微, 且其反函数 $U_1'^{-1}$ 存在. 通常而言,

$$U(c) = c^\gamma/\gamma \text{ 当 } \gamma < 1, \gamma \neq 0; \quad U(c) = \log c \text{ 当 } \gamma = 0, \tag{3.10}$$

其具有恒定的相对风险厌恶系数$-cU''(c)/U'(c) = 1 - \gamma$. 投资者的价值函数定义为

$$V(t,x,y) = \sup_{(C,L,M)\in\mathcal{A}(t,x,y)} J(t,x,y). \tag{3.11}$$

将 (3.6) 式与 (3.7) 式相加, 可知投资者的资产总价值 $Z_t = X_t + Y_t$ 满足:

$$dZ_t = \{rZ_t + (\alpha - r)\theta_t Z_t - C_t\}dt + \sigma\theta_t Z_t dB_t, \tag{3.12}$$

① 经济学中的风险厌恶指人们通常偏好一些结果确定的选择. 定义效用函数为 $U(x)$, 其中 x 表示个体可能获得的财物. 当 U 为凹函数时, 称个体为风险厌恶; 而当 U 为凸函数时, 称个体为风险喜好; 而当 U 为线性时, 称个体为风险中性. Arrow-Pratt 系数定义为: 绝对风险厌恶系数 $-U''(x)/U'(x)$, 相对风险系数 $-xU''(x)/U'(x)$.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 $\theta_t = \dfrac{Y_t}{X_t + Y_t}$ 是投资者投资在股票中的比例. 重新定义参数 $z = x + y$, 则价值函数可定义为

$$V(t,z) = \sup_{(C,L,M)\in\mathcal{A}(t,z)} E\left[\int_t^{\top} e^{-\beta(s-t)}U_1(C_s)ds + e^{-\beta(T-t)}U_2(Z_T)\middle| Z_t = z\right], \tag{3.13}$$

其中 $\mathcal{A}(t,z)$ 为可行集 (C,θ) 中满足对于任意 $t \leqslant s \leqslant T$, $Z_s \geqslant 0$ 的集合. 定义

$$\begin{aligned} p &= \frac{\alpha - r}{(1-\gamma)\sigma^2}, \quad c = \frac{1}{1-\gamma}\left[\beta - \gamma r - \frac{\gamma(\alpha - r)^2}{2(1-\gamma)\sigma^2}\right], \\ C_i(t) &= \frac{c}{1 - \phi_i e^{c(t-T)}}\ (i = 1,2), \quad \phi_1 = 1,\ \phi_2 = 1 - c. \end{aligned} \tag{3.14}$$

Merton (1969) 假设效用函数 U_1, U_2 满足 (3.10) 中 CRRA 形式, 求解随机控制问题 (3.13) 中的偏微分方程 (PDE), 结果表明在最优策略下, 以一个固定的比例 p (称为 Merton 比例) 投资于股票中, 而消费额应与投资者的总资产成一定比例. 关于解的细节问题在 3.4 节的补充材料 5 中给出. Karatzas 等 (1987) 以及 Cox 和 Huang (1989) 利用鞅方法替代通常的动态规划方法, 对于一般的 U_1 和 U_2 得到了类似的结果.

3.3.2 包含了交易成本的 Merton 问题

Merton (1969) 的问题中并没有考虑交易成本, 且交易是连续进行的, 因此 dL_t 和 dM_t 是 dt 的倍数, 而公式 (3.6)–(3.7) 可以代入到 (3.12) 式中. 此处并没有考虑 (3.6) 式和 (3.7) 式的复杂情形, 其原因在于这些复杂的情形都可以由本节中讨论的比例投资问题拓展而得. 假设投资者将总投资额的比例 $0 \leqslant \lambda < 1$ 和总卖出额的比例 $0 \leqslant \mu < 1$ 用于股票, 则投资者持有的头寸 (X_t, Y_t) 满足:

$$dX_t = (rX_t - C_t)dt - (1+\lambda)dL_t + (1-\mu)dM_t, \tag{3.15}$$

$$dY_t = \alpha Y_t dt + \sigma Y_t dB_t + dL_t - dM_t. \tag{3.16}$$

(3.15) 式中的 $1+\lambda$ 项表明在将 dL_t 投资于股票时, 应该从持有的债券中支付 λdL_t 的交易费用. 类似地, (3.15) 式中的 $1-\mu$ 项表明在卖出 dM_t 的股票时, 也需要支付数额为 μdM_t 的交易费用. 定义投资者的净资产为

$$Z_t = X_t + (1-\mu)Y_t \text{ 当 } Y_t \geqslant 0; \quad Z_t = X_t + (1+\lambda)Y_t \text{ 当 } Y_t < 0.$$

在考虑交易成本后, (3.8) 式相应地变成

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$$\partial_\lambda \mathcal{D} = \{(x,y) : x > 0,\ y < 0\ ,\ x + (1+\lambda)y = 0\},$$
$$\partial_\mu \mathcal{D} = \{(x,y) : x \leqslant 0,\ y \geqslant 0\ ,\ x + (1-\mu)y = 0\}.$$

在加入交易成本的修正后, 优化问题的目标依旧是寻找能够最大化 (3.9) 式中定义的期望效用 $J(0,x,y)$ 的 $(C,L,M) \in \mathcal{A}(t,x,y)$. 这个问题是一个奇异型随机控制问题, 这类问题中会存在一个不进行交易的集合, 其边界由买入边界 $\mathcal{B}_t$ 和卖出边界 $\mathcal{S}_t$ 来决定. 在最优策略中. 当股票价格 S_t 在买入边界 $\mathcal{B}_t$ 上或边界之下时, 立即买入股票, 而在价格满足 $S_t \geqslant \mathcal{S}_t$ 时立即卖出股票. 有关奇异型随机控制以及买入边界和突出边界的细节在 3.4 节的补充材料 6 中给出. 买入边界和卖出边界通常并没有解析解, Gennotte 和 Jung (1994)、Zakamouline (2005)、Dai 和 Zhong (2010)、Chen 和 Dai (2013) 以及 Palczewski 等 (2015) 对此问题给出了数值解. 在无限期的情况下, 即当 $(T = \infty)$ 时, 买入边界和卖出边界与时间 t 无关, 因此该问题会更容易求解. Magill 和 Constantinides (1976) 使用启发式算法求解此问题, 结果表明并不存在 "楔形的" 不发生交易区域. Davis 和 Norman (1990) 随后利用平滑拟合的准则和求解 "楔形" 区域的常微分方程的方法证明了 Magill 和 Constantinides (1976) 的问题. Akian 等 (1996)、Leland (2000) 和 Liu (2004) 将该问题拓展到包含多种风险资产的情况.

3.3.3 多期资本增长和波动率脉动

自 Kelly (1956)、Latane (1959) 和 Breiman (1961) 以来, 在多期模型和对数效用函数的设定下, 3.3.1 节中提到的 "终身投资组合选择" 问题有了新方法. 新方法在满足 $U_1 = 0$ 和 $U_2(c) = \log c$ 的情况下, 解决了离散情况下的 (3.13) 问题. 定义 Z_t 为时间 $t = 0, 1, \ldots, T$ 时的资本 (由于此处不存在消费, 因此资本等于财产). 定义第 i 个资产的价格为 P_t^i, 定义 $\boldsymbol{w}_t = (w_t^1, \ldots, w_t^m)^\intercal$ 为基于直到 t 时刻的信息集 $\mathcal{F}_t$ 决策的下一期的投资组合的权重向量. 定义 $\mathcal{R}_t^i = \dfrac{P_{t+1}^i}{P_t^i}$ 为下一期的总收益, 则有 $Z_{t+1} = Z_t \boldsymbol{w}_t^\intercal \mathbf{R}_t$, 其中 $\mathbf{R}_t = (\mathcal{R}_t^1, \ldots, \mathcal{R}_t^m)^\intercal$. Breiman (1961) 假设 $\mathbf{R}_t$ 独立同分布, 求解得到对数最优策略, 结果表明当满足在时间 t 投资组合的权重向量为

$$\boldsymbol{w}_t^* = \arg\max_{\boldsymbol{w}} E\left[\log(\boldsymbol{w}^\intercal \mathbf{R}_t) | \mathcal{F}_t\right] \tag{3.17}$$

的情况下, 在 $T \to \infty$ 时的资本增长率 $\dfrac{\log Z_T}{T}$ 近似最大. Algoet 和 Cover (1988) 证明该结果对于任意的 $\{\mathbf{R}_t,\ 0 \leqslant t < T\}$ 的联合分布都成立, 原因在于

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

对数最优策略的总资产 Z_t^* 满足如下的上鞅特性: 对于任意的多期投资组合选择策略的总资产 Z_t, $\{Z_t/Z_t^*, t \geqslant 0\}$ 都是一个非负的上鞅过程, 且

$$\limsup_{T\to\infty} T^{-1}\log\frac{Z_T}{Z_T^*} \leqslant 0 \text{ 以概率 1 成立.} \tag{3.18}$$

在 3.4 节的补充材料 8 中, 我们对于对数最优策略和信息理论的背景进行了进一步讨论.

Samuelson (1969, 245–246 页) 研究发现, 尽管 “从引理到定理的逻辑推导过程没有问题”, 但是使用此处的引理, 即在每一阶段都能最大化总收益 (从 (3.17) 式意义上来说) 的几何平均值的投资组合会大概率地相比于其他的投资组合策略的资产终值更高, 却 “在逻辑推断的开始就是错误的”, 因为此处只会在对数效用函数的条件下获得更高的期望效用, 在其他情况下并不成立. 在忽略优化问题的情况下, Dempster 等 (2007) 证明非时变的权重向量 $\boldsymbol{w} \geqslant \mathbf{0}$ (不允许卖空的情况下) 刻画了一个常数比例 (也称为固定混合) 策略, 当满足随机向量 $\mathbf{R}_t$, $t = 1, 2, \ldots$, 非退化、平衡且遍历时, 资本增长率 $\lim\limits_{T\to\infty} T^{-1}\log Z_T$ 以概率 1 存在. 此外, 资本增长率几乎处处高于 $(\rho_1, \ldots, \rho_m)\boldsymbol{w}$, 其中 $\rho_i = \lim\limits_{T\to\infty} T^{-1}\log\left(\prod_{t=1}^{\top}\mathcal{R}_t^i\right)$ 是第 i 个资产的增长率. 当非退化的随机变量 W 存在不确定性时 (或方差为正时), 常常使用 Jensen 不等式 $E(\log W) > \log(EW)$ 来解释 “波动率驱动的” 超额收益率 (也称为 “波动抽利”). Fernholz 和 Shay (1982) 以及 Luenberger (2013) 进一步表明, 在存在 $m = 2$ 个风险资产, 且价格过程为扩散过程时, 常数比例策略不仅具有波动率驱动的增长率, 同时还可以使投资组合的波动率下降. Dempster 等 (2007) 表明此处的波动率的下降并不是一直成立的, 并且给出了其成立的条件. 论文同时表明当投资再平衡中存在比例化的交易成本时, 只要交易成本足够小, 波动率驱动的超额收益率依旧存在.

3.3.4 多期均值–方差投资组合再平衡

即使在考虑交易成本的情况下, 在前面章节中提到的多期期望效用最大化和资本增长方面的文献, 依旧忽略了投资基金中积极型 (或消极型) 投资组合经理的实际问题, 即基金表现是利用 1.4 节和 3.1.3 节中提到的一些风险调整后的测度来进行周期性评估的. 因此, 大多数积极型投资经理仍然更倾向于使用 Markowitz 框架下的均值–方差优化模型. 此外, Levy 和 Markowitz (1979) 以及 Kroll 等 (1984) 讨论了如何使用均值和方差的函数来估计期望效用, 并应

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

用到投资组合选择中. 然而, 如何在多期框架下进行均值–方差投资组合的再平衡始终是一个难题. Grinold 和 Kahn (2000) 一书的第 21 章将此列为一个开放性问题, 书中指出 "积极型投资组合管理是一个动态的问题, ····· 在合理的框架下, 投资组合经理应该基于资产的动态变化, 分析现在情况和未来情况的相互作用, 然后在当下做出投资决策. 但这样的决策的一个基本的开放性问题是, (存在交易成本时) 我们应该在何时进行交易呢?"

Merton (1990) 讨论了连续时间情况下的均值–方差优化问题, Zhou 和 Li (2000) 针对该问题提出了一个线性二次规划 (LQ) 随机控制框架. Pliska (1997) 以及 Li 和 Ng (2000) 讨论了离散化形式多期问题. 然而, 由于动态规划中的 "维度诅咒" 的存在, 该方法只适用于几种特定的资产, 因此该方法在实际中很少使用, 具体可参见 Kritzman 等 (2007). 相反地, 在时刻变化的世界中, 将单期均值–方差分析方法拓展到多期情形的启发式算法是目前常用的方法, 该方法也可以扩展到更高维的情况. 特别地, Markowitz 和 van Dijk (2003) 提出应该使用 (a) "均值–方差替代" 的方法: 使用投资组合收益的均值和方差的线性组合来替代无限时间设定下基于当前信息的条件期望效用; (b) 近似动态规划 (ADP) 方法. 尽管 Markowitz 和 van Dijk 并没有使用 ADP 这个术语, 而是使用一个具体的例子来诠释其观点, 但论文中的求解过程实际上就是广义随机优化技术 ADP 的一个实例. 具体细节在 3.4 节的补充材料 7 中提供, 在该节中, 也对于 "均值方差替代" 方法及其与 ADP 方法的综合运用进行了讨论. Kritzman 等 (2007) 在现实的场景中进行了模拟计算, 在 Markowitz-van-Dijk (MvD) 方法和动态规划 (DP) 方法都可行的情况和一些标准业界探索情况下, 比较了两种方法结果的区别. 研究结果表明, 当 DP 方法的数值结果达到可信的准确度时, MvD 方法与 DP 方法的结果 "显著地接近", 而在标准业界探索情况下, MvD 方法的结果远远优于 DP 方法的结果. 论文同时重点提到了 MvD 论文标题中的 "时刻变化的世界" 的问题, 即 "机构投资者通常会使用均值–方差分析方法来确定最优投资组合权重. 由于世界始终在变化, 在投资策略执行后, 随着资产价格的变化, 投资组合与最优投资目标发生偏离, 最优权重立即变成次优的."

3.3.5 存在交易成本情况下的动态均值–方差投资组合优化

正如 Grinold (2005) 指出, 在忽略交易成本的情况下, 实施成本是采用投资策略的一个阻碍. Kroner 和 Sultan (1993)、Perold 和 Sharpe (1995) 以及 Engle 等 (1998) 针对此问题提出了相关的解决方案. 在 Markowitz 和 van

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

Dijk (2003) 以及 Kritzman 等 (2007) 中, 在 MvD 启发式算法中, 使用了如下的线性交易成本

$$C_t = \sum_{j=1}^{m} \kappa_j |w_t^j - w_{t-1}^j|, \tag{3.19}$$

其中 $\kappa_1, \ldots, \kappa_m$ 是既定的常数, 且投资组合中不允许进行卖空. Gârleanu 和 Pedersen (2013) 使用了更易于处理的交易成本方程, 即使用 $\Delta \boldsymbol{u}_t = \boldsymbol{u}_t - \boldsymbol{u}_{t-1}$ 的二次函数, 其中 u_t^j 是在 t 时刻投资组合中第 j 只股票的数量. 文中假设下一期相对于无风险利率 $\boldsymbol{R}_f$ 的超额收益率 $\tilde{\boldsymbol{R}}_{t+1}$ 与当前股票价格的乘积满足如下多因子模型:

$$\tilde{\boldsymbol{R}}_{t+1} = \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_{t+1}, \tag{3.20}$$

因此 $\tilde{\boldsymbol{R}}_{t+1}$ 的各个元素为 $\tilde{R}_{t+1}^j = P_t^j(R_{t+1}^j - R_f)$, $j = 1, \ldots, m$, 其中 t 时刻的因子 $\boldsymbol{f}_t$ 可观察到, 且误差项 $\boldsymbol{\varepsilon}_{t+1}$ 的协方差矩阵 $\boldsymbol{\Sigma}$ 已知. 假设这些因子由一个向量自回归模型决定, 模型的形式为 $\boldsymbol{f}_{t+1} = (\boldsymbol{I} - \boldsymbol{\Phi})\boldsymbol{f}_t + \boldsymbol{w}_{t+1}$. 论文中定义的优化模型是选择 $\boldsymbol{u}_t$ 来最大化收益率的现值

$$E\sum_{t=1}^{\infty}(1-\rho)^t \left[(1-\rho)(\boldsymbol{u}_t^{\intercal}\tilde{\boldsymbol{R}}_{t+1} - \gamma \boldsymbol{u}_t^{\intercal}\boldsymbol{\Sigma}\boldsymbol{u}_t) - \frac{1}{2}(\Delta \boldsymbol{u}_t)^{\intercal}\boldsymbol{\Lambda}\Delta \boldsymbol{u}_t\right], \tag{3.21}$$

其中交易成本矩阵 $\boldsymbol{\Lambda}$ 与风险矩阵 $\boldsymbol{\Sigma}$ 成比例, 比例为常数 $\lambda > 0$, 论文的附录 B 对此假设进行了定义. 在无限时间问题 (3.21) 中, 参数 $0 < \rho < 1$ 是折现因子, 参数 γ 可以定义为存在风险约束时的 Lagrange 乘子或投资者的风险厌恶系数. 该问题可以看作时间 t 时, 状态 $\boldsymbol{f}_t$ 和控制 $\Delta \boldsymbol{u}_t$ 下的马尔可夫决策问题. 在线性动态条件和成本为包含 $\boldsymbol{u}_t^{\intercal}\boldsymbol{\Sigma}\boldsymbol{u}_t$ 和 $(\Delta \boldsymbol{u}_t)^{\intercal}\boldsymbol{\Lambda}\Delta \boldsymbol{u}_t$ 的二次项时, 问题 (3.21) 可以直接使用动态规划方法来求解. 定义 $\bar{\rho} = 1 - \rho$, 则论文中命题 1 为价值方程的一个解析形式, 命题 2–4 则给出了最优策略. 3.4 节的补充材料 9 对此策略进行了讨论, 并讨论了其与随机控制中经典的 LQG 问题的联系. 论文利用该方法, 并使用样本估计值替代未知的参数, 然后选用了 1996 年 1 月 1 日–2009 年 1 月 23 日期间 15 种不同的商品期货进行了实证分析. 在分析时, 选用的因子包括 (a) 过去五天每个商品期货的价格变化的均值与标准差的比值向量; (b) 过去一年的价格变化向量; (c) 过去五年的价格变化向量. 至于转移矩阵 $\boldsymbol{\Lambda} = \lambda\boldsymbol{\Sigma}$, 文中使用了 Engle 等 (2012) 提出的对 λ 的估计和 $\boldsymbol{\Sigma}$ 的压缩估计量.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

3.3.6 参数不确定性情况下的动态投资组合选择

在第二章中, 我们讨论了资产回报率的均值和方差已知的情况下单期 Markowitz 优化问题, 在实际应用时, 使用样本的均值和方差的估计值来代替实际值. 第二章同时对于现有文献中解决该问题的不同方法及其难点进行分析, 特别是 Lai 等 (2011b, LXC) 提出的方法, 该方法将未知的参数作为贝叶斯框架或实证贝叶斯框架下的随机变量. 对于动态 (连续时间) 投资组合优化问题, Williams (1977)、Detemple (1986)、Dothan 和 Feldman (1986)、Gennotte (1986) 以及 Brennan (1998) 引入了序贯学习方法, 具体参见 3.4 节的补充材料 9. 该方法将未知的漂移项作为状态变量, 利用滤波方法逐步地进行求解, 然后将其代入 Merton (1971) 和其他假定资产价格的扩散过程已知的模型中.

特别地, Brennan (1998) 对于同时投资于一项无风险资产和一项风险资产的问题进行研究. 假定无风险资产的瞬时收益率为 r, 风险资产的价格服从几何布朗运动 $\dfrac{dP_t}{P_t} = \mu dt + \sigma dB_t$, 其中 σ 已知, 而 μ 未知. 假定 μ 为状态变量, 其先验分布为 $N(m_0, v_0)$. 作者发现投资者可以利用后验分布 $N(m_t, v_t)$, 利用序贯学习的方法来预测 μ, 其中 m_t 项由如下的 Kalman-Bucy 滤波方法求得:

$$dm_t = \frac{v_t}{\sigma^2}\left(\frac{dP_t}{P_t} - m_t dt\right), \quad \text{其中 } v_t = \frac{v_0\sigma^2}{v_0 t + \sigma^2}; \tag{3.22}$$

具体可参见 3.4 节的补充材料 9, 该节也给出了利用动态规划方程来选择 w_t $(0 \leqslant t \leqslant T)$, 以最小化投资者资产终值 Z_T 的期望效用的方法. 论文将该方法应用于实证研究中, 在 1926—1994 年这 69 年的时间段内, 选取 S&P 500 指数作为风险资产, 假定 $r = 5\%$, $m_0 = 13\%$, $\sqrt{v_0} = 0.0243$ 或 0.0452, 以及 $\sigma = 20.2\%$ (69 年间的年收益的标准差) 或 $\sigma = 14\%$ (相对而言 "更切合最近的经验" 的取值). 论文对于指数效用函数的动态规划方程使用有限期差分方法进行近似, 效用函数为 $U(z) = \dfrac{z^\gamma}{\gamma}$, 其中 $\gamma < 1, \gamma \neq 0$. 实证结果表明: 参数的不完全信息可能会导致两种不同的现象, 一是 "预测风险", 二是 "学习". 他指出收益率的期望的不确定性会增加风险, 且使得投资在风险资产中的比例发生一定的减少, 但是随着更多的收益率被观测到, 通过 "学习" 可以进一步了解均值参数的实际值, 增加用于对冲风险资产的比例.

Xia (2001) 在投资收益率是可预测且时变的情况下, 分析了预测风险和学习的影响. Xia (2001) 在投资收益率可预测且时变的情况下, 在一个更一般的框架下分析了预测风险和学习的影响, 指出 "与参数的不确定性相关的对冲需求在最优策略中起重要作用", 即 "与时间相关: 由于参数的不确定性产生了一

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

个与状态相关的对冲需求, 该需求可能随时间增加或减少, 因此最优股票配置在时间上可能增加、减少或非单调变化." 论文没有使用 Brennan (1998) 中所使用的资产价格的几何布朗运动模型, 而是使用了 $\frac{dP_t}{P_t} = \mu_t dt + \sigma dB_t$ 形式的定价模型, 其中 $\mu_t = \alpha + \boldsymbol{\beta}^\top \boldsymbol{f}_t$, $\boldsymbol{f}_t$ 是一个 $k \times 1$ 的可观测的因子向量, 服从 $d\boldsymbol{f}_t = (\boldsymbol{A}_{0,t} + \boldsymbol{A}_{1,t}\boldsymbol{\beta})dt + \boldsymbol{\Sigma}_t^{1/2} d\tilde{\boldsymbol{B}}_t$ 的线性随机动态过程, 其中 $\boldsymbol{A}_{1,t}$ 和 $\boldsymbol{\Sigma}_t$ 是已知的 $k \times k$ 矩阵, $\boldsymbol{A}_{0,t}$ 是一个 $k \times 1$ 向量, 且是 $(\boldsymbol{P}_t, \boldsymbol{f}_t)$ 的连续函数, $\tilde{\boldsymbol{B}}_t$ 是一个 k 维布朗运动过程, 且与 B_t 独立. 假定 $(\alpha, \boldsymbol{\beta}^\top)^\top$ 的先验分布为正态分布, 则在给定 $(\boldsymbol{P}_s, \boldsymbol{f}_s)(0 \leqslant s \leqslant t)$ 情况下的后验分布也是正态分布, 可由 Kalman-Bucy 滤波方法求得. 见 3.4 节的补充材料 9. Xia (2001) 的研究结果表明 "最优投资组合配置和预测变量的关系很大程度上取决于未来的学习和投资."

Cvitanić 等 (2006) 在 Brennan 的工作的基础上, 将其拓展到 m 个风险资产加上一个市场投资组合 (或者说基准投资组合), 价格过程 $P_{0,t}$ 满足:

$$\frac{dP_{0,t}}{P_{0,t}} = \mu_0 dt + \sigma_0 dB_t^0, \quad \frac{dP_{i,t}}{P_{i,t}} = \mu_i dt + \beta_i \sigma_0 dB_t^0 + \sigma_i dB_t^i, \tag{3.23}$$

$i = 1, \ldots, m$, 且 $\mu_i - r = \alpha_i + \beta_i(\mu_0 - r)$, 其中 $\mu_i - r$ 指相对于无风险资产的平均超额收益率. 假定 $\sigma_0, \ldots, \sigma_m$ 已知, 定义

$$\theta_0 = (\mu_0 - r)/\sigma_0, \quad \theta_i = \alpha_i/\sigma_i \quad (i = 1, \ldots, m), \tag{3.24}$$

假设 $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_m)^\top$ 的先验分布为 $N(\bar{\boldsymbol{\theta}}(0), \boldsymbol{Q}^\top \boldsymbol{D} \boldsymbol{Q})$, 论文使用 Kalman-Bucy 滤波方法求得 $\boldsymbol{\theta}$ 的后验分布为 $N(\bar{\boldsymbol{\theta}}(t), \boldsymbol{Q}^\top \boldsymbol{D}(t) \boldsymbol{Q})$, 其中 $\bar{\boldsymbol{\theta}}(t) = \boldsymbol{Q}^\top \bar{\boldsymbol{D}}(t) \cdot \{\boldsymbol{Q}\boldsymbol{B}^*(t) + (\bar{\boldsymbol{D}}(0))^{-1} \boldsymbol{Q}\bar{\boldsymbol{\theta}}(0)\}$, $\bar{\boldsymbol{D}}(t) = \mathrm{diag}(\delta_0(t), \ldots, \delta_m(t))$, $\delta_i(t)$ 和 $\boldsymbol{B}^*(t)$ 的定义如下. 首先, 注意到 $\boldsymbol{Q}^\top \boldsymbol{D} \boldsymbol{Q}$ 是 $\boldsymbol{\theta}$ 的先验分布的协方差矩阵的奇异值分解, 其中 $\boldsymbol{D} = \mathrm{diag}(d_0, \ldots, d_m)$ 是由特征值 d_i 组成的矩阵, $\boldsymbol{Q}$ 是由相对应的特征向量组成的正交矩阵. 定义 $\delta_i(t) = \dfrac{d_i}{1 + d_i t}$, $i = 0, \ldots, m$, $\boldsymbol{B}^*(t) = \bar{\boldsymbol{B}}(t) + \int_0^t \bar{\boldsymbol{\theta}}(s) ds$, 其中 $\bar{\boldsymbol{B}}(\cdot)$ 是滤波理论中的新息过程 (参见 Cvitanić 等 (2006) 第 1147 页, 特别是其中 (A1) 和 (A2)). 在效用函数为指数效用函数 $U(z) = z^\gamma/\gamma$ (其中 $\gamma < 0$) 或者对数效用函数 $U(z) = \log z$, 即对应于 (3.10) 中 $\gamma = 0$ 的情况时, Cvitanić 等 (2006) 对包含 $m + 1$ 个风险资产的优化问题给出了权重向量的解析解:

$$\boldsymbol{\pi}_t = \boldsymbol{V}^{-1/2} \boldsymbol{Q}^\top \boldsymbol{A}^{-1}(t) \boldsymbol{Q} \bar{\boldsymbol{\theta}}(t), \tag{3.25}$$

其中 $\boldsymbol{V}$ 是 $(B_1^0, B_1^1, \ldots, B_1^m)^\top$ 的协方差矩阵, $\boldsymbol{A}(t)$ 是一个对角矩阵, 对角项为

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$(1-\gamma)-\gamma(T-t)\delta_i(t)$, $i=0,\ldots,m$. 最优投资组合的权重向量 $\boldsymbol{w}_t$ 由子向量 $\boldsymbol{\pi}_t$ 和 $1-\boldsymbol{\pi}_t^\top\mathbf{1}$, 即无风险资产的权重项, 共同组成.

因此, 当价格过程的漂移项存在不确定性时, 滤波方法 (特别是 Kalman-Bucy 滤波方法) 在动态投资组合优化问题中起到重要作用. 然而, 波动率项的不确定性在之前的研究中并没有引起足够的重视, 前人的研究都把这一项作为已知项来处理. 另一方面, 在本书第二章中提到对于波动率的估计和建模在单期投资组合优化理论中的重要作用. 波动率的参数的不确定性没有引起足够重视的主要原因在于, 该方法需要使用非线性滤波方法. 这些方法在近年来被提出, 在 3.4 节的补充材料 10 中对这些方法进行了归纳, 并对于前面章节中提到的一系列建模和优化方法的背景和细节进行了补充. 在近年来的研究中, 这些方法被应用到解决存在交易成本和参数不确定性情况下的多期投资组合优化问题.

3.4 补充材料和注释

1. **SPX, SPY 等指数以及 ETF 作为基准投资组合** SPX 指 S&P 500 指数, 基于美国 500 家大公司的股票数据计算得到. 该指数由美国标准普尔金融服务公司 (S&P) 所有, 该公司是 McGraw-Hill 公司的子公司. 尽管指数中包含 500 家公司, 但是有些时候实际股票数可能多于 500 家, 因为有些公司可能会有多只股票. 由于该指数目标是反映美国经济中不同的部分, 当一个部分已有公司存在于指数时, 另一个满足加入指数所有条件的公司可能并不会加入指数. 该指数每天更新一次, 更新时按照 S&P 道 · 琼斯指数: 指数数学方法, 该方法参见 www.spdji.com 网站, 网站中提到 "最简单的资本化权重指数可以看作是一个包含指数中所有股票的投资组合, 数值通常非常大, 在十亿美元量级, 且需要利用一个合适的因子来进行比例缩放 ····· 为保证该指数的价值 (或水平) 在股票增减时保持不变, 该因子也需要进行调整, 以抵消指数的市值的差异的变化." S&P 指数中的 500 家大型公司被分成 11 个部分. 表 3.2 中列出了这些公司, 且同时给出了 2016 年 10 月 17 日时各部分的权重.

   很多指数基金和交易所交易基金通过持有与 S&P 500 指数中相同的股票的方法对该指数进行追踪, 各股票的权重与指数相同或成比例, 以匹配或提高其收益 (不考虑费用和开支情况下). 该现象也是该指数被用于

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

表 3.2 S&P 500 指数的各部分及其在 2016 年 10 月 17 日的权重

部分	权重	主要公司
信息技术	21.42%	苹果、微软、脸书
医疗	14.36%	强生、辉瑞、默克
金融	13.09%	伯克希尔 B 类、摩根大通、富国银行
可选消费	12.57%	亚马逊、康卡斯特、家得宝
基本消费	9.94%	宝洁、可口可乐、百事可乐
工业	9.71%	通用电气、3M、霍尼维尔
能源	7.26%	埃克森美孚、雪佛龙、斯伦贝谢
公共事业	3.23%	美国电力、暴雪、太平洋煤气电力
房地产	2.96%	西蒙地产集团、美国电塔、冠城国际
材料	2.83%	杜邦、陶氏化学、孟山都
通信	2.59%	美国电话电报公司、威瑞森电信、Level 3 通信公司
现金	0.03%	

基准投资组合, 用于评估投资组合经理的业绩的部分原因. 消极型投资组合管理试图追踪该指数, 而积极型投资组合管理则试图取得比指数更高的收益. 由于指数的权重和收益每天更新, S&P 已经积累了大量的数据和分析, 因此是一个非常好的基准.

SPY 是用于追踪 S&P 500 指数的交易所交易基金 (ETF) 的股票代码. 该 ETF 的初始时间为 1993 年 1 月 22 日. 在很长一段时间内, 该基金是世界上最大的 ETF 基金. 它的费用率是最低的之一, 目前的值为 0.09%. SPY 为投资者提供了一个对大型公司分散化投资的机会. S&P 500 是表示美国市场上交易情况最好的代理指标, 而 SPY 是世界上交易量最大的证券. 这么大的流动性使得大笔交易并不昂贵, 这也是期望交易时间较短的交易者和机构喜欢 SPY 之处. 考虑到交易者的投资偏好的不同, 市场上也存在其他的指数和相关联的 ETF. 举例来说, 道 · 琼斯工业指数 (DJIA) 是 S&P 道 · 琼斯指数中的另一个指数. 该指数包含在 NYSE 或 NASDAQ 上交易的最大的 30 家美国公司, 而 DIA 则是相对应的 ETF 基金的股票代码, 该基金在 1998 年提出, 属于美国道富环球投资组合管理公司的 SPDR 基金旗下.

2016 年 10 月 21 日的《华尔街杂志》发表了一篇关于 “积极/消极

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

巨头" 空间基金管理公司 (DFA) 的文章, 指出 "美国快速发展的主要公募基金公司并不是严格的积极或消极投资公司, 它实际上两者皆备." 该公司的创始人是 "指数基金的先行者", 但是 "在很久之前就总结认为遵守有效市场 (或 EMH) 规律的投资者" 利用积极型投资组合管理 "会比单纯的指数追踪基金获得更高的收益". 文章指出 "DFA 公司的投资策略为: 首先建立自己的指数, 通常包含小型公司, 然后等待一段时间, 需要的话会等待数周, 直到市场上有着急的卖方愿意以低于市场一般价的价格卖出股票. 这样的策略可以最小化, 甚至在一些情形下消除交易成本, 因此会得到一个不小的收益. 该公司通常会重点关注价格便宜、主动型交易者不愿购买的小型公司股票." 投资时参照本书 2.9 节补充材料 6 中提到的 Fama 和 French 的研究内容. 文章指出 "根据 DFA 公司的数据, 自从该公司 1981 年下半年成立以来,DFA 的旗舰基金获得了年均 11.8% 的收益率, 包含小型股票的罗素 2000 基金指数的年均收益率为 10.3%. 该公司的 50 只基金中的三分之二自成立以来一直高于相应的基准投资组合的收益率. 平均费用率为三分之一点."

2. **利用全局平滑算法求解 CCQP** Murray 和 Ng (2008) 提出了一个求解离散参数下的非线性优化问题的算法. Shek (2010) 将原问题的框架进行拓展应用到最优投资组合优化问题下. CCQP 问题和连续变量下求解全局最小值这两个问题有一定的等价性 (如参见 Ge 和 Huang (1989)). 然而, 如果只是在优化问题中引入惩罚函数, 并不足以保证问题的可解性, 因为在该情况下可能会产生大量的局部最小值而显著地增加原问题的复杂程度. 全局平滑算法是指在原问题中引入一个合理的惩罚函数的同时, 在原问题的目标函数加上一个严格凸函数. 接下来, 可以通过迭代方法来增加对于非整数值的惩罚函数, 减少平滑的次数.

   考虑如下的 CCQP 问题:

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) = \boldsymbol{c}^{\top}\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^{\top}\boldsymbol{H}\boldsymbol{x} \tag{3.26}$$

$$\text{满足 } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \tag{3.27}$$

$$y_i \geqslant x_i \geqslant 0,\ i = 1, \ldots, n, \tag{3.28}$$

$$\sum_i y_i = K,\ y_i \in \{0, 1\}, \tag{3.29}$$

   其中 $\boldsymbol{c}, \boldsymbol{x} \in \mathbb{R}^{n\times 1}$, $\boldsymbol{H} \in \mathbb{R}^{n\times n}$, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^{m\times 1}$, 以及 $m \leqslant n$. 则该问题可以转化为

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$$\min_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n} F(\boldsymbol{x},\boldsymbol{y}) = f(\boldsymbol{x}) + \underbrace{\frac{\gamma}{2}\sum_i y_i\,(1-y_i)}_{\text{惩罚项}} - \underbrace{\mu\sum_i (\log s_i + \log x_i)}_{\text{平滑项}}$$

$$\text{满足} \quad \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{e}^\top & \boldsymbol{0} \\ -\boldsymbol{I} & \boldsymbol{I} & -\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ s \end{bmatrix} = \begin{bmatrix} \boldsymbol{b} \\ K \\ \boldsymbol{0} \end{bmatrix}, \quad (3.30)$$

其中 $\boldsymbol{e}\in\mathbb{R}^{n\times 1}$ 是一个各项都为 1 的列向量, s 是松弛变量, γ 是非整数项相对应的惩罚函数的系数, μ 是全局平滑项相关的系数. 定义

$$\check{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{e}^\top & \boldsymbol{0} \\ -\boldsymbol{I} & \boldsymbol{I} & -\boldsymbol{I} \end{bmatrix}, \quad \check{\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ s \end{bmatrix}, \qquad \check{\boldsymbol{b}} = \begin{bmatrix} \boldsymbol{b} \\ K \\ \boldsymbol{0} \end{bmatrix},$$

则梯度矩阵和 Hessian 矩阵为

$$\check{\boldsymbol{g}} = \nabla F(\boldsymbol{x},\boldsymbol{y}) = \begin{bmatrix} g_1 - \dfrac{\mu}{x_1} \\ \vdots \\ g_n - \dfrac{\mu}{x_n} \\ \dfrac{\gamma}{2}(1-2y_1) \\ \vdots \\ \dfrac{\gamma}{2}(1-2y_n) \\ -\dfrac{\mu}{s_1} \\ \vdots \\ -\dfrac{\mu}{s_n} \end{bmatrix}, \quad \check{\boldsymbol{H}} = \nabla^2 F(\boldsymbol{x},\boldsymbol{y}) = \begin{bmatrix} \check{\boldsymbol{H}}_{11} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \check{\boldsymbol{H}}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \check{\boldsymbol{H}}_{33} \end{bmatrix},$$

其中

$$\boldsymbol{g} = \boldsymbol{c} + \boldsymbol{H}\boldsymbol{x}, \quad \check{\boldsymbol{H}}_{11} = \boldsymbol{H} + \operatorname{diag}\left(\frac{\mu}{x_i^2}\right),$$

$$\check{\boldsymbol{H}}_{22} = \operatorname{diag}(-\gamma_i), \quad \check{\boldsymbol{H}}_{33} = \operatorname{diag}\left(\frac{\mu}{s_i^2}\right).$$

Shek (2010) 梳理了全局平滑 CCQP 问题的算法的框架, 并对于将共轭梯度法适用到该问题的算法进行了研究.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

3. **利用局部松弛方法求解 CCQP 问题** Murray 和 Shek (2012) 提出了一种求解如下的 CCQP 问题的算法:

$$\min_{\boldsymbol{x}} f(x) = -\boldsymbol{c}^\top \boldsymbol{x} + \lambda \boldsymbol{x}^\top \boldsymbol{H} \boldsymbol{x}$$
$$\text{满足 } \sum_i x_i = 0,\ \sum_i I_{\{x_i \neq 0\}} = K,$$

在求解时, 论文对于不同的资产收益率之间的内在相似性进行分析, 然后定义一个降维的空间, 并将收益率投影聚类到这个空间内, 然后在该空间中求解一系列的松弛问题. 该方法模拟了参数连续问题求解中的线性搜索和信任域方法, 并将该问题进行局部近似, 从而改进每次迭代过程中的解的估计值. Murray 和 Shanbhag (2006, 2007) 使用了局部松弛方法的一种变异来求解基于格点的配电站选址问题, 研究表明论文中提出的方法相对于很多商业解决方案而言, 在数量增加时算法复杂程度的增加更慢. 该算法与前人的算法在很多方面有一定的区别. 首先, 与之前的算法中预先设定一个二维格点空间不同, 该算法将资产收益投资投影到一个基于历史收益的统计特征的多维空间中. 其次, 之前的算法使用的是物理距离, 而该算法将距离测度定义为因子载荷、风险厌恶和期望收益的一个函数. 另外, 考虑到搜索空间和聚类组的大小的基数, 在该算法中提出了一个概率方法, 用于在每次迭代过程中将聚类成员分配到每一个聚类中心处. 该算法的结构可以分为三个关键步骤:

- 识别初始聚类中心的集合 $\mathcal{S}_0$
  - 利用连续截断算法在每一次迭代过程中将权重较低的资产从投资组合中去掉, 直到投资组合中资产的数量等于合理的值.
- 在每一次迭代过程中, 利用算法求解在原始资产集 S 的子集 $\mathcal{S}_0 \subset \mathcal{S}_r \subseteq \mathcal{S}$ 下的二次规划问题, 其中 $\mathcal{S}_r$ 是包含在前面的步骤中识别的 K 个聚类中心中的资产集. 此处求解的二次规划问题包括:
  - *全局松弛的二次规划问题*

$$\min_x \left\{ -\boldsymbol{c}^\top \boldsymbol{x} + \frac{1}{2} \boldsymbol{x}^\top \boldsymbol{H} \boldsymbol{x} \right\}$$
$$\text{满足 } \boldsymbol{e}^\top \boldsymbol{x} = 0,\ \boldsymbol{x} \in \mathcal{S},$$

其中 $\boldsymbol{c}$ 和 $\boldsymbol{H}$ 分别是资产集 $\mathcal{S}$ 的期望收益和协方差矩阵, $\boldsymbol{e}$ 是一个各项都为 1 的列矩阵.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

– 聚类资产的二次规划问题

$$\min_{x}\left\{-\boldsymbol{c}_0^\top\boldsymbol{x}_0+\frac{1}{2}\boldsymbol{x}_0^\top\boldsymbol{H}_0\boldsymbol{x}_0\right\}$$
$$\text{满足 } \boldsymbol{e}^\top\boldsymbol{x}_0=0,\ \boldsymbol{x}_0\in\mathcal{S}_0,$$

其中 $\boldsymbol{c}_0$ 和 $\boldsymbol{H}_0$ 是聚类资产集合 $\mathcal{S}_0$ 的期望收益和协方差矩阵. 则可得最优权重向量 $\boldsymbol{x}_0^*$, 其中 $x_{0,i}^*$ 是权重向量的第 i 个元素, 对应于第 i 个聚类资产 $s_{0,i}$.

– 近邻松弛二次规划问题

$$\min_{\boldsymbol{x}_r}\left\{-\boldsymbol{c}_r^\top\boldsymbol{x}_r+\frac{1}{2}\boldsymbol{x}_r^\top\boldsymbol{H}_r\boldsymbol{x}_r\right\}$$
$$\text{满足 } \boldsymbol{e}_i^\top\boldsymbol{x}_r=x_{0,i}^*,\ i=1,\ldots,K,\ \boldsymbol{x}_r\in\mathcal{S}_r,$$

其中 $\boldsymbol{e}_i$ 是一个列向量, 对应于第 i 个聚类的资产项的值为 1, 其他项的值为 0. $\boldsymbol{c}_r$ 是期望收益向量, $\boldsymbol{H}_r$ 对应于资产的子空间的方差–协方差矩阵.

- 识别新的聚类中心以更新聚类.

4. **动态规划: 离散时间情形** 此处研究一个逐步选择最优控制以最小化 $E\{\sum_{t=1}^{N}c_t(x_t,u_t)+c_N(x_N)\}$ 的问题, 其中 x_t 为状态变量. c_t 是给定的方程, N 是该问题的 "时间跨度". 控制项 u_t 影响未来的状态 $x_{t+1},\ldots,x_N$, 且由包含 $x_t,u_{t-1},x_{t-1},\ldots,u_1,x_1$ 的信息集 $\mathcal{F}_t$ 决定. 其中 $c_t(x_t,u_t)$ 表示即时成本, 而 $c_N(x_N)$ 为终期成本. 在时间 N 并不会施加控制, 因为控制只会影响时间 N 之后的状态. 注意到 u_t 只与 $\mathcal{F}_t$ 有关, 因此我们可以使用全期望公式:

$$E\left[c_t(x_t,u_t)+V_{t+1}\right]=E\left\{E\left[c_t(x_t,u_t)+V_{t+1}|\mathcal{F}_t\right]\right\},\tag{3.31}$$

其中 V_{t+1} 是时间 $t+1$ 的价值函数, 后面会进行相关定义. 全期望公式说明, 可以使用如下的反向归纳算法来选择 u_t, $t=N-1,N-2,\ldots,1$, 并定义初始值为 $V_N=c_N(x_N)$ 的序列 V_t:

$$\begin{aligned}u_t&=\arg\min_{u\in U}\left\{c_t(x_t,u)+E(V_{t+1}|\mathcal{F}_t)\right\},\\V_t&=c_t(x_t,u_t)+E(V_{t+1}|\mathcal{F}_t).\end{aligned}\tag{3.32}$$

当 x_t 是一个可控的马尔可夫链, 即给定 x_s,u_s $(s\leqslant t)$ 时, x_{t+1} 的期望概率只与 x_t,u_t 有关时, 可以得到只由 x_t 决定的最优控制 u_t.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

无限时间折现问题 在有限时间情形下, 令 $N \to \infty$, 则当

$$E\left[\sum_{t=1}^{\infty} c_t(x_t, u_t)\right]$$

存在时, 即为无限期时间情形. 存在折现因子 $0 < \beta < 1$ 时, 可得到折现情况下的问题 $c_t(x,u) = \beta^t c(x,u)$. 当 x_t 为可控马尔可夫链, 状态空间为 S, 满足对于任意 $t \geqslant 0$, $u \in U$, $x \in S$, $A \subset S$, 平稳转移概率为

$$P\{x_{t+1} \in A | x_t = x,\ u_t = u\} = P_{x,u}(A), \tag{3.33}$$

则价值函数

$$V(x) := \sup_{u_1,u_2,\ldots} E\left[\sum_{t=1}^{\infty} \beta^t c(x_t, u_t) \middle| x_0 = x\right]$$

即为如下的动态规划问题的解:

$$V(x) = \inf_{u \in U}\left\{c(x,u) + \beta \int_S V(y) dP_{x,u}(y)\right\}. \tag{3.34}$$

序列 $\boldsymbol{u} = (u_1, u_2, \ldots)$ 称为控制策略. 由 (3.34) 可知, 最优控制策略是平稳的, 即 $u_t = g(x_t)$ 是一个非时变函数.

价值函数和控制策略的迭代 在求解 (3.34) 式中的价值函数 V 和控制策略时, 通常有两种方法. 其中一种称为价值迭代. 首先设置 V 的初始估计值 v_0, 接下来依次得到如下的近似序列:

$$v_{k+1}(x) = \min_{u \in U}\left\{c(x,u) + \beta \int_S v_k(y) dP_{x,u}(y)\right\}.$$

上式右边的最小值 $u = u_k(x)$ 为在第 k 次迭代中对平稳策略的估计值. 另一种方法称为价值迭代, 该方法可以用于状态空间 S 有限的情况. 定义状态空间 S 中各个状态分别为 $1, \ldots, m$, 且有

$$\begin{aligned} \boldsymbol{c}_g &= [c(1, g(1)), \ldots, c(m, g(m))]^\top, \\ \boldsymbol{P}_g &= \left[P_{x,g(x)}(y)\right]_{1 \leqslant x \leqslant m,\, 1 \leqslant y \leqslant m}. \end{aligned}$$

注意到 $\boldsymbol{P}_g$ 是平稳控制策略 g 下可控马尔可夫链的转移矩阵, 其中 g 由向量 $[g(1), \ldots, g(m)]^\top$ 定义. 则对于 g 进行最小化与对于 $\mathbb{R}^m$ 进行最小化是等价的. 考虑到 (3.34) 式的形式, 我们在 $\boldsymbol{x} \in \mathbb{R}^m$ 上定义仿射变换 $T_g(\boldsymbol{x}) = \boldsymbol{c}_g + \beta \boldsymbol{P}_g(\boldsymbol{x})$. 定义 $T(\boldsymbol{x}) = \min\limits_g T_g(\boldsymbol{x})$ 以及 $\boldsymbol{v} =$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$[V(1),\dots,V(m)]^\top$, 则 (3.34) 式可以写成 $\boldsymbol{v}=T(\boldsymbol{v})$. 由上可得到一个如下的策略迭代过程: 定义最优 g^* 的初始估计为 g_0, 在第 k 次迭代中, 先求得 g_k 值, 然后求解线性函数 $\boldsymbol{v}=T_{g_k}(\boldsymbol{v})$, 其解记为 $\boldsymbol{v}_k$. 当满足 $T(\boldsymbol{v}_k)=\boldsymbol{v}_k$ 时, 该次迭代结束, 且定义 $g^*=g_k$. 否则, 求解满足 $T_g(\boldsymbol{v}_k)=T(\boldsymbol{v}_k)$ 的 $(g(1),\dots,g(m))^\top$, 且定义 $g_{k+1}=g$.

最优停时 当控制 u_t 包含是否在时间 t 停止时, 该随机控制问题被称为最优停时问题. 与一般的随时控制问题不同, 停时规则并不会改变 x_t. 在此情况下, 最优策略问题可以简化为求解停时规则 τ. 假设停时前的成本函数为 $c(x_t)$, 停时的时刻为 $h(x_\tau)$, 且 $t>\tau$ 时 $c_t(x_t)=0$, 则此时转移矩阵 (3.33) 的形式如下:

$$P\{x_{t+1}\in A|x_t=x\}=P_x(A)\quad \text{对于 } 0\leqslant t<\tau,\ x\in S, A\subset S.$$

价值函数 $V(x):=\sup\limits_{\tau} E\left[\sum_{t<\tau}c(x_t)+h(x_\tau)|x_0=x\right]$ (其中上确界是对于所有停时规则 τ 而言) 满足如下的动态规划方程:

$$V(x)=\min\left\{h(x),c(x)+\int_S V(y)dP_x(y)\right\}. \tag{3.35}$$

等式 (3.35) 右侧是指当状态 x 可观测且成本为 $h(x)$ 时停止, 否则继续支付成本 $c(x)$ 且继续寻求最优值. 求解 (3.35) 式的最小值指选出期望成本更小的继续或停止策略. 最优停止准则是一个平稳策略, 其形式为 $\tau=\inf\{n\geqslant 0: h(x_t)=V(x_t)\}$.

5. **动态规划: 连续时间情形** 考虑连续时间马尔可夫过程 X_t, 其*无穷小生成元* (即连续时间情形下的转移概率) $\mathcal{A}$ 满足对于任意的一、二阶导数有限的函数 f, 下式成立

$$\mathcal{A}f(x)=\lim_{t\downarrow 0}\frac{1}{t}E[f(X_t)-f(x)],\quad \text{其中 } X_0=x,$$

特别地, 对于由如下的随机微分方程定义的扩散过程:

$$dX_t=\mu(X_t)dt+\sigma(X_t)dB_t, \tag{3.36}$$

其中 $|\mu(x)-\mu(y)|+|\sigma(x)-\sigma(y)|\leqslant K|x-y|$ 以及 $\mu^2(x)+\sigma^2(x)\leqslant K^2(1+|x|^2)$, 其生成元 $\mathcal{L}$ 可以写成如下形式:

$$\mathcal{L}f(x)=\mu(x)\frac{\partial}{\partial x}f(x)+\frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2}f(x). \tag{3.37}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

infoway.io

实时金融行情数据接口

低延迟 • 高可用 • 全覆盖 **专为量化交易设计的行情API**

全市场覆盖 一个接口查所有

股票API
涵盖A股、港股、美股等多个市场。

外汇API
包含56组主要货币对，高频更新。

加密货币API
超过100个币种，全天24小时实时报价。

商品期货API
贵金属、能源天然气、期货指数等

Infoway API是一款专门为量化交易设计的行情接口。支持逐笔成交/订单簿快照等颗粒度，兼容RESTful API和WebSocket，满足高频量化、资产管理系统、行情终端等场景需求。

全域市场一网打尽

每日处理4300万条市场数据流，开发者无需对接多个数据源即可构建多资产策略系统。

简单易接入

我们的API设计注重易用性，支持多种编程语言，确保无论您使用什么开发环境，都能快速将实时行情数据集成到您的应用中。

实时行情接口：www.infoway.io

免费试用！

注册后自动获得试用资格，无需实名认证。（**本活动长期有效**）

联系我们

官网地址：https://infoway.io
客服邮箱：hi@infoway.io



实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

由 Taylor 展开得到

$$f(X_t) \approx f(x) + f'(X_t)(X_t - x) + \frac{1}{2}f''(X_t)(X_t - x)^2, \tag{3.38}$$

一阶求导项为

$$\lim_{t\downarrow 0} \frac{1}{t} E_x \left[f'(X_t)(X_t - x)\right] = f'(x)\mu(x),$$

二阶求导项为

$$\lim_{t\downarrow 0} \frac{1}{t} E_x \left[f''(X_t)(X_t - x)^2\right] = f''(x)\sigma^2(x),$$

则可得 (3.37) 式的生成元, 其中 E_x 为初始状态 $X_0 = x$ 下的期望. 为了模拟控制过程 U_t 对漂移项和波动率的影响, 并将非时齐性引入随机过程中, 我们将原问题推广为如下的控制扩散过程:

$$dX_t = \mu(t, X_t, U_t)dt + \sigma(t, X_t, U_t)dB_t, \tag{3.39}$$

其中 U_t 在 $\mathbb{U}$ 上取值, 且对于 B_t 是不可预测的, 即在 $t' > t$ 时, $B_{t'} - B_t$ 与 $\{U_s, 0 \leqslant s \leqslant t\}$ 是独立的. 则相应的生成元是一个定义在合理的函数 f 的状态空间中的差分算子, 其形式为

$$\begin{aligned} \mathcal{A}^u &= \frac{\partial}{\partial t} + \mu(t,x,u)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(t,x,u)\frac{\partial^2}{\partial x^2} \\ &= \frac{\partial}{\partial t} + \mathcal{L}^u, \end{aligned} \tag{3.40}$$

满足 $\mathcal{A}^u f(t,x)$ 是由 (3.40) 给定的 $\dfrac{\partial f}{\partial t}$, $\dfrac{\partial f}{\partial x}$ 和 $\dfrac{\partial^2 f}{\partial x^2}$ 的线性组合. 特别地, 若对于可测函数 u 满足 $U_t = u(t, X_t)$, 则 U_t 称为马尔可夫控制准则, 而 u 称为马尔可夫控制函数.

无限时间折现问题 定义连续成本函数为 g, 则无限时间下马尔可夫控制准则的折现成本函数为

$$J(x,u) = E_x\left[\int_0^\infty e^{-\gamma s} g(X_s, u(X_s))ds\right].$$

随机控制问题的价值函数为

$$V(x) = \inf_u J(x,u), \tag{3.41}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中下确界是对于所有的马尔可夫控制函数而言. 使用反向归纳方法, $V(x)$ 可以定义为如下的递归形式:

$$V(x)=\inf_u E_x\left[\int_0^t e^{-\gamma s}g(X_s^u,u(t,X_s^u))ds+e^{-\gamma t}V(X_t^u)\right], \tag{3.42}$$

其中 X_s^u 指马尔可夫控制函数 u 相关的控制扩散过程. 对于最优控制函数 u^*,

$$\begin{aligned}V(x)&=E_x\left[\int_0^\infty e^{-\gamma s}g(X_s,u^*(X_s^{u^*}))ds\right]\\&=E_x\left[\int_0^t e^{-\gamma s}g(X_s,u^*(X_s^{u^*}))ds\right]\\&\quad+E_x\left[\int_t^\infty e^{-\gamma s}g(X_s,u^*(X_s))ds\right].\end{aligned}$$

其中第二项可以进一步简化为

$$\begin{aligned}&e^{-\gamma t}E_x\left[\int_t^\infty e^{-\gamma(s-t)}g(X_s,u^*(X_s))ds\right]\\&=e^{-\gamma t}E_x\left\{E\left[\int_0^\infty e^{-\gamma v}g(X_{t+v},u^*(X_{t+v}))dv\middle|X_t\right]\right\}\\&=e^{-\gamma t}E_x\{V(X_t^{u^*})\}.\end{aligned}$$

类似地, 对于 $u\neq u^*$,

$$V(x)\leqslant E\left[\int_0^t e^{-\gamma s}g(X_s^u,u(X_s^u))ds+e^{-\gamma t}V(X_t^u)\right]. \tag{3.43}$$

假设 V 为平滑过程, 将 Itô 公式应用于 $f(t,x)=e^{-\gamma t}V(x)$ 中可得

$$d(e^{-\gamma t}V(X_t^{u^*}))=\left\{e^{-\gamma t}[-\gamma V(X_t^{u^*})+\mathcal{A}^{u^*}V(X_t^{u^*})]\right\}dt+d\xi_t, \tag{3.44}$$

其中 $d\xi_t=e^{\gamma t}\sigma(X_t^{u^*})V'(X_t^{u^*})dB_t$. 因此 $\int_0^t d\xi_s$ 是一个鞅过程, 且 $e^{-\gamma t}\cdot V(X_t^{u^*})$ 等于

$$V(x)+\int_0^t\left\{e^{-\gamma s}\left[-\gamma V(X_s^{u^*})+\mathcal{A}^{u^*}V(X_s^{u^*})\right]\right\}\,ds+\int_0^t d\xi_s.$$

在不等式两侧都加上 $\int_0^t e^{-\gamma s}g(X_s^{u^*},u^*(X_s^{u^*}))ds$, 并取期望, 则有

$$E_x\left[\int_0^t\left\{e^{-\gamma s}\left[-\gamma V(X_s^{u^*})+\mathcal{A}^{u^*}V(X_s^{u*})+g(X_s^{u^*},u^*(X_s^{u^*}))\right]\right\}ds\right]=0. \tag{3.45}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

令 $t \downarrow 0$, 则有

$$-\gamma V(x) + \mathcal{A}^{u^*} V(x) + g(x, u^*(x)) = 0. \tag{3.46}$$

对于任意 $u \neq u^*$, 类似可得

$$-\gamma V(x) + \mathcal{A}^{u} V(x) + g(x, u(x)) \geqslant 0. \tag{3.47}$$

结合 (3.46) 式和 (3.47) 式, 则可得

$$\inf_u \{-\gamma V(x) + \mathcal{A}^u V(x) + g(x, u(x)\} = 0, \tag{3.48}$$

其中下确界是对所有马尔可夫控制函数而言. 非线性常微分方程 (3.48) 即为随机控制问题中的 Hamilton-Jacobi-Bellman (HJB) 方程.

有限时间情形下的 HJB 方程 在有限时间情形下, 对于一个不可预测的控制准则 U, 其成本函数为

$$J(t, x, U) = E\left[\int_t^T g(s, X_s, U_s)ds + h(X_T)\middle| X_t = x\right], \tag{3.49}$$

其中 g 是连续成本函数, h 是终时成本函数, X_t 为(3.39)式的控制扩散过程. HJB 方程为

$$0 = \inf_u \{\mathcal{A}^u V(t, x) + g(t, x, u)\}, \quad 0 \leqslant t < T, \tag{3.50}$$

其中 $\mathcal{A}^u$ 由 (3.40)式定义, 且边界条件为

$$V(T, x) = h(x); \tag{3.51}$$

具体参见 Øksendal (2003) 以及 Karatzas 和 Shreve (1991).

为具体阐述上述结果, 考虑 3.3.1 节中最大化 (3.13) 式的 Merton 问题, 其中总资产过程满足随机微分方程 (3.12). 价值函数的 HJB 方程[①]为

$$\max_{C,\theta} \left\{\left(\frac{\partial}{\partial t} + \mathcal{L}\right) V(t, z) + U_1(C) - \beta V(t, z)\right\} = 0, \tag{3.52}$$

① 7.3.2 节会对于 HJB 方程及其对动态规划准则的 "无穷小" 情况的阐述进行进一步讨论. HJB 方程假设价值函数对于 t 和 x 充分平滑. 对于此假设放宽时, 会导致 HJB 方程的 "黏性解". 当状态空间为离散空间, 即附录 B 中的连续时间马尔可夫链情形或者 (3.34) 的离散时间情形时, 价值函数同时满足一个类似的方程, 称为 "Bellman 方程".

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

满足终值条件 $V(T,z)=U_2(z)$, 其中 $\mathcal{L}$ 为(3.12)式的最小生成元:

$$\mathcal{L}=\frac{\sigma^2\theta^2z^2}{2}\frac{\partial^2}{\partial z^2}+[rz+(\alpha-r)\theta z-C]\frac{\partial}{\partial z}.$$

对于 C 和 θ 最大化, 则可得 $C=(U_1')^{-1}(V_z)$ 和 $\theta=-\left(\frac{V_z}{V_{zz}}\right)\frac{\alpha-r}{\sigma^2}z$ (其中下标表示偏微分, 如 $V_z=\frac{\partial V}{\partial z}$). 将 C 和 θ 的值代入(3.52), 则可得如下偏微分方程:

$$\frac{\partial V}{\partial t}-\frac{(\alpha-r)^2}{2\sigma^2}\frac{(\partial V/\partial z)^2}{\partial^2V/\partial z^2}+(rz-C^*)\frac{\partial V}{\partial z}+U_1(C^*)-\beta V=0, \quad (3.53)$$

其中 $C^*=C^*(t,z)=(U_1')^{-1}[V_z(t,z)]$.

6. **奇异随机控制和交易成本** HJB 方程可以扩展应用到奇异随机控制问题, 如 3.3.2 节中讨论的存在交易成本时的 Merton 问题, 在本节中会对此进行具体讨论. 偿付能力区域可以分为三个部分, 分别对应于 "买入股票"($\mathcal{B}$)、"卖出股票"($\mathcal{S}$) 和 "不发生交易"($\mathcal{N}$). 从区域 $\mathcal{B}$ 到买入区域边界 $\partial\mathcal{B}$, 或者从区域 $\mathcal{S}$ 到卖出区域边界 $\partial\mathcal{S}$ 的瞬时转移发生时, 相对应地, 投资组合的转移为 $\partial_\lambda\mathcal{D}$ 或 $\partial_\mu\mathcal{D}$ (即转移方向为 $(-1,(1+\lambda)^{-1})^\top$ 或 $(1,-(1-\mu)^{-1})^\top$). 由此可知, 对于 $(t,x,y)\in\mathcal{S}$, 价值函数满足 $V(t,x,y)=V(t,x+(1-\mu)\delta y,y-\delta y)$, 而对于 $(t,x,y)\in\mathcal{B}$, 价值函数满足 $V(t,x,y)=V(t,x-(1+\lambda)\delta y,y+\delta y)$, 因此当 $\delta y\to0$ 时,

$$V_y(t,x,y)=(1-\mu)V_x(t,x,y), \qquad (t,x,y)\in\mathcal{S}, \quad (3.54)$$

$$V_y(t,x,y)=(1+\lambda)V_x(t,x,y), \qquad (t,x,y)\in\mathcal{B}. \quad (3.55)$$

在 $\mathcal{N}$ 中, 价值函数满足 (3.53) 式. 7.2.3 节考虑了与做市策略相关的另一种奇异随机控制问题, 其中像 (3.53) 这样的 HJB 方程被变分或者拟变分不等式取代.

无限时间且包含平稳策略的情形 为简化计算, 考虑如下的无限时间 $T=\infty$ 情形, 满足价值函数和买入/卖出边界与 t 无关. 当效用函数为指数形式 $U(z)=\frac{z^\gamma}{\gamma}$, 且 $\gamma<1$, $\gamma\neq0$ 时, 利用同位相似性 $V(x,y)=y^\gamma\psi\left(\frac{x}{y}\right)$ (其中 $\psi(x)=V(x,1)$), HJB 方程可以简化为只包含一个状态变量的常微分方程 (ODE) 形式. 此时(3.53)–(3.55)式可以简化为如下 ODE 形式:

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$$b_3x^2\frac{d^2\psi}{dx^2}+b_2x\frac{d\psi}{dx}+\frac{1-\gamma}{\gamma}\left(\frac{d\psi}{dx}\right)^{\gamma/(\gamma-1)}+b_1\psi=0,\quad x_*<x<x^*, \tag{3.56}$$

$$\psi(x)=\begin{cases}\gamma^{-1}A_*(x+1-\mu)^\gamma, & x\leqslant x_*,\\ \gamma^{-1}A^*(x+1+\lambda)^\gamma, & x\geqslant x^*,\end{cases} \tag{3.57}$$

其中

$$b_1=\sigma^2\gamma(\gamma-1)/2+\alpha\gamma-\beta,\quad b_2=\sigma^2(1-\gamma)+r-\alpha,\ b_3=\frac{\sigma^2}{2}. \tag{3.58}$$

平滑拟合准则 平滑拟合准则与扩散过程相关的最优停时和奇异随机控制模型的凸 C^2-解的存在性和独立性相关, 具体参见 Benes 等 (1980)、Karatzas (1983) 和 Ma (1992). 该准则通过匹配边界上的一阶导数, 确定这些问题中的 "自由边界". Davis 和 Norman (1990) 为了求解 (3.56) 中的常数 x_*, x^*, A_*, A^* 和函数 ψ, 在 x_* and x^* 点将平滑拟合准则应用到 ψ'' 中, 以求解 A_* 和 A^* (这两项分别由 x_* 和 x^* 决定), 以及 x_* 和 x^*.

7. **近似动态规划问题 (ADP)** 尽管动态规划方法 (DP) 提供了一种随机优化的标准工具, 以最小化随时间变化的瞬时成本和最终成本, 但由于其计算复杂性随着状态空间维度的增加而指数级增加, 在实际中的应用受到一定的局限, 这个现象也被称为 "维度诅咒" (Bellman, 1957, ix 页). 举例来说, Åström (1983, 478 页) 在动态规划问题中利用反向归纳算法求解回归模型 $y_t=y_{t-1}+\beta u_t+\varepsilon_t$ 中最小化

$$\int E_\beta\left[\sum_{t=2}^{T}(y_t-y_t^*)^2\right]d\pi(\beta) \tag{3.59}$$

的最优策略, 其中 $\varepsilon_t\sim N(0,1)$, $y_t^*\equiv 0$, 而 π 是 β 的先验正态分布. 在 $T=30$ 时, 经过 180CPU 小时的数值计算, 当 β 的贝叶斯估计值 β_{t-1} 的精确度较低时, 制定最优策略需要相对较大且不规则的控制项 u_t 才能求解. 由于 DP 的计算复杂性, 目前并没有数值方法能够对于复杂的最优策略的贝叶斯风险 (3.59) 作计算, 从而与短视或确定性等价准则下的风险进行比较. 另外, 与评估贝叶斯风险不同, 频率论学派学者希望能对于不同的 α 和 β 值下的风险方程进行分析.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

Han 等 (2006) 指出 “受到交通和通信网络中的大型数值优化问题, 以及用于西洋双陆棋和纸牌游戏的计算机算法驱动, 机器学习/人工智能领域的相关文献中” 涉及的 ADP 领域的重要发展, 可以进行修改以用于 (3.59) 等回归模型的适应性控制研究. 论文中描述了如何应用起源于西洋双陆棋的 Rollout 算法, 以求解回归模型的适用性控制的近似最优策略问题.

基于 Rollout 算法的近似策略迭代 首先, 考虑有限状态空间 $\mathcal{X}$ 下的控制马尔可夫链 $\{X_t\}$, 其 t 时刻的控制 u_t 属于有限集 U. 令对应于控制 $u \in U$ 的转移矩阵为 $[p_{xy}(u)]_{x,y\in\mathcal{X}}$. 则无限空间下的马尔可夫决策问题是一个选择一系列的控制 $\boldsymbol{u} = (u_1, u_2, \dots)$ 来最小化总期望

$$c_{\boldsymbol{u}}(x) = E\left[\sum_{t=1}^{\infty} \delta^{t-1} g(X_t, u_t) \middle| X_0 = x\right] \tag{3.60}$$

的优化问题, 其中 $0 < \delta < 1$ 是折现因子, g 是成本函数. 对于所有控制策略 $\boldsymbol{u}$, $c_{\boldsymbol{u}}(x)$ 的最小值对应的 $V(x)$ 满足如下的 Bellman 方程:

$$V(x) = \inf_{u\in U}\left\{g(x,u) + \delta \sum_{y\in\mathcal{X}} p_{xy}(u) V(y)\right\}. \tag{3.61}$$

$c_{\boldsymbol{u}}(x)$ 的最小值通过平稳策略 $u_t = \mu(x_t)$ 来求解, 其中 $\mu(x)$ 是(3.61)式右边项的最小值 u. DP 中的策略迭代方法在第 k 次迭代时, 共进行两个步骤. 一是进行 “策略评估”, 计算平稳策略 (与 $\mu^{(k)} : \mathcal{X} \to U$ 相关的策略) 的成本 $c_{\mu^{(k)}}(x)$; 二是进行 “策略改进”, 定义 $\mu^{(k+1)}(x)$ 为下式的极小值 u:

$$g(x,u) + \delta \sum_{y\in\mathcal{X}} p_{xy}(u) c_{\mu^{(k)}}(y). \tag{3.62}$$

代价函数 c_μ 可以通过求解如下线性系统得到:

$$c_\mu(x) = g(x, \mu(x)) + \delta \sum_{y\in\mathcal{X}} p_{xy}(\mu(x)) c_\mu(y), \quad x, y \in \mathcal{X}. \tag{3.63}$$

注意到在策略改进步骤中, (3.62) 式可以写成

$$g(x,u) + E\left[\sum_{t=2}^{\infty} \delta^{t-1} g(X_t, u_t^{(k)}) \middle| X_1 = x,\ u_1 = u\right]. \tag{3.64}$$

当状态空间 $\mathcal{X}$ 很大时, 求解线性系统 (3.63) 对数值计算要求非常高. 同时, 存储平稳策略的 $\mu(x)$, $x \in \mathcal{X}$ 数据也非常困难. 除了求解 (3.63) 式之

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

外, 也可以利用蒙特卡罗模拟方法来计算平稳策略 $\boldsymbol{u}$ 执行期间涉及的状态的代价函数, 就像在 "Rollout" 算法中一样. "Rollout" 策略是指最小化 $k=0$ 时 (3.64) 式的单次策略改进迭代过程, 在计算时对于 $u_t^{(0)}$ 使用启发式策略. "迭代 Rollout" 策略包含 $k\geqslant 1$ 时的迭代过程, 其中 $k=1$ 指对启发式策略的 Rollout, $k=2$ 指对于上一步得到的策略的 Rollout, 依此类推. 在计算 (3.64) 中的 $u_t^{(k)}$ 项时, 需要进行蒙特卡罗模拟. 对于迭代的 Rollout 策略, 需要进行多层蒙特卡罗模拟. 因此, 找到一个好的启发式策略作为初始值非常重要, 因为最多只有几次迭代过程是可以计算的. 在利用蒙特卡罗方法计算代价函数时, (3.64) 式中的无限序列需要用有限和 $2\leqslant t\leqslant n_\delta$ 来替代. 除了 δ 非常接近 1 的情况以外, n_δ 在中等大小时 (<100), 即可使 δ^{n_δ} 充分小. 蒙特卡罗方法的具体细节可参见 Han 等 (2006) 的第四节.

有限时间问题的 Rollout 算法 Han 等 (2006) 的第三节中将上文提到的有限空间情形下的算法进行了改进. 在前述算法中, 动态规划问题利用反向归纳法, 在确定未来的控制 $u_{k+1}^*,\dots,u_n^*$ 后, 求解能最小化

$$h_{k-1}(u)+E\left[\sum_{i=k+1}^{n}h_{i-1}(u_i^*)\middle|\mathcal{F}_{k-1},\ u_k=u\right] \tag{3.65}$$

的 u_k^*, 从而得到控制 $u_1^*,\dots,u_n^*$. 在 $i\geqslant k+1$ 时, 由于 u_i^* 是过去的观测值和尚未观测到的 $y_k,u_{k+1}^*,y_{k+1},\dots,x_{i-1}^*,y_{i-1}$ 的一个复杂的非线性方程, 上式中的条件期望的计算是一个艰难的任务. 为了解决这个困难, Rollout 算法通过最小化 (3.65) 来对于最优策略 u_k^* 进行近似, 其中 $u_{k+1}^*,\dots,u_n^*$ 替换为某个已知的基准策略$\hat{u}_{k+1},\dots,\hat{u}_n$, 该算法理论上计算并不困难, 且近似值与实际值相差不大. 具体来说, 给定基准策略 $\hat{\boldsymbol{u}}=(\hat{u}_1,\dots,\hat{u}_n)$ 时, 令 $\hat{u}_1^{(1)}$ 为最小化下式的 u:

$$h_{k-1}(u)+E\left[\sum_{i=k+1}^{n}h_{i-1}(\hat{u}_i)\middle|\mathcal{F}_{k-1},\ \hat{u}_k=u\right], \tag{3.66}$$

其中 (3.66) 式第二项中的期望通常会由蒙特卡罗模拟方法得到. 策略 $\hat{\boldsymbol{u}}^{(1)}=(\hat{u}_1^{(1)},\dots,\hat{u}_n^{(1)})$ 被称为 $\hat{\boldsymbol{u}}$ 的 Rollout. Rollout $\hat{\boldsymbol{u}}^{(1)}$ 本身就可以作为一个基准策略, 以求得 $\hat{\boldsymbol{u}}^{(2)}$. 理论上来说, 该过程可以重复任意多次, 从而求得 $\hat{\boldsymbol{u}}^{(1)}$, $\hat{\boldsymbol{u}}^{(2)}$, $\hat{\boldsymbol{u}}^{(3)}$,.... Bayard (1991) 令 $R(\boldsymbol{u})=E[\sum_{i=1}^{n}h_{i-1}(u_i)]$, 证明无论选取怎样的基准策略, Rollout n 次都可以得到最优策略, Rollout

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

过程始终会改进基准策略, 即

$$R(\hat{\boldsymbol{u}}) \geqslant R(\hat{\boldsymbol{u}}^{(1)}) \geqslant R(\hat{\boldsymbol{u}}^{(2)}) \geqslant \cdots \geqslant R(\hat{\boldsymbol{u}}^{(n)}) = R(\boldsymbol{u}^*) \tag{3.67}$$

对任意 $\hat{\boldsymbol{u}}$ 都成立, 其中 $\boldsymbol{u}^*$ 为最优策略. 尽管 (3.67) 式说明对于基准策略进行 Rollout 即可对其改进, 而 Rollout n 次始终可得到动态规划问题的解, 但在实际中很难将一个 Rollout (此处定义为一个反向归纳算法, 该算法在每一步中都需要进行蒙特卡罗模拟和数值优化) 作为另一个 Rollout 的基准策略. Han 等 (2006) 研究发现该问题可以通过回归模型的适应性控制来解决, 令 (3.66) 式中的求和项 $\sum_{i=k+1}^{n}$ 的约束为 $\sum_{i=k+1}^{\min(n,k+m)}$, 即限制求和项最多为 m 项, 从而可以在有限的前向搜索框架下求解.

基于回归和蒙特卡罗模拟 (LS-MC) 的近似值迭代 (3.65) 式中的条件期望项是 u 的函数, 在动态规划问题 (DP) 中被称为**代价函数**. 此处介绍一个从机器学习方法中发展而来的 ADP 方法, 该方法基于两个条件期望相关的统计学概念. 第一个概念是, 对于给定的 u 和历史信息 $\mathcal{F}_{k-1}$, 条件期望即为期望, 因此在 $h_k(u_{k+1}^*), \ldots, h_{n-1}(u_n^*)$ 的生成过程已知时, 可以通过蒙特卡罗模拟方法求得. 第二个概念是, $h_i(u_{i+1})$ 是在给定 $\mathcal{F}_i$ 下的条件期望, 也是一个对于 $h_i(u_{i+1})$ 的回归方程 (或最小方差估计值), 回归量 (预测变量) 由 $\mathcal{F}_i$ 生成. 在样本量很大时 (由蒙特卡罗模拟生成), 回归方程可以通过基本的最小二乘法来估计, 该方法也是非参数估计中的常用方法. 将最小二乘回归 (LS) 方法和蒙特卡罗模拟法 (MC) 相结合, 即可得到对马尔可夫决策问题的 LS-MC 方法. 定义 $\{X_t, t \geqslant 0\}$ 为马尔可夫链, 其从状态 X_t 到状态 X_{t+1} 的转移矩阵取决于 t 时刻的行动 u_t. 令 $f_t(x,u)$ 为状态为 x, 行动为 u 时, 在 t 时刻的成本函数. 考虑如下的统计决策问题, 在每一阶段 k, 选择合适的 x, 以最小化在给定 $u_{k+1}, \ldots, u_n$ 时的代价函数:

$$Q_k(x,u) = E\left[f_k(x,u) + \sum_{t=k+1}^{n} f_t(X_t, u_t) \middle| X_k = x,\ u_k = u \right]. \tag{3.68}$$

注意到 Q_k 是状态 x 和行动 u 的函数, 它们也是机器学习文献中 "Q-学习" 技术的基本组成成分. 定义

$$V_k(x) = \min_x Q_k(x,u), \quad u_k^* = \arg\min_u Q_k(x,u). \tag{3.69}$$

上述函数可以通过动态规划问题 $V_n(x) = \min_u f_n(x,u)$ 的反向归纳方法

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

求得, 在 $n > k \geqslant 1$ 时,

$$V_k(x) = \min_u \{f_k(x,u) + E\left[V_{k+1}(X_{k+1})|X_k = x,\ u_k = u\right]\}, \qquad (3.70)$$

其中极小值等于 u_k^*. 假设状态空间维度有限, LS-MC 方法使用基本函数 ϕ_j, $1 \leqslant j \leqslant J$ 来得到 V_{k+1} 的近似值 $\hat{V}_{k+1} = \sum_{j=1}^{J} a_{k+1,j}\phi_j$, 使用该估计量以及蒙特卡罗方法, 对于代表值的每一个格点 u, 得到如下估计值:

$$E\left[V_{k+1}(X_{k+1})|X_k = x,\ u_k = u\right].$$

进而得到 V_k 的近似值 $\tilde{V}_k$, 以及 u_k^* 的估计值 $\hat{u}_k$. 此外, 利用由控制 $\hat{u}_k$ 生成的样本

$$\left\{\left(X_{k,b}, \tilde{V}_k(u_{k,b})\right) : 1 \leqslant b \leqslant B\right\} \qquad (3.71)$$

可以使用最小二乘法将 $\tilde{V}_k(u_{k,b})$ 对 $[\phi_1(X_{k,b}), \ldots, \phi_J(X_{k,b})]$ 进行回归, 得到 $\tilde{V}_k$ 的近似值 $\hat{V}_k = \sum_{j=1}^{J} a_{k,j}\phi_j$. LS-MC 方法在复杂的路径决定的金融衍生品的定价过程中也有重要应用, 具体可参见 Longstaff 和 Schwartz (2001), Tsitsiklis 和 Van Roy (1999), Lai 和 Wong (2004) 等.

增强学习 Powell 和 Ryzhov (2013) 指出, "ADP (包括增强学习、适应性动态规划和神经元动态规划方面的研究) 已成为广大的算法策略的保护伞", 大部分的 ADP 研究中包含 "快速且有效" 的抽样方法和蒙特卡罗抽样方法的学习函数 (或策略). 文中指出: "学习的起源既包括线下学习 (即利用计算机进行算法训练), 也包括线上学习 (即始终更新学习的状态). 学习的起源也包含算法的多个方面, 如学习策略中的参数, 学习价值函数以及学习如何将树进行扩展. 学习既包括积极型学习, 也包括消极型学习. 消极型学习策略收集相关信息, 并更新函数的信度. 在该策略中并不会主动尝试提升学习策略效率的方法. 积极型学习策略指考虑到收集到的信息会在后续进程中提升价值, 而愿意在当前采取次优行动的策略."

增强学习是积极型学习策略的一个领域, 该方法在计算机和人工智能领域得到广泛应用, 且在神经科学、博弈论、多代理系统、运筹学和控制系统等领域中也有相关应用. 正如论文 Kaelbling 等 (1996) 中指出, "增加学习的目标是迷人的, 该方法利用激励和惩罚来对于各个代理进行编程, 而不需要具体说明任务是如何实现的." 学习是一个 "与动态环境进行试错式交互" 的过程. 具体来说, "代理是通过感知环境和行动来与环境相关联的", 在每一次交互过程中, 收集到输入值 i、当前环境状态 s

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

之后的指示值和选择的行动 a 的信息, 来生成输出值. “行动影响环境的状态, 且状态转移值通过一个‘增强信号’标量来与代理进行通信.” 积极型学习采用从增强信号中得到的学习的形式来逐渐地改进行为的产出. 该过程的实现依赖于一系列的算法, 如 LS-MC 确定性等价算法、适应性启发式关键性差分或临时差分、Q-学习方法等, 具体可参见 Kaelbling 等 (1996)、Barto 等 (1983)、Sutton (1988)、Watkins 和 Dayan (1992)、Kumar 和 Varaiya (1986) 等.

蒙特卡罗树状搜索、策略和价值网络、深度学习 本书的两个副标题 “算法” 和 “优化” 的主要研究对象之一就是 ADP 问题, 特别是近年来发展迅速的增强学习问题. Silver 等 (2016) 对于该领域的发展进行了综述, 并讨论了增强学习方法在围棋游戏的计算机编程中的应用的发展. 后者 “在很长时间内, 由于庞大的搜索空间需求和在评估棋局的形势和移动步骤上的困难, 一直被认为是人工智能在经典游戏中最大的挑战.” 近年来研究的发展包括: (a) 起源于专业游戏的机器学习; (b) 价值网络的增强学习以估计棋局形势; (c) 策略网络的增强学习以选择接下来的步骤; (d) 蒙特卡罗树搜索法 (MCTS), 即使用蒙特卡罗 Rollout 算法来估计搜索树中各步骤和行动配对的价值, 从而当搜索树增大时能够使相关值更加精确; 利用策略网络和价值网络方法, 可以将搜索范围缩小到 Rollout 中的高概率行动集和样本行动集. 另外, “深度卷积神经网络” 方法被用于 “减少搜索树的有效深度和广度: 利用价值函数来评估棋局形势, 利用策略网络来评估样本行动.” 计算机程序 AlphaGo 就是在有效地将策略网络、价值网络和 MCTS 相结合的基础上设计的. AlphaGo 在 2015 年以 5:0 的比分击败了欧洲围棋冠军, 在 2016 年以 4∶1 的比分击败了韩国九段冠军棋手.

8. **信息理论和全局策略** 假设 $\boldsymbol{R}_t$ 独立同分布, 则由于 $\boldsymbol{R}_t$ 与 $\mathcal{F}_t$ 相互独立, 最优投资组合的权重向量 (3.17) 可以写成 $\boldsymbol{w}^* = \arg\max_{\boldsymbol{w}\in\mathcal{W}} E[\log(\boldsymbol{w}^\intercal \boldsymbol{R}_t)]$, 其与时间 t 无关. 因此, 近似最优策略即为*最优恒定再平稳投资组合* (BCRP), 即 $\boldsymbol{w}_t$ 的最优选择对于所有的 t 都是恒定的. 对于有限的 T, $\boldsymbol{w}_T^*$ 由下式给出:

$$\boldsymbol{w}_T^* = \arg\max_{\boldsymbol{w}\in\mathcal{W}} \prod_{t=1}^{T} \boldsymbol{w}^\intercal \boldsymbol{R}_t, \tag{3.72}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

此时最优资产为 Z_T^*. 然而, 这也意味着 $\boldsymbol{w}_T^*$ 是所有 $\boldsymbol{R}_t, t=1,\ldots,T$ 的函数, 使得 BCRP 是可预测的, 因而在没有后见之明的情况下不可行. 为解决该问题, Cover 和 Ordentlich (1996) 提出了 Dirichlet 权重投资组合如下:

$$\hat{\boldsymbol{w}}_{t+1}=\frac{\int_{\mathcal{W}}\boldsymbol{w}\left[\prod_{i=1}^{t}\boldsymbol{w}^{\intercal}\boldsymbol{R}_i\right]\mu_{\boldsymbol{\alpha}}(d\boldsymbol{w})}{\int_{\mathcal{W}}\left[\prod_{i=1}^{t}\boldsymbol{w}^{\intercal}\boldsymbol{R}_i\right]\mu_{\boldsymbol{\alpha}}(d\boldsymbol{w})},\quad t=1,2,\ldots, \tag{3.73}$$

其中权重为 m 维单纯形 $\{\boldsymbol{w}=(w_1,\ldots,w_m)^{\intercal}: w_j\geqslant 0, \sum_{j=1}^m w_j=1\}$ 上的 Dirichlet 分布:

$$\mu_{\boldsymbol{\alpha}}(d\boldsymbol{w})=\frac{\Gamma(\sum_{j=1}^m\alpha_j)}{\prod_{j=1}^m\Gamma(\alpha_j)}\left(\prod_{j=1}^m w_j^{\alpha_j-1}\right)d\boldsymbol{w}. \tag{3.74}$$

其中参数向量 $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_m)^{\top}$ 满足对于任意的 $j=1,\ldots,m$, $\alpha_j>0$.

在 $t=0$ 时, $\boldsymbol{w}_T^*$ 为 $\mu_{\boldsymbol{\alpha}}(d\boldsymbol{w})$ 的均值向量. 为了证明该策略的 "全局性", Ordentlich 和 Cover (1998)[①] 证明对于任意 T,

$$\max_{\{\tilde{\boldsymbol{w}}_t\}_{t=1}^T}\min_{\{\boldsymbol{R}_t\}_{t=1}^T}\frac{Z_T^*}{\tilde{Z}_T}=U_T, \tag{3.75}$$

其中最大值是对于所有非预测性投资组合 $\{\tilde{\boldsymbol{w}}_t\}_{t=1}^T$ 而言. $\tilde{Z}_T{=}Z_0\prod_{t=1}^T\tilde{\boldsymbol{w}}_t\boldsymbol{R}_t$ 是终值资产, 且有

$$U_T=\sum_{n_1+\cdots+n_m=T}\binom{T}{n_1,\ldots,n_m}2^{-T\mathcal{I}(n_1/T,\ldots,n_m/T)}, \tag{3.76}$$

其中 $\mathcal{I}(p_1,\ldots,p_m)=-\sum_{j=1}^m p_j\log_2 p_j$ 是信息理论中的 Shannon 熵, 括号中的项为多项式系数. (3.75) 式中的极大极小值是指投资者与 (最差情形下) 对于市场之间的博弈来最小化比率 $\dfrac{Z_T^*}{\tilde{Z}_T}$ 的值, 其中投资者使用非预测性的策略, 而对于市场则选择一个市场序列 $\{\boldsymbol{R}_t\}_{t=1}^T$. 换言之, 投资者希望最小化未能使用最优预测投资组合的最大后悔值, 该投资组合的权重为 (3.72), 且只在时间 T 时可行. 根据 (3.76) 式可知 $U_T=O(T^{(m-1)/2})$. Cover 和 Ordentlich (1996) 选取 $\boldsymbol{\alpha}=\left(\dfrac{1}{2},\ldots,\dfrac{1}{2}\right)^{\top}$, 表明对

① 与最小最大化方法不同, Ordentlich 和 Cover (1998) 使用了最大最小化方法来分析 $(Z_T^*/\tilde{Z}_T)$ 的倒数. 在该方法下, (3.75) 的等价形式即为求 $(\tilde{Z}_T/Z_T^*)$ 的最小最大值, 且等式右侧的值为 U_T^{-1}. Ordentlich 和 Cover (1998) 将这个最大最小化方法称为寻找最优后验投资组合的未预测性策略, 而 Cover 和 Ordentlich (1996) 以及 Cover (1996) 称等价策略为最小化最大后悔值.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

于所有的 T 和 $\{\boldsymbol{R}_t\}_{t=1}^T$, 都有 $\dfrac{Z_T^*}{\hat{Z}_T} \leqslant (T+1)^{m+1}$. 结合该结合和 (3.75) 式, 可知

$$\lim_{T\to\infty} \sup_{\{\boldsymbol{R}_t\}_{t=1}^T} \frac{1}{T} \log \frac{Z_T^*}{\hat{Z}_T} = 0. \tag{3.77}$$

Cover (1996) 称能够取得 (3.75) 的最小最大值的投资组合为 "全局投资组合", 因为该投资组合为信息论中的全局数据压缩算法的匹配结果. 论文指出, (3.75) 与数据压缩算法的 "最小最大冗余价值" 相同, 该算法可以在未知分布集上取得 "对于未知的最小最大冗余源的熵极限" (因此是 "全局" 极值). 注意到 (3.73) 式实际上将 (3.72) 式中的最大化过程替换为一系列的加权均值. 该流程与 de la Peña 等 (2009) 的第 11 与 17 章中提到的拟最大化方法类似, 该方法使用拟最大似然值将似然函数 (特别是广义似然函数) 推断与贝叶斯方法相结合.

9. **Kalman 滤波, LQG 控制和分离准则** 在 3.3.6 节中, 对于存在参数不确定性时的动态投资组合问题的研究进行了综述. 在贝叶斯方法中, 假定未知参数 (或其变换) 服从正态先验分布, 然后使用 Kalman 滤波 (或 Kalman-Bucy 滤波) 方法来迭代求解后验均值和协方差矩阵. 现有的文献侧重于多期投资组合选择中的序贯学习方面的研究, 包括指数效用函数下的投资组合选择的 "预测风险", 和 "边做边学" 以降低参数不确定性, 但是这些研究都没有考虑交易成本情况. 在 3.3.5 节中, 我们讨论了 Gârleanu 和 Pedersen (2013), 该论文中假设存在二次形式的交易成本, 超额收益服从多因子回归模型, 且因子服从自回归过程. 文中的多期投资组合优化问题 (3.21) 与此处我们接下来讨论的求解线性状态空间下控制模型的 Kalman 滤波方法相关.

存在输入时的线性状态空间模型和 Kalman *滤波* Lai 和 Xing (2008) 一书的 5.3 节对于线性状态空间模型进行了介绍: $\boldsymbol{x}_{t+1} = \boldsymbol{F}_t\boldsymbol{x}_t + \boldsymbol{w}_{t+1}$, $\boldsymbol{y}_t = \boldsymbol{G}_t\boldsymbol{x}_t + \boldsymbol{\varepsilon}_t$, 其中 $\boldsymbol{x}_t$ 是未观测的线性变化的状态向量, $\boldsymbol{y}_t$ 是状态 $\boldsymbol{x}_t$ 的变换与附加噪声项 $\boldsymbol{\varepsilon}_t$ 之和. 假设存在相互独立的零均值高斯向量 $\boldsymbol{w}_{t+1}$, 以及相互独立、且与状态变量独立的 $\boldsymbol{\varepsilon}_t$. 假设 $\mathrm{Cov}(\boldsymbol{w}_t) = \boldsymbol{\Sigma}_t$, $\mathrm{Cov}(\boldsymbol{\varepsilon}_t) = \boldsymbol{V}_t$, 则此时给定观测值 $\boldsymbol{y}_s$ 下, $\boldsymbol{x}_t$ $(s \leqslant t)$ 的条件分布为正态分布, 且 Kalman 滤波方法给出了计算均值 $\hat{\boldsymbol{x}}_{t|t}$ 和预测误差 $\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t}$ 的协方差矩阵的递归求解公式. Kumar 和 Varaiya (1986) 的第 7 章指出, 这

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

些公式可以推广到如下的线性控制系统中:

$$\begin{aligned} \boldsymbol{x}_{t+1} &= \boldsymbol{F}_t \boldsymbol{x}_t + \boldsymbol{\Gamma}_t \boldsymbol{u}_t + \boldsymbol{w}_{t+1}, \\ \boldsymbol{y}_t &= \boldsymbol{G}_t \boldsymbol{x}_t + \boldsymbol{\Lambda}_t \boldsymbol{u}_t + \boldsymbol{\varepsilon}_t, \end{aligned} \tag{3.78}$$

其中 $\boldsymbol{u}_t$ 是时间 t 时的未预测性输入 (控制). 实际上, 本章第三节介绍了如何利用线性状态方程和测量方程来得到 $\boldsymbol{u}_t \equiv \mathbf{0}$ 情况的 Kalman 递归方程:

$$\begin{aligned} \hat{\boldsymbol{x}}_{t+1|t} &= \boldsymbol{F}_t \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{\Gamma}_t \boldsymbol{u}_t + \boldsymbol{K}_t (\boldsymbol{y}_t - \boldsymbol{G}_t \hat{\boldsymbol{x}}_{t|t-1} - \boldsymbol{\Lambda}_t \boldsymbol{u}_t), \\ \boldsymbol{P}_{t+1} &= (\boldsymbol{F}_t - \boldsymbol{K}_t \boldsymbol{G}_t) \boldsymbol{P}_t (\boldsymbol{F}_t - \boldsymbol{K}_t \boldsymbol{G}_t)^\top + \boldsymbol{\Sigma}_{t+1} + \boldsymbol{K}_t \boldsymbol{V}_t \boldsymbol{K}_t^\top, \end{aligned} \tag{3.79}$$

其中 $\boldsymbol{K}_t = (\boldsymbol{F}_t \boldsymbol{P}_t \boldsymbol{G}_t^\top)(\boldsymbol{G}_t \boldsymbol{P}_t \boldsymbol{G}_t{}^\top + \boldsymbol{V}_t)^{-1}$ 是 Kalman 增益方程. 对于 $\boldsymbol{P}_t = \mathrm{Cov}(\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t-1})$, 上式中第二个递归方程称为 Riccati 方程, 第一个递归方程则是用于求解 Kalman 预测值 $E[\boldsymbol{x}_{t+1}|(\boldsymbol{y}_s, \boldsymbol{u}_s), s \leqslant t]$. Kalman 滤波值 $\hat{\boldsymbol{x}}_{t|t}$ 由下式给定:

$$\hat{\boldsymbol{x}}_{t|t} = \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{P}_t \boldsymbol{G}_t^\top (\boldsymbol{G}_t \boldsymbol{P}_t \boldsymbol{G}_t^\top + \boldsymbol{V}_t)^{-1} (\boldsymbol{y}_t - \boldsymbol{G}_t \hat{\boldsymbol{x}}_{t|t-1} - \boldsymbol{\Lambda}_t \boldsymbol{u}_t).$$

LQG 控制问题 标题中的 "L" 和 "G" 代表线性 (Linear) 高斯 (Gaussian) 模型 (3.78), 而 "Q" 则代表二次 (Quadratic) 控制误差的成本和控制成本. 对于一个多输入多输出 (MIMO, 此处指代多元变量 $\boldsymbol{u}_t$ 和 $\boldsymbol{y}_t$) 系统的随机控制问题, 需要选取非预测性的控制 $\boldsymbol{u}_t$, $1 \leqslant t \leqslant T$ 来最小化

$$E\left\{ \sum_{t=1}^{T} \boldsymbol{x}_t^\top \boldsymbol{Q}_t \boldsymbol{x}_t + \sum_{t=1}^{T-1} \boldsymbol{u}_t^\top \boldsymbol{R}_t \boldsymbol{u}_t \right\}. \tag{3.80}$$

相关联的动态规划问题, 可以使用反向归纳法进行求解. Kumar 和 Varaiya (1986, 7.5 节) 给出了计算最优 $\boldsymbol{u}_t^*$ 的解析反向递归公式如下:

$$\boldsymbol{u}_t^* = -\boldsymbol{L}_t \hat{\boldsymbol{x}}_{t|t-1}, \quad \text{其中 } \boldsymbol{L}_t = (\boldsymbol{\Gamma}_t^\top \boldsymbol{S}_{t+1} \boldsymbol{\Gamma}_t + \boldsymbol{R}_t)^{-1} \boldsymbol{\Gamma}_t^\top \boldsymbol{S}_{t+1} \boldsymbol{F}_t, \tag{3.81}$$

其中 $\boldsymbol{S}_t$ 由反向 Riccati 方程 $\boldsymbol{S}_T = \boldsymbol{Q}_T$ 求得, 且对于 $1 \leqslant t < T$,

$$\boldsymbol{S}_t = \boldsymbol{F}_t^\top \left(\boldsymbol{S}_{t+1} - \boldsymbol{S}_{t+1} \boldsymbol{\Gamma}_t [\boldsymbol{\Gamma}_t^\top \boldsymbol{S}_{t+1} \boldsymbol{\Gamma}_t + \boldsymbol{R}_t]^{-1} \boldsymbol{\Gamma}_t^\top \boldsymbol{S}_{t+1} \right) \boldsymbol{F}_t + \boldsymbol{Q}_t.$$

分离准则和确定性等价准则 注意到随机误差项 $\boldsymbol{w}_t$ 和 $\boldsymbol{\varepsilon}_t$ 的协方差矩阵 $\boldsymbol{\Sigma}_t$ 和 $\boldsymbol{V}_t$ 并没有出现在最优控制 $\boldsymbol{u}_t^*$ 中. 因此不存在噪声项的情况

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

与估计值不存在误差的情况的控制准则是一致的, 这被称为确定性等价准则, 该准则对于分离准则成立时的随机控制问题也成立. 分离准则在随机系统中最优反馈控制准则的设计问题中, 将估计值与控制相分离. 具体而言, 控制问题可以在不存在未知参数和未观测状态的情况下求解. 而在存在未预测状态时, 分离准则首先对状态进行估计, 然后将估计值代入到参数和状态已知的最优控制准则中进行求解.

连续时间情况的推广和 Kalman-Bucy 滤波　为简化问题, 令 $\boldsymbol{\Lambda}_t = \mathbf{0}$, 这也是实际中的情形. 此时, (3.78) 式可以扩展到如下的连续时间情形:

$$\begin{aligned} d\boldsymbol{x}_t &= (\boldsymbol{F}_t\boldsymbol{x}_t + \boldsymbol{\Gamma}_t\boldsymbol{u}_t)dt + \boldsymbol{\Sigma}_t^{1/2}d\boldsymbol{B}_t, \\ d\boldsymbol{y}_t &= \boldsymbol{G}_t\boldsymbol{x}_t dt + \boldsymbol{V}_t^{1/2}d\boldsymbol{B}_t^*, \end{aligned} \tag{3.82}$$

其中 $\boldsymbol{B}_t$ 和 $\boldsymbol{B}_t^*$ 是相互独立的布朗运动. Kalman-Bucy 滤波, 即(3.79)式的连续时间情形, 可以定义为如下的常微分方程组 (ODEs):

$$\begin{aligned} d\hat{\boldsymbol{x}}_t &= (\boldsymbol{F}_t\hat{\boldsymbol{x}}_t + \boldsymbol{\Gamma}_t\boldsymbol{u}_t)dt + \boldsymbol{K}_t(d\boldsymbol{y}_t - \boldsymbol{G}_t\hat{\boldsymbol{x}}_t dt), \\ \frac{d}{dt}\boldsymbol{P}_t &= \boldsymbol{F}_t\boldsymbol{P}_t + \boldsymbol{P}_t\boldsymbol{F}_t^\top + \boldsymbol{\Sigma}_t - \boldsymbol{P}_t\boldsymbol{G}_t^\top\boldsymbol{V}_t^{-1}\boldsymbol{G}_t\boldsymbol{P}_t, \end{aligned} \tag{3.83}$$

其中 $\boldsymbol{K}_t = \boldsymbol{P}_t\boldsymbol{G}_t^\top\boldsymbol{V}_t^{-1}$. 连续时间下的 LQG 问题的目标是, 寻找非预测性控制来最小化

$$E\left[\boldsymbol{x}_T^\top\boldsymbol{Q}_T\boldsymbol{x}_T + \int_0^T(\boldsymbol{x}_t^\top\boldsymbol{Q}_t\boldsymbol{x}_t + \boldsymbol{u}_t^\top\boldsymbol{R}_t\boldsymbol{u}_t)dt\right]. \tag{3.84}$$

最优控制准则为 $\boldsymbol{u}_t^* = -\boldsymbol{L}_t\hat{\boldsymbol{x}}_t$, 其中 $\boldsymbol{L}_t = \boldsymbol{R}_t^{-1}\boldsymbol{\Gamma}_t^\top\boldsymbol{S}_t$ 且 $\boldsymbol{S}_t$ 由终值情形为 $\boldsymbol{S}_T = \boldsymbol{Q}_T$ 时的 ODE 决定.

10. **求解多期投资组合优化问题的新方法**　3.3 节回顾了多期投资组合优化方面的文献. 在第二章研究的单期问题中, 对于投资组合未来的收益率矩阵的建模和预测问题受到广泛关注, 且在近年来得到了重要的发展. 但对于多期问题, 该问题却一直被忽视. 在 1.4 节和 2.7 节中提到, 投资基金的业绩是基于风险调整的测度在每一个周期中进行评估, 因此在 2.7 节中提到的单期问题中, 在选取 λ 时, 是在有效前沿上选取能最大化业绩测度的 λ, 对于多期问题, 自然也可以类似地进行分析. 然而, 在 3.3 节提到的文献中, 大部分都没有考虑对未来的协方差矩阵的预测. 举例来说, 在 3.3.6 节对于存在参数不确定性时的动态投资组合选择问题的研究中, 涉及的参数是平均收益, 并利用 Kalman 滤波方法来求得, 而协方差

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

参数则是假设已知, 或可以根据现有的数据进行估计. 正如 3.3 节的最后一段指出, Kalman 滤波方法并没有被用于对波动率参数的不确定性进行建模, 原因在于此时会涉及非线性滤波问题. 前面的补充材料部分重点介绍了 Kalman 滤波方法在线性状态空间模型的状态预测和控制问题中的作用, 其中线性状态空间模型由线性状态和测量方程来刻画. 当状态方程和测量方程中有一个非线性时, 滤波的复杂程度会大大增加. 利用序贯蒙特卡罗方法对该问题的求解方法在近年来得到了重大发展, 该方法也称为适应性粒子滤波, 可用于对波动率进行估计, 具体可参见 Chan 和 Lai (2013, 2016)、Lai 和 Bukkapatanam (2013) 以及 Lai 和 Xing (2016) 的第 6 章.

非线性滤波方法说明新方法可以用于解决多期投资组合优化中的一些长期未解决的问题. 除了提供更多的 3.3 节涉及的文献的细节和讨论外, 这些补充材料也对于多期投资组合优化问题提供了一种新的技术思路. 特别地, Lai 和 Gao (2016a, b) 以及 Lai 等 (2016a) 中利用奇异随机控制、近似动态规划、适应性滤波和蒙特卡罗树搜索等方法, 提出了一种解决与 3.3 节第一段中提到的一些长期未解决问题相关联的多期投资组合优化问题的新方法.

3.5 习题

1. 在存在比例化的交易成本时, 会产生一个 “大爆炸式” 的控制问题. 此时会存在一个价格的 “不交易区域”, 当价格在该区域中时, 保持不变; 当价格不在该区域时, 会通过密集交易, 在价格太低时买入, 或在价格过高时卖出, 从而使价格回到该区域中. 解释上述现象产生的原因. 阅读 Magill 和 Constantinides (1976) 以及 Davis 和 Norman (1990), 阐述论文中发现的 “不交易区域”.

2. 阅读 Markowitz 和 van Dijk (2003), 阐述论文中提到的启发式方法. 该方法将单期 Markowitz 均值–方差投资组合再平衡问题进行改进以应用于存在多个证券和交易成本时的多期再平衡问题.

3. 阅读 Cvitanić 等 (2006), 解释论文中求解最优投资组合权重向量 (3.25) 的方法. 根据 3.4 节补充材料 9, 给出一种新的最优投资优化问题的求解方法, 新方法中需要使用 Kalman-Bucy 滤波, 并使用补充材料中提到的二次成本准则 (3.83).

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

第四章　电子交易中的计量经济学

1.1 节和 1.7 节指出, 自 20 世纪 80 年代以来, 随着交易所技术的发展, 高分辨率数据的可得性和数据的质量都显著增加, 资产的买卖逐渐由线下交易转至连续、低延迟的双向拍卖系统交易. 随着高分辨率市场数据的出现, 一系列新的量化研究方法应运而生. 特别地, 高频计量经济学, 即与电子交易平台中的量化交易相关的计量经济学建模研究, 成为目前金融计量经济学研究中一个活跃的领域. 本章对于高频计量经济学中的一些主要的方法和模型进行回顾. 4.1 节首先对于电子交易所中的交易进行介绍, 重点对于限价委托簿 (LOB)、买卖策略的提交和交易发生情况进行介绍. 高频计量经济学的重点是上述最后一方面 (即交易发生的时间、价格和数量), 以及交易数据的时间序列特征. 而下一章则对高频交易策略进行具体介绍. 高频交易策略对于未来交易进行预测, 在预测时利用限价委托簿中的数据, 并对于数据进行统计研究和建模. 4.2 节介绍了单个资产的交易价格的基本计量经济学模型. 对于风险资产的有效价格过程的二次变差 (也称为已实现方差) 进行预测, 是金融计量经济学中的重要问题, 4.3 节对于该问题进行了研究. 4.4 节对于高频交易中资产价格的多元联合分布模型进行研究. 除了已实现方差外, 即时波动率和联合波动率在数量金融中也起着重要的作用. 4.5 节介绍了一个估计已实现方差等测度的替代方法——Fourier 估计量. 4.6 节讨论了交易之间的时间间隔和交易数据的其他统计量的计量经济学模型, 并讨论了高频交易的波动和资产价格的长期波动率的关系. 4.7 和 4.8 节给出了相应的补充材料和习题.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

4.1 电子交易和交易数据

交易是指两个不同主体对于资源进行交换的过程, 交易中产生的数据和信息即为交易数据. 从远古时期开始, 交易数据就被交易者广泛利用. Lo 和 Hasanhodzic (2010) 指出, 根据巴比伦时期的石板的记录, 在公元前 747 年到公元前 651 年之间, 商品交易者就将交易数据与天文观测相结合, 用于制定季节性的交易策略. 正如第一章中提到, 自最小交易单位的发明已来, 交易池中的逐笔交易数据被广泛使用. 然而, 对于每一笔交易进行详细记录的成本很高 (参见 Freeman (2006)), 且记录中也可能存在一定的错误. 错误可能来源于交易过程中的人工失误或在公开喊价系统中进行交易时的数据转移失误 (参见 Gorham 和 Singh (2009)). 而在现在的电子交易中, 交易和交易数据与传统的交易有着显著的不同.

具体而言, 以 \$114.50 的价格购买一股苹果公司的股票 (代号 AAPL) 即为一笔交易. 在该交易中, 包含流动性最高的资产, 即现金 \$114.50, 与流动性较差的资产, 即一股苹果公司的股票之间的交换. 在大多数的订单或报价驱动[①]的电子交易中, 当交易所的匹配程序判定买方和卖方就一定金额的交易达到匹配, 且其中一方已经在限价委托簿中提交了订单时, 即视为交易发生. 限价委托簿中包括离散化的尚未交易的资产限价买卖合约, 如 AAPL 股票或电子迷你期货的合约等. 表 4.1 描述了 “以 \$ 114.50 的价格, 在 NASDAQ 购买 150 股 AAPL 股票的限价合同” 的情形, 表中给出了交易发生前和交易匹配后的限价委托簿的情形. 匹配完成后, 交易会作为研究中的一个数据点. 流动性资产

表 4.1 买入 150 股限价订单, 限价为 \$114.50, GTC,[②] 产生 100 股在 \$114.50 价格上交易的限价订单, 而剩下的 50 股成为 LOB 中的最优买入股

		114.52	145				114.53	350
		114.51	51				114.52	145
		114.50	100	$\Longrightarrow$			114.51	51
80	114.49				50	114.50		
150	114.48				80	114.49		

① 订单驱动是指交易者可以提交订单, 而未进行交易的订单则保留在限价委托簿中. 报价驱动是指限价委托簿只包括经销商提交的订单. 此处的经销商也指专家或做市商, 专门负责报价, 即使在极端市场情形下也要履行职责. 大多数的交易所可以看成两者的结合.

② GTC 指一直有效, 直到取消.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

在一秒钟内可以发生多次交易, 因此交易中会产生丰富的数据用于建模. 第五章和第八章对于限价委托簿中的交易数据、匹配算法和统计建模进行了详细的介绍. 本章则只研究交易发生后, 会在交易所立即公布的交易数据.

4.2 高频数据建模

本节首先介绍了 Roll 提出的微观结构模型, 该模型对于交易价格变化序列的负一期自回归系数进行分析, 从而验证了有效市场假说 (EMH). 本节对于 Roll 模型进行了一般化, 首先将标的资产的价格取对数, 然后加入微观结构的噪声项, 并结合 Aït-Sahalia 等 (2005) 以及 Zhang 等 (2005) 中的模型, 提出新的模型.

4.2.1 Roll 提出的买入–卖出反弹模型

Roll (1984) 利用市场微观结构的买入–卖出反弹特性, 在有效市场假说下, 求解得到在 2.1.2 节中提到的高频交易中价格变化的 MA(1) 自回归模型, 如下所示:

$$P_i = P_i^* + cI_i, \tag{4.1}$$

其中 $2c$ 是市场中的买卖价差, $I_i = 1$ (或 -1) 指代买入 (或卖出), P_i^* 和 P_i 分别是 EMH 下的有效价格和第 i 次交易发生时的交易价格. 在 Roll 模型中, 假设市场达到均衡时, I_i 独立同分布,

$$P\{I_i = 1\} = P\{I_i = -1\} = 1/2,$$

且 P_i^* 的增量独立且与 I_i 独立. 定义第 i 次交易发生时的价格变化为

$$D_i = P_i - P_{i-1} = P_i^* - P_{i-1}^* + c(I_i - I_{i-1}), \tag{4.2}$$

则有

$$\mathrm{Cov}(D_i, D_{i-1}) = -c^2; \quad \mathrm{Cov}(D_i, D_{i-j}) = 0, j \geqslant 2. \tag{4.3}$$

该结果也与交易价格变化的 MA(1) 特性相吻合. 4.7 节的补充材料 6 介绍了利用市场微观结构的其他特性的替代模型, 并对于 EMH 下的 MA(1) 自回归性质进行了解释.

4.2.2 包含噪声项的市场微观结构模型

本节首先对于 (4.1) 式进行了如下修正和一般化. 定义 X_t 为时间 t 内的有效价格的对数, 定义 $t_1 < \cdots < t_n$ 为 $[0, T]$ 内的交易时间的集合, 令 Y_t 为时

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

间 t 时的交易价格的对数. 除了类似于 (4.1) 式, 将 P_i 和 P_i^* 直接替代为其对数 (对数方法的优势在于可以解释价格的市场摩擦的加法效应, 如买卖价差是交易价格的一定比例), 此处将 Roll 模型扩展为

$$Y_{t_i} = X_{t_i} + \varepsilon_i, \tag{4.4}$$

其中 ε_i 是市场微观结构的噪声项, 且满足 $E(\varepsilon_i) = 0$. 噪声项的来源包括:

- 买入–卖出反弹, 但不包括 Roll 提出的过度限制假设下不随时间变化的恒定买入–卖出反弹.
- 不同的电子通信网络间的同步性误差.
- 记录的延迟和记录错误, 当交易的记录时间发生在实际时间之后, 则会导致延迟.
- 大宗交易并没有得到即时响应.①
- 交易价格 Y_{t_i} 是离散化的, 而 X_t 则是连续化的.
- 订单流和库存控制效应的战略组成成分.
- 交易所和数据供应商在控制活动时期的后期修正处理, 如对于上期价格的外推.

假设 X_t 增量独立, ε_i 独立同分布, 均值为 0, 方差为 v_ε, 且与 X_t 独立. 则对于 Roll 模型的一般形式 (4.4), 在 $j \geqslant 2$ 时, $\mathrm{Cov}(D_i, D_{i-1}) = -v_\varepsilon < 0$ 和 $\mathrm{Cov}(D_i, D_{i-j}) = 0$ 仍然成立, 其中 $D_i = Y_{t_i} - Y_{t_{i-1}}$.

连续时间有效价格过程 P_t^*, 特别是其方差项, 是期权定价理论和风险管理的基石 (参见 Lai 和 Xing (2008) 一书的第 8、10 和 12 章). 尽管高频交易将市场数据带入了连续化的框架中, 即交易时间 t_i 的集合是 $[0, T]$ 的一个 "紧" 子集, 其网格②接近于 0. Roll 模型 (4.1) 第一次在微观结构中加入噪声项, 以考虑买入–卖出反弹的影响, 研究表明有效价格的变动并不能根据时间序列 P_{t_i}

① 大宗交易是指交易量大于等于交易所预先设定的阈值的交易, 通常私下协商交易, 而不会在电子交易市场中进行交易. 举例来说, 在卖出大量股票份额 (如 10000 股) 时, 对冲基金可以与另一个基金公司通过经纪交易商进行大宗交易, 使得卖出股票的公司可以获得更高的卖价, 而买入股票的公司则可以获得市场费率的折扣.

② 4.3 节的第一段对于网格进行了正式定义.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

的统计模型来预测. Aït-Sahalia 等 (2005) 以及 Zhang 等 (2005) 在几何布朗运动 (GBM) $dP_t^*/P_t^* = \mu dt + \sigma dB_t$ 和更一般的 $X_t = \log P_t^*$ 的扩散过程:

$$dX_t = \mu_t dt + \sigma_t dB_t \tag{4.5}$$

的背景下进行研究, 其中 B_t 是标准布朗运动. 假设在模型 (4.4) 中存在附加的市场结构噪声项 ε_i, Roll 研究表明, 如果直接根据 Y_{t_i} 来估计几何布朗运动的参数, 会导致参数不一致的情况. 只利用聚合时间 (如逐天或逐月) 收盘价数据, 而将大部分的数据废弃时, 反而可以得到一致的参数, 但该方法并没有对于可得的数据进行有效利用. 一致性问题在研究更一般的模型 (4.5), 且 μ_t 和 σ_t 为时变参数时, 更为明显. 在接下来的一节中, 我们讨论了三种可以解决该问题的方法.

4.3 X_t 的积分方差的估计

首先, 注意到 (4.4) 式并没有对于连续时间定价过程 X_t 的模型进行明确定义. 当尚未观测的随机过程 X_t 给定时, 根据观测值 Y_{t_i} 来预测未观测状态变量 X_t $(0 \leqslant t \leqslant T)$ 是一个滤波问题 (参见 3.4 节). 与估计理论价格 X_t 的轨迹相比, 预测其*积分方差* (或*二次变差*) 在金融中是一个更重要的问题. 积分方差的定义如下:

$$[X]_T = \operatorname*{p\text{-}lim}_{\operatorname{mesh}(\Pi)\to 0} \sum_{i=1}^{n} (X_{t(i)} - X_{t(i-1)})^2, \tag{4.6}$$

其中 Π 是 $[0,T]$ 的一个分割 $0 = t_0 < \cdots < t_n = T$, 而 $\operatorname{mesh}(\Pi) = \max\limits_{1\leqslant i\leqslant n}(t_i - t_{i-1})$. (4.6) 式的依概率极限的存在性可根据连续时间半鞅理论得到, 该理论也被广泛认为是有效价格过程建模的最一般的模型. 附录 A 对于本节和后续章节中使用的半鞅理论进行了总结.

在不存在微观结构噪声时 (即 (4.4) 式中 $\varepsilon_i = 0$, 从而有 $X_{t_i} = Y_{t_i}$), 积分方差 (4.6) 的一个一致估计量为已实现方差, 即

$$RV^{(n)} = \sum_{i=2}^{n} \left(Y_{t_i} - Y_{t_{i-1}}\right)^2, \tag{4.7}$$

注意到 $\max\limits_{2\leqslant i\leqslant n}(t_i - t_{i-1})$ 在高频交易中通常非常小. 但是, 当 ε_i 为独立非退化随机变量时, $RV^{(n)}$ 不再是一致的. 实际上, 假设 ε_i 独立同分布, 均值为 0, 方差为 v_ε, 且 $E(\varepsilon_i^4) < \infty$, 假设 X_t 是一个与 ε_i 独立的扩散过程, Zhang 等 (2005)

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

以及 Hansen 和 Lunde (2006) 发现:

$$RV^{(n)}/n \xrightarrow{P} 2v_\varepsilon. \tag{4.8}$$

因此, 存在微观结构噪声项 ε_i 时, $RV^{(n)}$ 是积分方差的有偏估计量, 且偏差随着样本数的增加而线性增加. 对于这些作者们考虑的包含扩散过程 (4.5) 的更一般的半鞅过程, (4.8) 式的证明如下. 注意到

$$\begin{aligned}\sum_{i=2}^{n}(Y_{t_i}-Y_{t_{i-1}})^2 &= \sum_{i=2}^{n}\left(X_{t_i}+\varepsilon_i-X_{t_{i-1}}-\varepsilon_{i-1}\right)^2\\ &= \sum_{i=2}^{n}(X_{t_i}-X_{t_{i-1}})^2+2\sum_{i=2}^{n}(X_{t_i}-X_{t_{i-1}})(\varepsilon_i-\varepsilon_{i-1})\\ &\quad+\sum_{i=2}^{n}(\varepsilon_i-\varepsilon_{i-1})^2.\end{aligned}$$

根据 (4.6) 式和鞅收敛定理 (参见附录 A) 可知, 前面两个求和项为 $O_p(1)$. 因此, 根据大数定理可得 $n^{-1}\sum_{i=2}^{n}(\varepsilon_i-\varepsilon_{i-1})^2 \xrightarrow{P} 2v_\varepsilon$, 即可证得 (4.8).

本节剩下的部分对于微观噪声结构模型 (4.4) 以及 X_t 满足扩散模型 (4.5) 时的不同的 $[X]_T$ 进行了回顾, 此时 $[X]_T=\int_0^T\sigma_t^2dt$. 4.3.1–4.3.3 节介绍了使用粗抽样频率的方法. 4.3.4 和 4.3.5 节分别介绍了已实现核估计量和预平均方法. 4.3.6 节介绍了拟最大似然方法.

4.3.1 稀疏抽样方法

稀疏抽样是指将收益率在时间上加总, 从而使用低频收益率的平方值之和来估计 $[X]_T$ 的方法. 具体而言, 与使用 $RV^{(n)}$ 不同, 该方法基于包含收益 $Y_{t_{ki}}-Y_{t_{k(i-1)}}$ 的 $n_k\sim n/k$ 个样本的估计量, 如下:

$$[Y]_T^{(k)}=\sum_{0\leqslant k(i-1)\leqslant n-k}(Y_{t_{ki}}-Y_{t_{k(i-1)}})^2. \tag{4.9}$$

Aït-Sahalia 等 (2005) 以及 Zhang 等 (2005) 证明

$$[Y]_T^{(k)}-[X]_T-2n_kv_\varepsilon \overset{\mathcal{L}}{=} (1+o_p(1))\gamma(n_k)N(0,1), \tag{4.10}$$

其中 $\overset{\mathcal{L}}{=}$ 是指分布等价, $\gamma^2(n_k)$ 是指基于样本量为 n_k 的子样本的两项之和, 其中一项为市场微观结构噪声 ε_i 引起的误差, 另一项为将 $[X]_T$ 中的连续时间 t 离散化为 t_{ki} 时带来的误差. 论文中利用 (4.10) 式, 提出了 $v_\varepsilon\to 0$ 时 n_k 的近似最优选择, 并证明最优子样本的样本量 n^* 与 $v_\varepsilon^{-2/3}$ 同阶.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

4.3.2 基于子样本的平均法

Zhang 等 (2005) 指出, 稀疏抽样方法的缺点在于并没有对于数据进行完全利用. 作为对稀疏抽样方法的改进, 论文中将 $\{t_1,\dots,t_n\}$ 分成 k 个互斥子集 $S_j=\{t_{j-1+ki}:i=1,\dots,n_j\}$, $j=1,\dots,k$. 注意到当 $j=1$ 时, 该问题简化为上一节中的子样本问题, 且有 $n_j\sim n_k\sim n/k$. 类似于 (4.9), 定义 $[Y]_T^{(j,k)}$, 并用 S_1 来替代 S_j. 本质来说, 所有的数据点都在 $[X]_T$ 的平均估计量 $[Y]_T^{\text{avg}}$ 中得到利用, 估计量为

$$[Y]_T^{\text{avg}}=\frac{1}{k}\sum_{j=1}^{k}[Y]_T^{(j,k)}. \tag{4.11}$$

令 $\bar{n}=k^{-1}\sum_{j=1}^{k}n_j$, 论文发现在 (4.5) 及一些其他情况下,

$$[Y]_T^{\text{avg}}-[X]_T-2\bar{n}v_\varepsilon\overset{\mathcal{L}}{=}(1+o_p(1))\left[4\frac{n_k}{k}E[\varepsilon^4]+\frac{4}{3n_k}\int_0^T\sigma_t^4dt\right]^{1/2}N(0,1). \tag{4.12}$$

由于 $\bar{n}\sim n_k\sim\dfrac{n}{k}$, 可得 (4.12) 右边项的分布的标准差最小的 k 值与 $n^{2/3}$ 同阶, 其中 n_k 的阶数为 $n^{1/3}$.

4.3.3 两种时间尺度的方法

上述方法的一个主要问题在于, (4.10) 中的 $2n_kv_\varepsilon$ 和 (4.12) 中的 $2\bar{n}v_\varepsilon$ 都是有偏的, 而由于渐近理论中要求 $n\to\infty$ 和 $n_k\to\infty$, 因此偏差是无界的. 为了解决这个问题, Zhang 等 (2005) 引入了两种时间尺度的方法, 第一种尺度是指 "稀疏" 时间尺度, 对应于上一节中的样本量为 $n_j\sim\dfrac{n}{k}$ 的子样本, 即在 k 个子样本上的均值为 $[Y]_T^{(j,k)}$, 从而得到有偏估计量(4.11)的样本. 第二种尺度是指数据的原始时间尺度, 结合(4.8)式, 可得到 v_ε 的一致估计量为

$$\hat{v}_\varepsilon=\frac{RV^{(n)}}{2n}, \tag{4.13}$$

其偏差为 $2\bar{n}\hat{v}_\varepsilon=\dfrac{\bar{n}}{n}RV^{(n)}$. 利用两种时间尺度的方法, 可以得到如下偏差修正估计量:

$$\widehat{[X]}_T=[Y]_T^{\text{avg}}-\frac{\bar{n}}{n}RV^{(n)}. \tag{4.14}$$

Zhang 等 (2005) 证明在一定情况下,

$$\left(\frac{k}{n_k}\right)^{1/2}(\widehat{[X]}_T-[X]_T)\xrightarrow{\mathcal{L}}N(0,8v_\varepsilon^2), \tag{4.15}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

且 $k \sim n^{2/3}$ 时, 该估计量等于 $n^{1/6}(\widehat{[X]}_T - [X]_T) \xrightarrow{\mathcal{L}} N(0, 8v_\varepsilon^2)$.

4.3.4 核平滑方法: 已实现核估计量

(4.4) 式中的附加噪声项 ε_i 表明, 非参数估计 (如参见 Lai 和 Xing (2008), 第 7 章) 中的平滑方法可以用于消除估计 $[X]_T$ 时 (4.4) 式中的附加噪声项 ε_i 的影响. 然而, 该问题相对于非参数估计中估计条件均值 $E(Y|X)$ 而言难度更大, 原因在于 (a) X 是未观测到的; (b) 该问题需要估计 $[X]_T$, 而不是估计 $f(X) = E(Y|X)$. 但是, 非参数估计中通过局部平均来消除噪声项, 以及将 X 的连续分布进行离散化的方法 (如核平滑方法) 在此处仍然适用. 该方法首先在 Zhou (1996) 以及 Hansen 和 Lunde (2006) 中使用, 但论文中提出的估计量并不一致. Barndorff-Nielsen 等 (2008) 对该问题给出了决定性解法, 论文中沿行 (4.7) 中广泛使用的已实现方差的命名,将 $[X]_T$ 的核估计量称为*已实现核估计量*. 文中将时间区间 $[0,T]$ 分割为 $n = \dfrac{T}{\delta}$ 网格,利用时间汇总的对数收益率数据 $Y^{(\delta)} = \{Y_{i\delta} - Y_{(i-1)\delta}, i = 1, \ldots, n\}$, 并分析各网格起始时间和最终时间对数价格的差别, 从而求得已实现核估计量. 对于 $h = -H, \ldots, -1, 0, 1, \ldots, H$, 定义

$$\gamma_h(Y^{(\delta)}) = \sum_j (Y_{j\delta} - Y_{(j-1)\delta})(Y_{(j-h)\delta} - Y_{(j-h-1)\delta}), \tag{4.16}$$

其中当 $h \geqslant 0$ 时, $\sum_j$ 在 $h+1 \leqslant j \leqslant n$ 上求和, 而在 $h < 0$ 时, $\sum_j$ 在 $1 \leqslant j \leqslant n - |h|$ 上求和. 注意到 $h = 0$ 时, (4.16) 式简化为 (4.7). 文中提出 $[X]_T$ 的已实现核估计量如下:

$$K(Y^{(\delta)}) = \underbrace{\gamma_0(Y^{(\delta)})}_{\text{已实现方差}} + \underbrace{\sum_{h=1}^{H} \kappa\left(\frac{h-1}{H}\right)\left(\gamma_h(Y^{(\delta)}) + \gamma_{-h}(Y^{(\delta)})\right)}_{\text{已实现核修正项}}, \tag{4.17}$$

其中 κ 是一个在 $[0,1]$ 上二重可微的方程, 且满足 $\kappa(0) = 1$ 和 $\kappa(1) = 0$ (这也称为 "平顶" 核函数).

假设 X_t 服从扩散模型 (4.5), 且 ε_i 与过程 $\{X_t\}$ 相互独立, 论文指出, 令 $H = cn^{2/3}$, 当 $n \to \infty$ $\left(\text{因而 } \delta = \dfrac{T}{n} \to 0\right)$ 时,

$$n^{1/6}\{K(Y^{(\delta)}) - [X]_T\} \xrightarrow{\mathcal{L}} N\left(0, 4c^* T \int_0^T \sigma_t^2 dt\right), \tag{4.18}$$

其中 c^* 由 c, $(\kappa'(0))^2 + (\kappa'(1))^2$, $\int_0^1 \kappa^2(x)dx$ 和 $v_\varepsilon^2/(T\int_0^T \sigma_t^4 dt)$ 确定. 另外, 当

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$\kappa'(0)=\kappa'(1)=0$ 时, 取值 $H=cn^{1/2}$, 可以得到更快的收敛速度:

$$n^{1/4}\{K(Y^{(\delta)})-[X]_T\}\xrightarrow{\mathcal{L}}N\left(0,4\tilde{c}T\int_0^T\sigma_t^2dt\right), \tag{4.19}$$

其中 $\tilde{c}$ 也是由 $\int_0^1(\kappa'(x))^2dx$ 和 $\int_0^1(\kappa''(x))^2dx$ 确定. 论文同时分析了 ε_i 由论文 5.4 节中提到的自回归过程所生成的情况. 文中指出, 基于独立同分布 ε_i 模型所得到的已实现核估计量仍然可以使用, 且 "已实现核估计量的收敛速度并不会受序列相关性影响, 但此时渐近分布会有所不同".

Fan 和 Wang (2008) 使用了另一个核平滑方法, 基于交易数据求解 $[X]_T$ 的一致估计量, 在求解时使用市场微观结构噪声模型 (4.4), 其中 X_t 服从更强正则性条件的模型 (4.5).

4.3.5 预平均方法

存在市场微观结构噪声时, 使用 $Y_{t_i},Y_{t_{i+1}},\ldots,Y_{t_{i+k-1}}$ 的均值 $\bar{Y}_{t_i}$ 可以显著减小附加噪声项的方差. 因此, 在 (4.7) 式中, 使用 $\bar{Y}_{t_i}$ 来替代 Y_{t_i} 可以更好地估计半鞅 X_t. Jacod 等 (2009) 假设 X_t 服从扩散过程 (4.5), 且存在附加的正则性条件, 利用如下的 "预平均方法" 对于该观点进行了论证. 论文中定义均匀分布的时间 $t_i=i\delta_n$ (其中 $n=\left\lceil\dfrac{T}{\delta_n}\right\rceil$), 对于特定的 $\theta>0$, 选取 $k_n=\dfrac{\theta}{\sqrt{\delta_n}}+o(\delta_n^{-1/4})$. 在定义了 t_i 和 $k=k_n$ 的取值后, 分别定义 Y_{t_i} 为 Y_i^n, 定义 $\bar{Y}_{t_i}$ 为 $\bar{Y}_i^n$. 令 g 为 $[0,1]$ 上的连续差分 (权重) 函数, 差分为 Lipschitz 差分, 且 $g(0)=g(1)=0$. 令 $\psi_1=\int_0^1(g'(u))^2du$, $\psi_2=\int_0^1g^2(u)du>0$. 论文中提出了 $[X]_T$ 的估计值

$$\begin{aligned}\hat{C}_T=&\frac{\sqrt{\delta_n}}{\theta\psi_2}\sum_{i=1}^{n-k_n+1}\left\{\sum_{j=1}^{k_n-1}g\left(\frac{j}{k_n}\right)(\bar{Y}_{i+j}^n-\bar{Y}_{i+j-1}^n)\right\}^2\\&-\frac{\psi_1\delta_n}{2\theta^2\psi_2}\sum_{i=1}^n(Y_i^n-Y_{i-1}^n)^2,\end{aligned} \tag{4.20}$$

并证明 $n^{-1/4}(\hat{C}_T-[X]_T)$在 $\delta_n\sim\dfrac{T}{n}\to0$ 时, 极限分布为正态分布. (4.20) 式的第二项为偏差修正项. 在偏差修正后, 根据第一个求和项中的移动平均值 $\bar{Y}_{i+j}^n$ 进行预平均估计, 估计值以 $n^{-1/4}$ 的速度收敛于 $[X]_T$.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

4.3.6 从波动率的 MLE 估计量到 $[X]_T$ 的 QMLE 估计量

尽管 4.3.3–4.3.5 节中提到的渐近理论和非参数估计的模拟研究在进行研究时表现良好, 但正如 Gatheral 和 Oomen (2010) 指出, 这些方法相比于其他方法而言, 如相对于其他修正方法, 或基于模拟真实市场的关键特征的 "零智能" 限价委托簿市场的模拟数据所得的已实现方差 (4.7), 理论上并没有显著优势. 论文指出, 即使是在 (4.5) 中的 X_t 被错误定义为 $dX_t = \sigma dB_t$ 时, Hansen 等 (2008) 中的最大似然估计量仍然是最优估计量之一. Xiu (2010) 将拟最大似然估计方法 (QML) 中的统计理论用于分析模型 (4.4) 中的参数 σ 和 v_ε 的最大似然估计量的特征, 在模型中 ε_i 独立同分布, 服从 $N(0, v_\varepsilon)$, 且有 $dX_t = \sigma dB_t$. QML 方法会在 4.7 节的补充材料 4 中具体介绍. 为简化分析, 论文假设 t_i 在 $[0, T]$ 中等间距分布, 满足 $t_i = i\delta$, 其中 $T = n\delta$.

定义 $\Delta_i = Y_{i\delta} - Y_{(i-1)\delta}$ 和 $\boldsymbol{\Delta} = (\Delta_2, \ldots, \Delta_n)^\top$, 则在该参数模型中 $\Delta_i = \sigma(B_{i\delta} - B_{(i-1)\delta}) + \varepsilon_i - \varepsilon_{i-1}$, $\mathrm{Var}(\Delta_i) = \sigma^2\delta + 2v_\varepsilon$, 且模型的对数似然方程为

$$l(v_\varepsilon, \sigma^2) = -\frac{1}{2}\log\det(\boldsymbol{\Omega}) - \frac{n}{2}\log(2\pi) - \frac{1}{2}\boldsymbol{\Delta}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}, \tag{4.21}$$

其中 $\boldsymbol{\Omega}$ 是 MA(1) 三对角带状矩阵:

$$\boldsymbol{\Omega} = \begin{pmatrix} \sigma^2\delta + 2v_\varepsilon & -v_\varepsilon & 0 & \ldots & 0 \\ -v_\varepsilon & \sigma^2\delta + 2v_\varepsilon & -v_\varepsilon & \ddots & \vdots \\ 0 & -v_\varepsilon & \sigma^2\delta + 2v_\varepsilon & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -v_\varepsilon \\ 0 & \ldots & 0 & -v_\varepsilon & \sigma^2\delta + 2v_\varepsilon \end{pmatrix}. \tag{4.22}$$

Xiu (2010) 首次应用中心极限定理证明, 在 ε_i 的真实分布不是正态时, 最大似然估计量 $(\hat{v}_\varepsilon, \hat{\sigma}^2) = \arg\max_{(v_\varepsilon, \sigma^2)} l(v_\varepsilon, \sigma^2)$ 仍然是一致的. 特别地, 当 $n \to \infty$ 时 (此时 $\delta \sim \dfrac{T}{n} \to 0$), $dX_t = \sigma dB_t$ 且 ε_i 独立同分布, 均值为 0, 方差为 v_ε 时, 满足

$$\begin{pmatrix} n^{1/4}(\hat{\sigma}^2 - \sigma^2) \\ \sqrt{n}(\hat{v}_\varepsilon - v_\varepsilon) \end{pmatrix} \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{4.23}$$

其中

$$\boldsymbol{\Sigma} = \begin{pmatrix} 8\sqrt{v_\varepsilon}\sigma^3 T^{-1/2} & 0 \\ 0 & 2v_\varepsilon^2 + \mathrm{cum}_4[\varepsilon] \end{pmatrix}, \tag{4.24}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$\mathrm{cum}_4[\varepsilon]$ 是 ε_i 的四阶累积量. 此处与正态分布情形唯一的不同在于 $\mathrm{cum}_4[\varepsilon] = E(\varepsilon^4) - 3v_\varepsilon^2$ 不一定等于 0. 注意到 $\hat{\sigma}^2$ 以 $n^{-1/4}$ 的速度 (而不是以正常速度 $n^{-1/2}$) 收敛于 σ^2, 原因在于 $X_{i\delta} - X_{(i-1)\delta}$ 的方差为 $\sigma^2\delta \sim \sigma^2 \dfrac{T}{n}$. Xiu (2010) 接下来分析了 X_t 服从扩散过程 (4.5) 时, $\hat{\sigma}^2$ 是否是 $\dfrac{1}{T}\int_0^T \sigma_t^2 dt$ 的一个好的估计量. 在正则条件下, 论文发现当 $n \to \infty$ 时, 满足

$$\begin{pmatrix} n^{1/4}\left(\hat{\sigma}^2 - \dfrac{1}{T}\displaystyle\int_0^T \sigma_t^2 dt\right) \\ \sqrt{n}(\hat{v}_\varepsilon - v_\varepsilon) \end{pmatrix} \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{V}),$$

其中

$$\boldsymbol{V} = \begin{pmatrix} \dfrac{5\sqrt{v_\varepsilon}\displaystyle\int_0^T \sigma_t^4 dt}{T\left(\displaystyle\int_0^T \sigma_t^2 dt\right)^{1/2}} + \dfrac{3\sqrt{v_\varepsilon}\left(\displaystyle\int_0^T \sigma_t^2 dt\right)^{3/2}}{T^2} & 0 \\ 0 & 2v_\varepsilon^2 + \mathrm{cum}_4[\varepsilon] \end{pmatrix};$$

具体可参见论文的 4.2 节和 4.3.1 节. 论文的 4.3.2 节和 4.3.3 节中指出该结果也适用于随机抽样时间 t_i 以及非高斯和序列相关的 ε_i.

4.4 多资产协变差估计

附录 A 中的连续时间鞅理论说明, 二次变差 $[X]_T$ 可以推广到两个半鞅过程 X_t 和 $\tilde{X}_t$ 的协变差中, 其中两个半鞅过程分别对应于投资组合中两个不同风险资产的有效对数价格的时间序列. 本节讨论了求解存在微观结构噪声和异步性的多资产交易中的 $[X, \tilde{X}]_T$ 的问题.

4.4.1 异步性和 Epps 效应

自 20 世纪 60 年代以来, 交易过程 (即标的资产的供给和需求、新闻效应和信息传输机制) 中的异步性对于多元金融交易数据的影响一直是金融计量经济学中的重要问题. 具体参见 Fisher (1966) 和 Campbell 等 (1997) 中的综述. Scholes 和 Williams (1977) 基于同一天中两个资产的最后一笔交易并不会同时发生的事实, 以及从属随机过程理论 (参见 2.6.2 节), 证明根据日收盘价数据计算得到的 CAPM 模型参数的最大似然估计值是向下有偏的. 利用工具变量方法, 论文提出了 β 的估计值, 以解决日收盘价的异步性问题. 在此之

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

后, Epps (1979) 利用实证研究, 说明缩短抽样周期可以使不同股票的对数收益率之间的瞬时相关关系趋向于 0 方向减小. 这个现象称为 Epps 效应, 也可以适用于汇率收益中, 具体参见 Guillaume 等 (1997) 和 Muthuswamy 等 (2001). 在市场微观结构噪声模型 (4.4) 中, 时间聚合技术被用于估计 $[X]_T$. 该技术也可以拓展用于同步不同资产的交易数据中. 在接下来的章节中, 介绍了同步技术, 并提出了估计量 $[X, \tilde{X}]_T$ 来消除 Epps 效应.

4.4.2 同步技术

对于 $i = 1, \ldots, M$, 定义 $Y_{i,t}$ 为时间 $t \in [0, T]$ 时第 i 个资产的对数交易价格. 为了同步这 M 个资产的数据, Aït-Sahalia 等 (2010) 定义了广义抽样时间 τ_j, $1 \leqslant j \leqslant n$, 满足

(a) $0 = \tau_0 < \tau_1 < \cdots < \tau_n = T$,

(b) $(\tau_{j-1}, \tau_j] \cap \{t_{i,k} : k = 1, \ldots, n_i\} \neq \varnothing$ 对于所有的 $i = 1, \ldots, M$,

(c) $\max\limits_{1 \leqslant j \leqslant n} \delta_j \to 0$ 依概率收敛, 其中 $\delta_j = \tau_j - \tau_{j-1}$.

对于 (b), 在 $i = 1, \ldots, M$ 时, 存在时间区间 $\tilde{t}_{i,j} \in (\tau_{j-1}, \tau_j]$, 在该区间中 $Y_{i,\tilde{t}_{i,j}}$ 可观测, 则可得到同步数据集 $\{Y^{\tau}_{i,\tau_j} : i = 1, \ldots, M; j = 1, \ldots, n\}$ 满足 $Y^{\tau}_{i,\tau_j} = Y_{i,\tilde{t}_{i,j}}$. 多种同步性方法被用于定义 τ_j. 举例来说, Zhang (2011) 提出的"前笔交易"方法对应于任意 j, 满足 $\delta_j \equiv \delta$ 的情况. Barndorff-Nielsen 等 (2011) 提出的 "更新时间" 方法, 对于 τ_j 进行了迭代定义如下:

$$\tau_{j+1} = \max_{1 \leqslant i \leqslant M} t_{i, N_i(\tau_j)+1},$$

其中 $\tau_1 = \max\{t_{i,1} : i = 1, \ldots, M\}$, $N_i(t)$ 是到时间 t 时, 第 i 个资产观测到的数据个数. "前笔交易" 和 "更新时间" 方法对于第 i 个资产在 τ_j 时或 τ_j 前的一笔交易, 都记为 Y^{τ}_{i,τ_j}.

正如 4.4.1 节指出, 日收盘价数据可以通过令 τ_j 为市场收盘时间来获取.[①] 在所有的同步方法中, 都不可避免地会遗漏一部分数据. 当其中一个资产是非流动性资产时, 数据缺失的比例会特别显著. 在使用广义抽样时间的所有方法中, "更新时间" 方法对于数据的利用率最高. 同时, 我们也需要与选取前一笔交易数据不同, 而是有些任意地选取交易价格的其他广义抽样方法进行研究. 对该问题的更多讨论可参见 4.7 节的补充材料 9.

① 与选取前笔交易不同, 香港交易所将收盘价定义为交易日最后一分钟的五个名义价格 (15 秒一次) 的中位数. 此处使用名义价格而不是交易价格, 是因为即使非流动性资产在最后一分钟没有交易, 也能得到相关数据.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

4.4.3 协方差和相关系数估计的 QMLE 方法

考虑 $M=2$ 的情形. 利用 4.4.2 节中提到的广义抽样时间方法, 得到同步数据 $\{(Y^{\tau}_{1,\tau_j}, Y^{\tau}_{2,\tau_j})\}_{j=1}^{n}$. 对于 $i=1,2$; $j=1,\ldots,n_i$, 有

$$Y^{\tau}_{i,\tau_j} = X^{\tau}_{i,\tau_j} + \varepsilon_{i,j}, \tag{4.25}$$

其中 $\varepsilon_{i,j}$ 指代独立同分布的微观结构噪声项, 其均值为 0, 方差为 $v_{\varepsilon,i}$, 则 $\{\varepsilon_{1,j}\}$ 和 $\{\varepsilon_{2,j}\}$ 也相互独立. 注意到

$$[X_1, X_2]_T = \frac{1}{4}\left\{[X_1+X_2]_T + [X_1-X_2]_T\right\}, \tag{4.26}$$

Aït-Sahalia 等 (2010) 将 4.3.6 节中的拟最大似然估计 (QMLE) 方法应用于估计 $Z_{1,\tau_j} = Y^{\tau}_{1,\tau_j} + Y^{\tau}_{2,\tau_j}$ 和 $Z_{2,\tau_j} = Y^{\tau}_{1,\tau_j} - Y^{\tau}_{2,\tau_j}$, 则可得

$$\widehat{[X_1, X_2]}_T = \frac{1}{4}\left\{\widehat{[X_1+X_2]}_T + \widehat{[X_1-X_2]}_T\right\}.$$

类似于 (4.5) 式, 假设有效价格对于 $i=1,2$, 服从扩散过程 $dX_{i,t} = \mu_{it}dt + \sigma_{it}dB_{it}$, 其中 B_{1t} 和 B_{2t} 是相关的布朗运动过程, 且 $E(dB_{1t}dB_{2t}) = \rho_t dt$. 因此, 在 τ_j 为平均间隔时 (即对于任意 j, $\tau_j - \tau_{j-1} \equiv \delta$), 由 4.3.6 节中的渐近理论可知当 $n\to\infty$ 时,

$$n^{1/4}\left(\frac{1}{T}\widehat{[X_1, X_2]}_T - \frac{1}{T}\int_0^T \rho_t\sigma_{1t}\sigma_{2t}dt\right) \xrightarrow{\mathcal{L}} N(0, v_1).$$

结合 Delta 方法 (参考 Lai 和 Xing (2008), 59 页) 可知, 当 $n\to\infty$ 时,

$$n^{1/4}\left(\frac{\widehat{[X_1, X_2]}_T}{\sqrt{\widehat{[X_1]}_T\widehat{[X_2]}_T}} - \frac{\int_0^T \rho_t\sigma_{1t}\sigma_{2t}dt}{\sqrt{\int_0^T \sigma_{1t}^2dt\int_0^T \sigma_{2t}^2dt}}\right) \xrightarrow{\mathcal{L}} N(0, v_2).$$

参见 Aït-Sahalia 等 (2010), 该论文的附录中也给出了 v_1 和 v_2 的显式解. 当 τ_j 满足 $\delta_j = \tau_j - \tau_{j-1}$ 独立同分布, 且均值为 $\bar{\delta}$ 时, Aït-Sahalia 等 (2010) 证明 $\widehat{[X_1, X_2]}_T$ 和 $\dfrac{\widehat{[X_1, X_2]}_T}{\sqrt{\widehat{[X_1]}_T\widehat{[X_2]}_T}}$ 分别是协变过程 $[X_1, X_2]_T$ 和相关过程 $\dfrac{[X_1, X_2]_T}{\sqrt{[X_1]_T[X_2]_T}}$ 的一致估计量, 收敛速度为 $O_p(\bar{\delta}^{1/4})$. 在 4.7 节的补充材料 1–4 中, 利用交易数据的 MA(1) 相关结构, 对于 QMLE 在实际中的应用, 以及其与 Hansen 等 (2008) 的联系进行了进一步的阐述.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

4.4.4 多元已实现核和双尺度的估计量

Zhang (2011) 利用 "前笔交易" 方法, 将数据进行同步, 从而得到 $\boldsymbol{Y}_{\boldsymbol{\tau}} = \{\boldsymbol{Y}_{\tau_j} : j = 1, \ldots, n\}$, 其中 $\boldsymbol{Y}_{\tau_j} = (Y^{\tau}_{1,\tau_j}, Y^{\tau}_{2,\tau_j})$. 类似于 4.3.1–4.3.3 节, 定义稀疏抽样估计量

$$[Y^{\tau}_1, Y^{\tau}_2]^{(k)}_T = \sum_{0 \leqslant k(i-1) \leqslant n-k} \left(Y^{\tau}_{1,\tau_{ki}} - Y^{\tau}_{1,\tau_{k(i-1)}}\right)\left(Y^{\tau}_{2,\tau_{ki}} - Y^{\tau}_{2,\tau_{k(i-1)}}\right),$$

平均估计量

$$[Y^{\tau}_1, Y^{\tau}_2]^{\text{avg}}_T = \frac{1}{k}\sum_{j=1}^{k}[Y^{\tau}_1, Y^{\tau}_2]^{(j,k)}_T,$$

和双尺度已实现方差 (TSRC) 估计量

$$\widehat{[X_1, X_2]_T}^{\text{TSRC}} = [Y^{\tau}_1, Y^{\tau}_2]^{\text{avg}}_T - \frac{\bar{n}}{n}\text{Cov}^{(n)}, \tag{4.27}$$

其中 $\bar{n} = k^{-1}\sum_{j=1}^{k} n_j$,

$$\text{Cov}^{(n)} = \sum_{j=1}^{n}\left(Y^{\tau}_{1,\tau_j} - Y^{\tau}_{1,\tau_{j-1}}\right)\left(Y^{\tau}_{2,\tau_j} - Y^{\tau}_{2,\tau_{j-1}}\right).$$

类似于单变量情形, 论文表明 TSRC 是一致估计量, 且当选取 $k \sim n^{2/3}$ 时, 收敛速度的阶数为 $n^{1/6}$.

Barndorff-Nielsen 等 (2011) 利用 "更新时间" 方法对于数据进行同步, 得到 $\boldsymbol{Y}_{\boldsymbol{\tau}} = \{\boldsymbol{Y}_{\tau_j} : j = 1, \ldots, n\}$. 论文中定义 (4.16) 的矩阵一般化形式为

$$\boldsymbol{\Gamma}_h\left(\boldsymbol{Y}_{\tau}\right) = \sum_{j=h+1}^{n}\left(\boldsymbol{Y}_{\tau_j} - \boldsymbol{Y}_{\tau_{j-1}}\right)\left(\boldsymbol{Y}_{\tau_{j-h}} - \boldsymbol{Y}_{\tau_{j-h-1}}\right)^{\top}, \ h \geqslant 0 \tag{4.28}$$

且在 $h < 0$ 时, $\boldsymbol{\Gamma}_h(\boldsymbol{Y}_{\tau}) = [\boldsymbol{\Gamma}_{-h}(\boldsymbol{Y}_{\tau})]^{\top}$. 定义 $[X_1, X_2]_T$ 的多元已实现核 (MRK) 估计量为

$$\boldsymbol{K}\left(\boldsymbol{Y}_{\tau}\right) = \sum_{h=-n}^{n}\mathcal{K}\left(\frac{h}{H}\right)\boldsymbol{\Gamma}_h\left(\boldsymbol{Y}_{\tau}\right), \tag{4.29}$$

其中核函数 $\mathcal{K}(\cdot)$ 在单变量情形下与 4.3.4 节中的核函数 $\kappa(\cdot)$ 类似. 特别地, 为了保证 $\boldsymbol{K}(\boldsymbol{Y}_{\tau})$ 的正定性, Barndorff-Nielsen 等 (2011) 只选取满足如下条件的 $\mathcal{K}(\cdot)$:

(a) $\mathcal{K}(0) = 1, \mathcal{K}'(0) = 0$;

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

(b) $\mathcal{K}$ 二阶可导, 且导数连续;

(c) $\int_0^\infty \mathcal{K}^2(x)dx < \infty$, $\int_0^\infty [\mathcal{K}'(x)]^2 dx < \infty$, 且 $\int_0^\infty [\mathcal{K}''(x)]^2 dx < \infty$;

(d) 对于任意 $\lambda \in \mathbb{R}$, 满足 $\int_{-\infty}^{\infty} \mathcal{K}(x)\exp\{\sqrt{-1}x\lambda\}dx \geqslant 0$.

根据 Bochner 定理, 条件 (d) 可以保证 $\boldsymbol{K}(\boldsymbol{Y}_\tau)$ 的正定性, 具体参见 4.7 节的补充材料 11. 条件 (a) 中, 由 $\mathcal{K}(0) = 1$ 可知, $\boldsymbol{\Gamma}_0(\boldsymbol{Y}_\tau)$ 的权重为单位权重. $\mathcal{K}(0) = 1$ 和 $\mathcal{K}'(0) = 0$ 表明, 当 $|h|$ 很小时, $\boldsymbol{\Gamma}_h(\boldsymbol{Y}_\tau)$ 的权重值接近 1. 在类似于 4.3.4 节的假设下, Barndorff-Nielsen 等 (2011) 证明了 MRK 估计量的渐近正态性和一致性, 并指出在 $H = cn^{3/5}$ 时, 收敛速度的阶数为 $n^{1/5}$. 论文中同时指出, 对于序列相关性不同的 $(\varepsilon_{1,j}, \varepsilon_{2j})$, MRK 统计量是稳健的.①

前面提到的估计量都需要基于 4.4.2 节中的显式同步方法, 而 Hayashi 和 Yoshida (2005) 提出了如下的 $[X_1, X_2]_T$ 的估计量, 该方法可以利用所有的可得数据:

$$\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \left(Y_{1,t_{1,i}} - Y_{1,t_{1,i-1}}\right)\left(Y_{2,t_{2,j}} - Y_{2,t_{2,j-1}}\right)\mathbf{1}_{\{O_{ij}\neq\varnothing\}}, \tag{4.30}$$

其中 $O_{ij} = (t_{1,i-1}, t_{1,i}] \cap (t_{2,j-1}, t_{2,j}]$. Aït-Sahalia 等 (2010) 指出, (4.30) 式可以写成:

$$\sum_{i=1}^{n_1} \left(Y_{1,t_{1,i}} - Y_{1,t_{1,i-1}}\right)\left(Y_{2,t_{2,i+}} - Y_{2,t_{2,(i-1)-}}\right),$$

其中 $t_{2,i+} = \min\{t_{2,k} : t_{2k} \geqslant t_{1,i}\}$, $t_{2,i-} = \max\{t_{2,k} : t_{2k} \leqslant t_{1,i}\}$. 由此可知, 存在一个可以将 $(t_{2,(j-1)+}, t_{2,j-})$ 中的数据点忽略的隐式同步方法. 另外, Hayashi 和 Yoshida (2005, 2008) 中要求不存在微观结构噪声项, 从而可以建立相关的渐近理论, 从而证明在异步交易的到达时间服从独立泊松分布时, 估计量 (4.30) 是 $\sqrt{n}$ 一致的. 然而, 正如 Voev 和 Lunde (2007) 指出, (4.30) 在市场微观结构模型 (4.25) 下是有偏的, 且其方差在观测样本量趋于无穷时是发散的.

4.5 傅里叶方法

正如 4.4.2 节指出, 当 M 个资产中存在非流动性资产时, 同步性方法会导致大量数据缺失. 另外, 由于同步性方法使用时间区间为 $(\tau_j, \tau_{j+1}]$, 其中 $\tau_{j+1} - \tau_j = \delta$, 当 δ 很小时, 由同步交易价格计算得到的对数收益率大部分都是 0. 为了直接从异步交易数据中估计协变异, Malliavin 和 Mancino (2002) 提出

① Barndorff-Nielsen 等 (2011) 中建议使用估计量的均值以控制初始时间和终值时间点的边缘效应. 使用该方法时, 需要对于 (4.28) 式中求和项的上确界进行相应的调整.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

了傅里叶方法. 在本节中, 首先介绍 $M = 1$ 情况下的傅里叶方法和相关的理论, 接下来拓展到多资产情况.

4.5.1 $[X]_T$ 和瞬时波动率的傅里叶估计

Malliavin 和 Mancino (2002, 2009) 假设价格变化服从过程 (4.5), 其中 μ_t 和 σ_t 为满足如下条件的适应性过程:

$$E\left[\int_0^T \mu_t^2 dt\right] < \infty, \quad E\left[\int_0^T \sigma_t^4 dt\right] < \infty. \tag{4.31}$$

假设市场不存在微观结构噪声项, 则在各个时间点 $t_1 < \cdots < t_n \in [0, T]$ 的观测值满足 $Y_t = X_t$. $[X]_T = \int_0^T \sigma_t^2 dt$ 称为 X_t 的积分方差 (或二次变异), σ_t 则称为瞬时波动率 (或即时波动率). 令 $\Sigma_t = \sigma_t^2$, 则论文的主要思路为: (a) 利用 Itô 随机积分公式, 将 Σ_t 的傅里叶系数表达为 dX_t 的函数; (b) 对于定义傅里叶系数的随机积分方程的积分项进行分段常函数近似; (c) 对于 Σ_t 的傅里叶展开序列进行有限和逼近.

首先, 定义最终时间 T 为 2π (需要的情况下可以进行比例调节), 则傅里叶系数可以定义为

$$\begin{aligned} a_0(dX) &= \frac{1}{2\pi}\int_0^{2\pi} dX_t, \\ a_k(dX) &= \frac{1}{\pi}\int_0^{2\pi} \cos(kt) dX_t, \quad \text{对于} k > 0, \\ b_k(dX) &= \frac{1}{\pi}\int_0^{2\pi} \sin(kt) dX_t, \quad \text{对于} k \geqslant 0. \end{aligned} \tag{4.32}$$

由于 X_t 只在离散时间点 t_i 可观测, 对于 (4.32) 式中的积分项进行 Lebesgue-Stieltjes 近似 (对于任意的 $s \in (t_i, t_{i+1}]$, 令 $\tilde{X}_s = X_{t_i}$), 可得如下估计量:

$$\hat{a}_0(dX) = \frac{X_{2\pi} - X_0}{2\pi}, \tag{4.33}$$

$$\hat{a}_k(dX) = \frac{X_{2\pi} - X_0}{\pi} + \frac{1}{\pi}\sum_{i=1}^{n-1} \left[\cos(kt_i) - \cos(kt_{i+1})\right] X_{t_i}, \quad \text{对于} k > 0, \tag{4.34}$$

$$\hat{b}_k(dX) = \frac{1}{\pi}\sum_{i=1}^{n-1} \left[\sin(kt_i) - \sin(kt_{i+1})\right] X_{t_i}, \quad \text{对于} k \geqslant 0. \tag{4.35}$$

将 Itô 公式应用于扩散过程 (4.5), 则可将 $\Sigma_t = \sigma_t^2$ 的傅里叶系数写成如下

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

形式:

$$a_0(\Sigma) = \underset{N_x\to\infty}{\text{p-lim}} \frac{\pi}{N_X+1} \sum_{j=0}^{N_X} \left[a_j^2(dX) + b_j^2(dX)\right], \tag{4.36}$$

$$a_k(\Sigma) = \underset{N_x\to\infty}{\text{p-lim}} \frac{2\pi}{N_X+1} \sum_{j=0}^{N_X} \left[a_j(dX)a_{j+k}(dX)\right], \quad \text{对于 } k>0, \tag{4.37}$$

$$b_k(\Sigma) = \underset{N_x\to\infty}{\text{p-lim}} \frac{2\pi}{N_X+1} \sum_{j=0}^{N_X} \left[a_j(dX)b_{j+k}(dX)\right], \quad \text{对于 } k\geqslant 0, \tag{4.38}$$

其中 p-lim 表示依概率收敛. 求解 (4.36)–(4.38) 的关键步骤在 4.8 节的问题 4 中给出. 将 $a_i(dX)$ 和 $b_j(dX)$ 分别替代为 $\hat{a}_i(dX)$ 和 $\hat{b}_j(dX)$, 并对于 (4.36)–(4.38) 中的 $\sum_{i=0}^{N_X}$ 求和项进行有限和近似, 可得到 Σ_t 的傅里叶系数中的 $\hat{a}_0(\Sigma)$, $\hat{a}_k(\Sigma)$ 和 $\hat{b}_k(\Sigma)$, 并可以得到聚合方差 $\widehat{[X]}_T^F = 2\pi\hat{a}_0(\Sigma)$ 的傅里叶估计. 注意到, 利用 Fourier-Fejér 反演公式

$$\Sigma_t = \underset{N_S\to\infty}{\text{p-lim}} \sum_{k=0}^{N_S} \left(1-\frac{k}{N_S}\right) \left[a_k(\Sigma)\cos(kt) + b_k(\Sigma)\sin(kt)\right] \tag{4.39}$$

可以得到 $t\in(0,2\pi)$ 上的瞬时波动率. Malliavin 和 Mancino (2002) 在不存在市场微观结构噪声的情况下, 给出了 Σ_t 的估计值为

$$\hat{\Sigma}_t = \sum_{k=0}^{N_S} \left(1-\frac{k}{N_S}\right) \left[\hat{a}_k(\Sigma)\cos(kt) + \hat{b}_k(\Sigma)\sin(kt)\right], \tag{4.40}$$

论文指出, 当合理选取 N_X 和 N_S 时, $\hat{\Sigma}_t$ 对于 X_t 市场微观结构噪声是稳健的, 该结果会在下面具体解释.

4.5.2 傅里叶估计量的统计特征

Malliavin 和 Mancino (2009) 研究表明, $\widehat{[X]}_T^F$ 与 $RV^{(n)}$ 有如下的相关关系. 注意到 $\widehat{[X]}_T^F$ 可以写成如下形式:

$$\widehat{[X]}_T^F = \sum_{i=1}^{n}\sum_{j=1}^{n} D_{N_X}(t_i-t_j)(X_{t_i}-X_{t_{i-1}})(X_{t_j}-X_{t_{j-1}}), \tag{4.41}$$

其中

$$D_{N_X}(t) = \frac{1}{2N_X+1} \frac{\sin\left[\left(N_X+\frac{1}{2}\right)t\right]}{\sin\left(\frac{t}{2}\right)} \tag{4.42}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

为调整后 Dirichlet 核. 将 (4.41) 式中 $i=j$ 项与 $i\neq j$ 项分离, 则可将 (4.41) 式改写为

$$\widehat{[X]}_T^F = RV^{(n)} + \sum_{i=1}^{n}\sum_{\substack{j=1;\\ i\neq j}}^{n} D_{N_X}(t_i - t_j)(X_{t_i} - X_{t_{i-1}})(X_{t_j} - X_{t_{j-1}}).$$

对于任意 i, 当 $t_i - t_{i-1} = \delta$ 时, 上式中双重求和项可以改写为

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{\substack{j=1;\\ i\neq j}}^{n} D_{N_X}((i-j)\delta)(X_{i\delta} - X_{(i-1)\delta})(X_{j\delta} - X_{(j-1)\delta})\\ =&\sum_{h=1}^{H} D_{N_X}(h\delta)\left[\gamma_h(X) + \gamma_{-h}(X)\right],\end{aligned}$$

其中 γ_h 为 (4.16) 式中定义的 h 阶自回归方差项. 因此, 双重求和项与 (4.17) 式中已实现核估计量中的修正项是类似的. 两者主要的区别在于傅里叶估计量中的 Dirichlet 核与频率次数 N_X 和观测样本之间的时间跨度有关. 因此, $\widehat{[X]}_T^F$ 与已实现核估计量类似, 但其并不需要样本均匀分布.

Dirichlet 核也可以用于分析存在 (4.4) 式中的微观结构噪声时, $\widehat{[X]}_T^F$ 的偏差. 对于该问题, Mancino 和 Sanfelici (2008) 指出:

$$E\left(\widehat{[X]}_T^F - [X]_T\right) = 2nv_\varepsilon\left[1 - D_{N_X}\left(\frac{2\pi}{n}\right)\right]. \tag{4.43}$$

由上可知, 在条件 (4.4) 下, $\widehat{[X]}_T^F$ 是有偏的, 但是随着 $\dfrac{N_X^2}{n}\to 0$ 和 $n\to\infty$, 该偏差趋近于 0. 论文中也对于 $\mathrm{MSE} := E(\widehat{[X]}_T^F - [X]_T)^2$ 进行分析, 得到了相关特性. 在求解 MSE 时, 基于 (4.4) 式中的微观结构噪声模型, 然后在不存在微观结构噪声的情况下求解. 论文同时发现, 当 $\dfrac{N_X^2}{n}\to 0$ 和 $n\to\infty$ 时, 存在噪声项时 MSE 的值比不存在噪声项时 MSE 的值大 $C + o(1)$, 其中 C 是与 v_ε, $E(\varepsilon^4)$ 和 $[X]_T$ 相关的一个正的常数. 为了增加估计量在有限样本时的表现, 论文提出了选取 N_X 的一个实证准则, 该准则与 Bandi 和 Russell (2006) 以及 Barndorff-Nielsen 等 (2008) 中已实现核估计中窗宽的选择方法类似.

傅里叶瞬时波动率估计量是根据 Fourier-Féjer 反演公式 (4.39) 所得, 根据该关系可以得到多个近似结果. 举例来说, Malliavin 和 Mancino (2009) 证明在一定的正则条件下, 当 Σ_t 连续时, 满足

$$\sup_{t\in[0,2\pi]} |\hat{\Sigma}_t - \Sigma_t| \to 0.$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

具体可参见该论文的推论 3.5. Mancino 和 Recchioni (2015) 也证明了在一些附加假设下, 对于固定的 t, $\hat{\Sigma}_t$ 的渐近正态性. 一些其他的理论和实证研究表明, 傅里叶估计量即使在存在内生和序列相关的微观结构噪声项时也是稳健的. 相对于其他的波动性估计量而言, 傅里叶估计量对未来的预测结果相当或更好, 参见 Mancino 和 Sanfelici (2008, 2011a,b) 以及 Mancino 和 Recchioni (2015).

4.5.3 瞬时协方差的傅里叶估计量

假设投资组合的对数价格 X_{it}, $i=1,\ldots,M$, $0\leqslant t\leqslant T$, 满足

$$dX_{it}=\mu_{i,t}dt+\sum_{k=1}^{d}\sigma_{k,t}^{(i)}dB_{k,t}, \tag{4.44}$$

其中 $\boldsymbol{B}_t=(B_{1,t},\ldots,B_{d,t})^\top$ 是一个 d 维布朗运动. 类似于 (4.31) 式, 假设对于 $i=1,\ldots,M$ 和 $k=1,\ldots,d$,

$$E\left[\int_0^T(\mu_{i,t})^2dt\right]<\infty,\quad E\left[\int_0^T(\sigma_{k,t}^{(i)})^2dt\right]<\infty.$$

定义 $\boldsymbol{\Sigma}_t^{1/2}$ 为对数价格过程 $\boldsymbol{X}_t=(X_{1,t},\ldots,X_{M,t})^\top$ 的瞬时协方差矩阵, $\Sigma_t^{(i,j)}$, $i,j=1,\ldots,d$ 为矩阵 $\boldsymbol{\Sigma}_t$ 的第 (i,j) 个元素. 对于非负定矩阵, 其平方根可以由奇异值分解得到 (参见 Lai 和 Xing (2008, 43 页)). 在模型 (4.44) 下, $\Sigma_t^{(i,j)}=\sum_{k=1}^{d}\sigma_{k,t}^{(i)}\sigma_{k,t}^{(j)}$, 且相应的二次协方差为

$$[X_i,X_j]_T=\int_0^T\Sigma_t^{(i,j)}dt.$$

与单变量情形类似, 此处令 $T=2\pi$. 因此, 对于 $i=1,\ldots,M$, dX_i 的傅里叶系数可以根据 (4.33)、(4.34) 和 (4.35)式求得. 类似地, 瞬时协方差过程 $\Sigma_t^{(i,j)}$ 的傅里叶系数可以根据 dX_i 的傅里叶系数求得, 如下:

$$\begin{aligned}
a_0(\Sigma^{(i,j)})&=\operatorname*{p-lim}_{N\to\infty}\frac{\pi}{N_X+1}\sum_{q=0}^{N_X}\left[a_q(dX_i)a_q(dX_j)+b_q(dX_i)b_q(dX_j)\right],\\
a_r(\Sigma^{(i,j)})&=\operatorname*{p-lim}_{N\to\infty}\frac{\pi}{N_X+1}\sum_{q=0}^{N_X}\left[a_q(dX_i)a_{q+r}(dX_j)+a_q(dX_j)a_{q+r}(dX_i)\right],\\
b_s(\Sigma^{(i,j)})&=\operatorname*{p-lim}_{N\to\infty}\frac{\pi}{N_X+1}\sum_{q=0}^{N_X}\left[a_q(dX_i)b_{q+s}(dX_j)+a_q(dX_j)b_{q+s}(dX_i)\right],
\end{aligned}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 $r > 0$ 且 $s \geqslant 0$. 另外, 类似于单变量情形, Malliavin 和 Mancino (2002) 利用 Fourier-Fejér 反演定理的多元形式, 提出了如下的傅里叶瞬时协方差估计量:

$$\hat{\Sigma}_t^{(i,j)} = \sum_{k=0}^{N_S} \left(1 - \frac{k}{N_S}\right) \left[\hat{a}_k(\Sigma^{(i,j)}) \cos(kt) + \hat{b}_k(\Sigma^{(i,j)}) \sin(kt)\right], \tag{4.45}$$

其中 $t \in (0, 2\pi)$.

在本节的最后, 我们对于 Malliavin 和 Mancino (2002,2009) 两篇论文进行相关评述. 定义 $[\widetilde{X_i, X_j}]_T$ 为基于同步方法对于聚合协方差的估计量. 根据 $[\widetilde{X_i, X_j}]_T$ 求解瞬时估计量的一个显然的方法是:

$$\tilde{\Sigma}_t^{(i,j)} = \frac{[\widetilde{X_i, X_j}]_{t+h} - [\widetilde{X_i, X_j}]_t}{h}, \tag{4.46}$$

但是由于 $h > 0$, 该方法受到很大的数值可行性限制. 该问题产生的根本原因在于同步方法中 "差分化过程" 这个隐式概念. 注意到在傅里叶方法中, 并不需要进行同步性, 因为该方法的估计量是基于 "聚合" 变量所得. Malliavin 和 Mancino (2002, 2009) 将傅里叶方法的优异表现归结于在估计瞬时协方差时, 使用聚合变量而不是差分变量.

4.6 与 TAQ 相关的其他计量经济学模型

NYSE 交易和报价 (TAQ) 数据库包含所有在 NYSE, AMEX 和 NASDAQ 上市的股票的日内交易数据 (包括交易和报价数据). 在本节中, 我们提到的 TAQ 数据是指与上述交易所或其他类似交易所得到的日内交易数据, 并讨论了上一节中 (4.5) 和 (4.6) 外的高频交易计量经济学中的几类模型. 首先, 讨论时间间隔为 $t_i - t_{i-1}$ 的模型. 在 4.4.4 节的最后一段中提到, Hayashi 和 Yoshida (2005, 2008) 假设到时间 t 时发生交易的次数 $N(t)$ 服从泊松分布, 即时间间隔独立同分布, 服从指数分布. 然而, 由于资本市场只在交易时间内发生交易, 在交易日开盘和收盘时更容易发生交易, 这个与时间间隔独立同分布的特性相违背. 在 4.6.1 节中, 利用自回归条件间隔 (ACD) 模型对于交易日各交易之间的时间间隔时间序列进行建模. 在 4.6.2 节中, 利用自激励点过程 (SEPP) 对于 $N(t)$ 进行建模. 接下来, 使用标值点过程将交易价格的变化分解为两个部分: (a) 交易发生次数的点过程; (b) 交易价格的变动模型, 即具体而言,

$$P_t - P_0 = \sum_{i=1}^{N(t)} D_i, \tag{4.47}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 P_t 为时间 t 时的交易价格, 而 D_i 为第 i 次交易时价格的变动. 4.6.3 节讨论了 D_i 的计量经济学建模. 4.6.4 和 4.6.5 节讨论了 $N(t)$, D_i 或其他标值点的联合分布模型. 4.6.6 节讨论了将 TAQ 数据的计量经济学模型用于未来波动率的方法, 并将其应用于投资组合优化、衍生品定价和风险管理. 4.6.7 节利用带跳的半鞅理论对于有效价格过程进行了建模.

4.6.1 交易之间间隔的 ACD 模型

对于 4.2.2 节中提到的交易时间 $t_i \in [0,T]$, Engle 和 Russell (1998) 提出了如下模型来对于交易时间间隔 $\delta_i = t_i - t_{i-1}$ 进行建模和分析. 在分析时, 选用日内交易数据, 0 和 T 分别指当天的开盘时间和收盘时间, 共收集了连续 m 个交易日的数据. 论文中首先利用如下的非参数回归模型, 分析了日内交易时间间隔的周期性:

$$\log(\delta_i) = g(t_i) + \eta_i, \tag{4.48}$$

其中 g 为平滑样条或回归样条 (参见 Lai 和 Xing (2008, 7.2 节)), η_i 独立同分布, 且均值为 0, 方差有限. 定义调整后的时间间隔为 $\delta_i^* = \dfrac{\delta_i}{\exp\{\hat{g}(t_i)\}}$, ACD 模型假设其服从 $\delta_i^* = \psi_i \varepsilon_i$ 的模型, 其中 ε_i 是独立同分布的正随机变量, 均值为 1. 当 ψ_i 服从如下 GARCH(r,s) 模型:

$$\psi_i = \omega + \sum_{j=1}^{r} \gamma_j \delta_{i-j}^* + \sum_{j=1}^{s} \beta_j \psi_{i-j}, \tag{4.49}$$

且参数 $\omega, \gamma_1, \ldots, \gamma_r, \beta_1, \ldots, \beta_s$ 非负时, ACD 模型即为 ACD(r,s) 模型. 对于 ε_i 的分布, 可以选取均值为 1 的指数分布, 或者形状参数为 α 的标准 Weibull 分布 (参见 Lai 和 Xing (2008, 第 293 页)), 分别称为指数 ACD (EACD) 模型或 Weibull ACD (WACD) 模型. 类似于 GARCH (r,s) 模型, ACD (r,s) 可以写成具有非高斯鞅差分新息项的 ARMA 模型. 特别地, 当满足宽平稳条件 $\sum_{j=1}^{r} \gamma_j + \sum_{j=1}^{s} \beta_j < 1$ 时,

$$E(\delta_i^*) = \frac{\omega}{1 - (\sum_{j=1}^{r} \gamma_j + \sum_{j=1}^{s} \beta_j)}.$$

Engle 和 Russell 利用独立同分布的 $\varepsilon_i = \dfrac{\delta_i^*}{\psi_i}$ 的联合分布函数, 得到了 EACD(r,s) 模型下 $(\omega, \gamma_1, \ldots, \gamma_r, \beta_1, \ldots, \beta_s)$ 的似然函数, 和 WACD (r,s) 模型下 $(\alpha, \omega, \gamma_1, \ldots, \gamma_r, \beta_1, \ldots, \beta_s)$ 的似然函数.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

在宽平稳或其他正则条件下, 可以求得最大 (条件) 似然估计量的渐近理论, 具体参见 Engle 和 Russell (1998). 对于 EACD(1,1) 模型, 令 $u_i = d_i\sqrt{\delta_i^*}$, 其中 d_i 与 δ_i^* 独立, 且以 0.5 的概率分别取值为 -1 和 1. 注意到 u_i 的期望为 0, 而方差为 ψ_i. 将 d_i 当成正态分布模型, 会得到与 GARCH 模型相同的似然函数. 由于似然函数只涉及 u_i^2 项 (等于观测值 δ_i^*), 可以使用 GARCH 软件来求解拟最大似然估计量 (QML), 该估计量是一致的, 且渐近正态 (参见 Lai 和 Xing (2008, 第 294–296 页)).

4.6.2 自激励点过程模型

在 ACD 模型中, 隐含条件是假设 (4.48) 式中的 g 是外生变量. 一个去除该假设的替代方法是直接利用*自激励点过程* (SEPP) 来对于 $N(t)$ 的强度函数进行建模 (而不是对于交易之间的时间间隔进行建模). Hautsch (2012) 以及 Bauwens 和 Hautsch (2009) 对于随机强度函数进行了回顾. 定义 $N(t)$ 的密度函数为

$$\lambda(t) = \lim_{s\downarrow 0} \frac{1}{s} E\left[N(t+s) - N(t)|\mathcal{F}_t\right], \tag{4.50}$$

其中 $\mathcal{F}_t$ 是直到 t 时刻的信息集 (σ 域), $N(t) = \max\{i : t_i \leqslant t\}$. 在附录 C 中指出, 强度过程包含了 $N(t)$ 的关键信息. 具体而言, $N(t) - \Lambda(t)$ 是一个零均值、局部二次可积的鞅过程, 其中 $\Lambda(t) = \int_0^t \lambda(s)ds$ 为聚合强度方程. 强度函数 $\lambda(t)$ 取决于直到 t 的信息集 (但不包括 t 时刻), 且为如下形式时:

$$\lambda(t) = \phi\left(\mu(t) + \int_{u<t} h(t-u)dN(u)\right), \tag{4.51}$$

则 $N(t)$ 为*自激励*的. 其中 $\phi : \mathbb{R} \to \mathbb{R}_+$, 且核函数 h 是 $\mathbb{R}_+$ 上的一个非负函数, 利用时刻 t 之前的事件发生时刻 $t_i(< t)$ 来对于时刻 t 时的自激励强度函数进行建模.

(4.51) 的一个特例是 $\phi(x) = x$, $\mu(t) \equiv \mu$. Hawkes (1971) 讨论了核函数为 $h(x) = \alpha e^{-\beta x}$ 的情形, 可得到 Hawkes 过程 $N(t)$ 的一元情况下的强度过程为

$$\lambda_\theta(t) = \mu + \int_{u<t} \alpha e^{-\beta(t-\mu)} dN(u) = \mu + \sum_{i:t_i<t} h_\theta(t-t_i), \tag{4.52}$$

其中 $\theta = (\mu, \alpha, \beta)$ 为参数向量, 可基于连续时间序列 $t_i \in [0, T]$, 利用最大似然估计方法求得. 当 (4.50) 式中的强度过程的参数 μ 满足 $\mu(t) \equiv \mu$ 和 $h(s) = h_\theta(s)$ 时, 利用 `R` 语言包 `ppstat` 和其最近的改进版 `NHPoisson`, 可以对于 Hawkes 过程和更一般的 SEPP 过程进行最大似然拟合, 并进行相关的诊

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

断检测. 附录 C 对于似然函数和相应的 MLE 统计量进行了详细说明, 同时对于其他包含模拟生成 SEPP 过程的函数的程序包进行了说明. 附录 C.4 对于模拟方法和软件进行了总结.

由于 Hawkes 模型的关系, 前面的讨论集中在对于 $\mu_\theta(t) \equiv \mu$ 的情形进行分析. 在 Chen 和 Hall (2013) 中, 分析了时变的强度函数 $\mu_\theta(t)$ 在分析日间交易间隔时的实用性, 并使用澳大利亚证券交易所的数据证明了其观点. 澳大利亚证交所的正常交易时间为澳大利亚东部标准时间 10:00–16:00. 论文中利用 B-样条 (参见 Lai 和 Xing (2008, 第 169 页)) 方法进行建模, 选取的内部结点为 11:00, 12:00, 13:00, 14:00 和 15:00, 在指数衰减函数情况下选取的核函数为 $h_\theta(t) = \gamma_1 e^{-\gamma_2 t}$, 在多项式衰减函数情况下选取的核函数为 $h_\theta(t) = \gamma_1(1+t)^{-(1+\gamma_2)}$. 论文表明, 在两种核函数下, SEPPVB(其中 VB 指时变背景下) 模型对于数据的拟合效果都比 ACD 模型更好, 且指数衰减函数下的 SEPPVB 模型的表现略优于多项式衰减函数下的模型. `R` 语言包 `ppstat` 和 `NHPoisson` 也可以用于时变的 $\mu_\theta(t)$ 情形.

4.6.3 D_i 的分解和广义线性模型

回顾之前内容, 我们可知交易价格的变动 $D_i = P_{t_i} - P_{t_{i-1}}$ 是离散的, 且 $D_i = 0$ 的概率很大. Rydberg 和 Shephard (2003) 在此基础上进行了如下的分解:

$$D_i = J_i A_i M_i, \tag{4.53}$$

以对 D_i 进行建模, 其中 $J_i = I_{\{D_i \neq 0\}}$. 在 $J_i = 1$ 时, A_i 是 D_i 的符号, M_i 是 D_i 的量级, 单位为最小交易单元. 论文中将给定 $\mathcal{F}_{i-1}$ 且 $y > 0$ 时 D_i 的密度函数表示为

$$\begin{aligned} &P(D_i = y|\mathcal{F}_{i-1}) \\ &= P(J_i = 1|\mathcal{F}_{i-1}) \times P(A_i = 1|J_i = 1, \mathcal{F}_{i-1}) \times P(M_i = y|A_i = 1, J_i = 1, \mathcal{F}_{i-1}). \end{aligned} \tag{4.54}$$

在 $y < 0$ 时, 可将上式中 $A_i = 1$ 替代为 $A_i = -1$. J_i 和 A_i 的二元性表明, 可以使用 Logistic 回归模型来估计 (4.54) 式中的 $P(J_i = 1|\mathcal{F}_{i-1})$ 和 $P(A_i = 1|J_i = 1, \mathcal{F}_{i-1})$. 令

$$\operatorname{logit}[P(J_i = 1|\mathcal{F}_{i-1})] = \boldsymbol{\beta}^\top \boldsymbol{x}_i, \tag{4.55}$$

$$\operatorname{logit}[P(A_i = 1|J_i = 1, \mathcal{F}_{i-1})] = \boldsymbol{\gamma}^\top \boldsymbol{z}_i, \tag{4.56}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 $\text{logit}(P) = \log \dfrac{P}{1-P}$, $\boldsymbol{x}_i$, $\boldsymbol{z}_i$ 和 $\boldsymbol{w}_i$ 是包含 $\mathcal{F}_{i-1}$ 的有限维向量, $\boldsymbol{\beta}$ 和 $\boldsymbol{\gamma}$ 是参数向量. 具体可参见 Lai 和 Xing (2008) 的 4.1.2 节, 该节也对于广义线性模型进行了介绍. $(M_i - 1)|(A_i, J_i = 1, \mathcal{F}_{i-1})$ 的一个简单的概率模型是几何分布模型, 其密度函数为 $\lambda_i(1-\lambda_i)^k$, $k = 0, 1, \ldots$, 参数 λ_i 与 $\boldsymbol{w}_i$ 的关系如下:

$$\text{logit}(\lambda_i) = \begin{cases} \boldsymbol{\theta}^\top \boldsymbol{w}_i, & \text{当 } A_i = 1, \\ \boldsymbol{\alpha}^\top \boldsymbol{w}_i, & \text{当 } A_i = -1. \end{cases} \tag{4.57}$$

因此, Rydberg-Shephard 模型的对数似然函数的形式为 $\sum_{i=1}^{n} f(D_i|\mathcal{F}_{i-1})$, 其中参数 $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\theta}$ 和 $\boldsymbol{\alpha}$, 函数 f 由 (4.55)–(4.57)式定义. 未知参数可以由最大似然估计法进行估计. $M_i - 1$ 的几何分布模型也可以由负二项分布来替代.

Rydberg 和 Shephard (2003) 利用 $J_{i-1}, A_{i-1}, M_{i-1}, \ldots, J_{i-q}, A_{i-q}, M_{i-q}$ 和 AIC[①], 对于 (4.55)–(4.57) 式中的参数 $\boldsymbol{x}_i$, $\boldsymbol{z}_i$ 和 $\boldsymbol{w}_i$ 进行估计. Tsay (2010, 250–253页) 在分析时加入了 $J_{i-1}, A_{i-1}, M_{i-1}$, 交易间的时间间隔 δ_{i-1}, t_{i-1} 时的交易量 (除以 100 的值), 以及 t_i 前的买卖价差. 可以发现, (4.54)–(4.57)式其实就是一个包含协变量的广义线性模型.

4.6.4 McCulloch 和 Tsay 分解方法

Rydberg 和 Shephard (2003) 中给出了 D_i 的一个分解方法, 而在 Tsay (2010) 中建立了一个 J_i 的拟合模型, 模型表明在非活跃交易 (即不引起价格变动的交易) 之后, 很可能会跟随着非活跃交易. 为了更有效地体现非活跃交易的集聚效应, McCulloch 和 Tsay (2001) 提出了如下的分解方法:

$$P_t - P_0 = \sum_{i=1}^{\tilde{N}(t)} \tilde{D}_i, \tag{4.58}$$

其中 $\tilde{N}(t) = \sum_{i=1}^{N(t)} J_i$ 是截至 t 时间共计发生的活跃交易数, $\tilde{D}_i$ 是第 i 次活跃交易引起的价格变动. 定义 $\tilde{t}_i$ 为第 i 次活跃交易的交易时间, 则 $\tilde{\delta}_i = \tilde{t}_i - \tilde{t}_{i-1}$ 为两次连续的活跃交易之间的时间间隔. 定义 $\tilde{Z}_i$ 为 $(\tilde{t}_{i-1}, \tilde{t}_i)$ 时间内发生的非活跃交易的次数, 则有 $\tilde{Z}_i = \sum_{j=N(\tilde{t}_{i-1})+1}^{N(\tilde{t}_i)-1} (1 - J_j)$. 因此, 高频交易数据 (δ_j, D_j), $j = 1, \ldots, N(T)$ 可以用 (4.58) 表示, 类似地可定义 $(\tilde{\delta}_i, \tilde{Z}_i, \tilde{D}_i)$, $j = 1, \ldots, \tilde{N}(T)$.

考虑到 $P(\tilde{Z}_i = 0) > 0$ 的值可能非常大, McCulloch 和 Tsay (2001) 令 $\tilde{Z}_i = \tilde{J}_i \tilde{U}_i$, 其中 $\tilde{J}_i = I_{\{\tilde{Z}_i > 0\}}$, $\tilde{U}_i = |\tilde{Z}_i|$. 另外, 由于 $P(\tilde{D}_i = 0) = 0$, $\tilde{D}_i$ 可以直

① AIC 是 Akaike 信息准则的缩写; 参见 Lai 和 Xing (2008, 122 页).

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

接由符号变量 $\tilde{A}_i = \text{sign}(\tilde{D}_i)$ 和量级变量 $\tilde{M}_i = |\tilde{D}_i|$ 定义. 类似于 4.6.3 节, $\tilde{J}_i$ 和 $\tilde{A}_i$ 可由 Logistic 回归得到, 而 $(\tilde{M}_i - 1)|(\tilde{A}_i, \mathcal{F}_{i-1})$ 和 $(\tilde{U}_i - 1)|(\tilde{J}_i = 1, \mathcal{F}_{i-1})$ 可由包含几何分布、负二项分布或泊松分布的广义线性模型所得. 最后, 与 4.6.3 节中将时间间隔与价格变化分别建模不同, McCulloch 和 Tsay (2001) 使用如下的快速衰减时间模型来对于 $\tilde{\delta}_i|\mathcal{G}_{i-1}$ 进行建模:

$$\log(\tilde{\delta}_i) = \beta_0 + \beta_1 \log(\tilde{\delta}_{i-1}) + \beta_2 \tilde{M}_{i-1} + \varepsilon_i, \tag{4.59}$$

其中 ε_i 独立同分布, 服从 $N(0, \sigma^2)$ 分布. 论文中将用于对 $\tilde{J}_i$, $\tilde{A}_i$, $\tilde{U}_i$ 和 $\tilde{M}$ 进行建模的 (4.59) 式、(4.58) 式和广义线性模型称为*价格和条件时间间隔模型* (PCD), 原因在于其提供了一种将时间间隔变化和价格变动进行联合建模的新思路. 注意到, 在该模型下, $\text{Cov}(\tilde{A}_i, \tilde{A}_{i-1}) < 0$ 即为*连续活跃*交易的买卖反弹.

McCulloch 和 Tsay (2001) 中讨论的另一个重要问题是如何将 "无间隔" 的时间序列模型用于跨交易日的高频交易建模. 注意到 Rydberg 和 Shephard (2003) 中只使用了开盘后 15 分钟的数据, 并将其与前一天的交易数据串联, 从而消除交易日间隔对于恒定参数模型的影响. 在 Engle 和 Russell (1998) 的 ACD 建模中也使用了类似的方法. ACD 方法基于不同交易日交易模式 g 恒定的假设, 利用 (4.48) 来估计白天的交易模式 g. 与假设不同交易日参数恒定不同, McCulloch 和 Tsay (2001) 定义第 j 个交易日的广义线性模型的参数向量为 $\boldsymbol{\Theta}_j$, 并假设其独立同分布, $\boldsymbol{\Theta}_j \sim N(\boldsymbol{\Theta}^*, \boldsymbol{\Sigma}^*)$, 其中 $\boldsymbol{\Theta}^*$ 和 $\boldsymbol{\Sigma}^*$ 分别是先验分布的均值向量和协方差矩阵. 先验分布可由马尔可夫链蒙特卡罗 (MCMC) 方法求得. 实际上, 论文中的分层模型与广义线性模型系统类似, 因此可以通过最大似然估计方法求得, 在求解时似然函数可由 Laplace 渐近二次方程和重要性抽样方法求得, 具体参见 Lai 等 (2006).

4.6.5 点过程和其标值的联合建模

Engle (2000) 中通过分析点过程 $N(t)$ 和其标值 (如价格变化 D_i) 在给定 $\mathcal{F}_{i-1}$ 时的联合条件分布, 建立了联合模型. 其中 $\delta_i = t_i - t_{i-1}$, y_i 是第 i 次交易的标值 (即表达利率的变量). 由于 $N(t)$ 本质上是一个到达时间间隔为 δ_i 的更新过程, 其也可以用于本节标题中提到的联合模型. 在 Engle (2000) 中, 讨论了如何将不同的市场微观结构特征用于对 y_i 进行合理建模. 举例来说, y_i 可以表示第 i 次交易的 P_{t_i}, D_i 或交易量, 而时间 t_i 的交易资产或其他相关资产的买入、卖出价格可以用标值向量 $\boldsymbol{y}_i$ 来求得. 论文中, 同时分析了利用分类变

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

量, 如交易方的特点或 (在混合交易市场中的) 下单机制等, 来对于 y_i 进行建模. 文中进一步指出 "经济学的假设往往与标值的分布有关", 特别是对于未来固定时间或在发生一定数量的交易之后的标值的分布, 以及 "通过对于交易的到达率进行直接量度, 并分析其对于市场中其他观测变量的分布的影响, 以对于交易时间和价格过程进行联合分析".

在 Engle (2000) 的第六节中, 引入了 UHF-GARCH 模型, 介绍了上述联合建模方法的实际应用, 其中 UHF 指 "超高频". 定义 r_i 为 t_{i-1} 到 t_i 区间内的收益, 则每次交易的条件波动率为

$$V_{i-1}(r_i|x_i) = h_i. \tag{4.60}$$

第 i 次交易的条件波动率 σ_i 通常是以年化的形式给出, 定义为 $V_{i-1}\left(\frac{r_i}{\sqrt{\delta_i}}|\delta_i\right) = \sigma_i^2$. 根据 (4.60) 和上式, 可知 $h_i = \delta_i\sigma_i^2$. 利用 ARMA(1,1)-GARCH(1,1) 模型 (参见 Lai 和 Xing (2008, 156 页)) 对 $\frac{r_i}{\sqrt{\delta_i}}$ 进行建模, 可得

$$\begin{aligned} \frac{r_i}{\sqrt{\delta_i}} &= \rho\frac{r_{i-1}}{\sqrt{\delta_{i-1}}} + e_i + \phi e_{i-1}, \\ \sigma_i^2 &= \omega + \alpha e_{i-1}^2 + \beta\sigma_{i-1}^2, \end{aligned} \tag{4.61}$$

其中新息项为 e_i. 该模型也是 Engle 讨论的最基本的 UHF-GARCH 模型. 包含对 δ_i 进行 EAD 建模的更复杂的 UHF-GARCH 模型可参见 Engle (2000).

4.6.6 结合低频波动率和高频波动率的已实现 GARCH 等预测模型

2.6.3 节提到了 `R` 语言包 `rugarch`, 其中第一个 `r` 代表 "已实现". 已实现 GARCH (RealGARCH) 模型最初由 Hansen 等 (2012) 提出, 用于将高频交易中波动率的测度与低频交易 (如日交易) 中的波动率相联系. 论文指出:

> 标准的 GARCH 模型通常使用日收益率 (通常使用收益率的平方) 来得到当前的波动率信息, 且该信息被用于对下一期的波动率进行预测, 而单个的收益率只能对于当时波动率提供一个微弱的信号. 由此可见, GARCH 模型通常并不适用于波动率剧烈变化的情形. 其原因在于, GARCH 模型通常并不能及时地更新信息 …… 目前, 高频交易金融数据是广泛可得的, 而现有文献介绍了一系列波动率的 "已实现测度" …… 相对于收益率的平方而言, 这些测度对下一期波动率的预测提供了更有价值的信息. 因此, 已实现测度在

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

对于未来波动率的建模和预测中非常有效. 对于包含已实现测度的 GARCH 模型 (也称为 GARCH-X 模型) 的预测对于上述观点进行了有效说明.

然而, RealGARCH 模型和 GARCH-X 模型有一定的区别, 前者是一个 "完全" 模型 (数据生成机制), 且刻画了已实现测度 x_t 的随机变化. 举例来说, RealGARCH(1,1) 的对数形式为 $r_t = \sqrt{h_t} z_t$, 其中 z_t 服从独立同分布的正态分布, 且 GARCH 模型和测量方程分别为:

$$\begin{aligned} \log h_t &= \omega + \beta \log h_{t-1} + \gamma \log x_{t-1}, \\ \log x_t &= \xi + \varphi \log h_t + \tau(z_t) + u_t, \end{aligned} \tag{4.62}$$

其中 u_t 独立同分布, 且有 $E(u_t) = 0$, $\mathrm{Var}(u_t) = \sigma_u^2$. $\tau(z)$ 是杠杆函数, 用于刻画 z_t 对波动率的非对称影响, 且有 $E[\tau(z)] = 0$. Hansen 等 (2012) 中令 $\tau(z) = \tau_1 z + \tau_2(z^2 - 1)$, 该函数可以在 GARCH 方程中引入 EGARCH 结构 (参见 Lai 和 Xing (2008, 152–153 页)). 高频交易风险测度 x_t 的不确定性, 作为一些高频交易实体 (如 $[X]_T$) 的近似或测度, 体现在 (4.62) 式的潜在变量 u_t 中. 此处假设 $\left(z_t, \dfrac{u_t}{\sigma_u}\right)^\top \sim N(\mathbf{0}, \boldsymbol{I})$, 则可完成对于 RealGARCH 的说明.

对于对数线性方程 (4.62) 的一个替代形式是利用如下的线性变换:

$$h_t = \omega + \beta h_{t-1} + \gamma x_{t-1}, \quad x_t = \xi + \varphi h_t + \tau(z_t) + u_t. \tag{4.63}$$

在此之前, Engle 和 Gallo (2006) 引入了一种类似的模型, 但其中 x_t 被替代为 $(h_{R,t}, h_{RV,t})$, R_t 表示交易日内的价格极差 (最高价减去最低价), RV_t 为已实现方差.

$$\begin{aligned} h_t &= \omega + \alpha r_{t-1}^2 + \beta h_{t-1} + \delta r_{t-1} + \varphi R_{t-1}^2, \\ h_{R,t} &= \omega_R + \alpha_R r_{t-1}^2 + \beta_R h_{t-1} + \delta_R r_{t-1} + \varphi R_{t-1}^2, \\ h_{RV,t} &= \omega_{RV} + \alpha_{RV} RV_{t-1} + \beta_{RV} h_{RV,t-1} + \delta_{RV} r_{t-1} \\ &\quad + \theta_{RV} RV_{t-1} I_{\{r_{t-1}<0\}} + \varphi_{RV} r_{t-1}^2, \end{aligned}$$

其中 $R_t^2 = h_{R,t} z_{R,t}^2$, $RV_t = h_{RV,t} z_{RV,t}^2$, 以及 $(z_t, z_{R,t}, z_{RV,t})^\top \sim N(\mathbf{0}, \boldsymbol{I})$. 该模型称为乘法误差模型 (MEM). 在此之后, Shephard 和 Sheppard (2010) 提出了一种相对于 (4.63) 更复杂的模型, 形式如下:

$$\begin{aligned} h_t &= \omega + \alpha r_{t-1}^2 + \beta h_{t-1} + \gamma x_{t-1}, \quad x_t = \mu_t z_{RK,t}^2, \\ \mu_t &= \omega_R + \alpha_R x_{t-1} + \beta_R \mu_{t-1}, \end{aligned}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 $(z_t, z_{RK,t})^\top \sim N(\mathbf{0}, \boldsymbol{I})$, x_t 为 $[X]_T$ 的已实现核估计量. Shephard 和 Sheppard 将论文中的高频波动率模型简称为 HEAVY 模型.

4.6.7 有效价格过程中的跳跃和幂变差

微观结构噪声通常被用于量化高频交易与标准布朗运动半鞅有效价格过程之间的偏差, 偏差的来源包括 4.2.2 节中提到的买入–卖出反弹和其他的微观结构因素. 而带跳的几何布朗运动过程则通常用于分析资产价格的大幅变化, 该模型的应用最初起源于 20 世纪 60 年代对于低频交易数据的分析, 参见 Press (1967)、Merton (1976) 以及 Ball 和 Torous (1983). 这类模型的最简单形式为

$$dX_t = \mu dt + \sigma dB_t + J_t dN_t, \tag{4.64}$$

其中 X_t 是时间 t 时的资产对数价格, B_t 是标准布朗运动, μ 和 $\sigma > 0$ 分别是漂移项和波动率参数. N_t 是一个到达率参数为 λ 的泊松过程, 而 $J_t \sim N(\beta, \eta)$ 是跳跃幅度的对数. 在该模型下, 收益率可以表示为泊松混合形式下的正态分布, 因此其厚尾性可以由 (4.64) 表示. 跳跃幅度 J_t 的非零均值 β 也可以用于对收益的偏度自由性进行建模.

受到资产定价模型和利率的期限结构模型的应用的驱动, (4.64) 式被扩展为如下形式:

$$dX_t = \mu dt + \sigma dB_t + dL_t, \tag{4.65}$$

其中 L_t 表示独立稳定增量的 Lévy 单纯跳跃过程, 参见 Madan 等 (1998)、Carr 和 Wu (2004)、Eberlein 等 (1998) 以及 Eberlein 和 Raible (1999). 在 (4.64) 式中, $dL_t = J_t dN_t$ 的情况对应于 L_t 为跳跃幅度服从正态分布的复合泊松过程. 更一般地, 当 L_t 是一个跳跃幅度独立同分布的复合泊松过程时, 称为*有限活跃*且跳跃相对稀疏. 所有其他的 Lévy 单纯跳跃过程称为*无限活跃*, 即跳跃事件通常在每个开区间内都会无限发生, 且大部分跳跃幅度很小. Aït-Sahalia (2004) 将金融经济学中与 (4.65) 相关的研究方向分为以下三大类:

(a) 对参数或置信 (预测) 区间的预测;

(b) 基于离散数据测试当前是否有跳跃过程发生;

(c) 在发生跳跃时, 对于二次变异和相关量的统计特征进行分析.

上述三个方向的研究对于 (4.65) 进行了丰富的应用. 首先, 跳跃扩散过程 (4.64) 的二次变差的形式为

$$[X]_T = \sigma^2 T + \sum_{i=1}^{N_T} J_{s_i}^2, \tag{4.66}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 s_i 为 N_t 的第 i 次到达时间. 即 $\sigma^2 T$ 为二次变差的连续部分, 而 $\sum_{i=1}^{N_T} J_{s_i}^2$ 为非连续部分, 参见附录 A. 与只分析二次变差不同, Lepingle (1976) 引入了 (绝对) 幂变差, 类似于 (4.6), 定义为

$$ {}_p[X]_T = \operatorname*{p-lim}_{\text{mesh}(\Pi)\to 0} \sum_{i=1}^{n} |X_{t(i)} - X_{t(i-1)}|^p. \tag{4.67} $$

论文中证明在模型 (4.65) 下, 在 $p \in (0,2)$ 上, ${}_p[X]_T$ 中跳跃部分的贡献为 0 (在标准化之后); 而对于 $p > 2$, 贡献为 ∞.

Aït-Sahalia 和 Jacod (2012) 定义 (4.67) 的离散时间形式为

$$ B(p,u,\delta) = \sum_{i=1}^{n} |X_{i\delta} - X_{(i-1)\delta}|^p I_{\{|X_{i\delta} - X_{(i-1)\delta}| \leqslant u\}}, $$

其中 u 为截断水平, $\delta = \dfrac{T}{n}$ 是通常使用的抽样区间. 论文在 $\delta \to 0$ 和 $u \to \infty$ 的条件下, 对于 $B(p,u,\delta)$ 的渐近行为进行分析, 并提出了针对 (a) 和 (b) 的三种统计策略:

- 增大幂 $p > 2$, 以放大跳跃幅度;
- 利用截断水平 u 来消除大幅跳跃;
- 在不同的抽样频率下, 比较幂变差.

特别地, 论文中给出了 (i) 在原假设为 "X_t 的跳跃过程为无限活跃" 时, 检验有限活跃的概率; (ii) 检验连续部分的存在性, 两个假设的检验统计项. Sahalia 和 Jacod (2011, 2012, 2015) 同样利用这些策略, 在存在舍入误差和微观结构噪声情况下, 建立了稳健估计量.

4.7 补充材料和评论

1. **微观结构噪声模型和 $n^{-1/4}$-收敛** (4.21) 中的似然函数 $l(v_\varepsilon, \sigma^2)$ 由如下正态微观噪声模型得到: 对于 $i = 1, \ldots, n$,

$$ Y_{i\delta} = X_{i\delta} + \varepsilon_i, \tag{4.68} $$

其中 Y_t 为观测到的交易价格, $\delta = T/n$ 为时间聚合区间长度, X_t 为模型 $dX_t = \sigma dB_t$ 下的有效价格过程, 以及 ε_i 独立同分布, 服从 $N(0, v_\varepsilon)$ 分布, 且与 X_t 独立. 上述模型和 MLE 的渐近性已在空间统计学中研究, 参见

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

Stein (1990, 1999) 和 Chen 等 (2000). Chen 等 (2000) 对于 Ornstein-Uhlenbeck (OU) 过程 $dX_t = -\alpha X_t dt + \sigma\sqrt{2\alpha}dB_t$ ($\alpha > 0$) 也进行了研究. 微观结构噪声项在空间统计学中称为 "测量误差" 或 "积累效应". 在 $dX_t = \sigma dB_t$ 和 $dX_t = -\alpha X_t dt + \sigma\sqrt{2\alpha}dB_t$ 两种情形下, Chen 等 (2000) 和 Xiu (2010) 表明, σ^2 的 QMLE 是一致的, 且以 $n^{1/4}$ 的速度收敛, 即对于固定的 T (不失一般性, 可令 $T = 1$), 在 $n \to \infty$ 时, $n^{1/4}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{\mathcal{L}} N(0, \psi)$. 该性质在空间统计学中称为内填式渐近, 且对于两种模型 ψ 不同. 注意到 OU 过程的平稳分布是方差为 σ^2 的高斯分布. 两种情形下, QMLE $\hat{v}_\varepsilon$ 具有相同的渐近分布 $\sqrt{n}(\hat{v}_\varepsilon - v_\varepsilon) \xrightarrow{\mathcal{L}} N(0, 2v_\varepsilon^2)$, 且都与 $\hat{\sigma}^2$ 渐近独立. 令 $\xi_i = (X_{i/n} - X_{(i-1)/n})/(\sigma/\sqrt{n})$, 由于 $\Delta_i = Y_{i/n} - Y_{(i-1)/n}$, 则在 $T = 1$ 和 $dX_t = \sigma dB_t$ 时, 模型 (4.68) 可以写成

$$\Delta_i = \sqrt{v_\varepsilon}(e_i - e_{i-1}) + \sqrt{\frac{\sigma^2}{n}}\xi_i, \tag{4.69}$$

其中 ξ_i 和 $e_i = \dfrac{\varepsilon_i}{\sqrt{v_\varepsilon}}$ 独立同分布, 服从 $N(0,1)$ 分布, 当 $n \to \infty$ 时, 移动平均项 $\sqrt{v_\varepsilon}(e_i - e_{i-1})$ 支配 Δ_i. 在此情况下, 用于估计 σ^2 的数据提供的信息量远远小于用于估计 v_ε 的数据, 因此, σ^2 的估计量的收敛速度为 $n^{-1/4}$, 并不等于 v_ε 的估计量的收敛速度 $n^{-1/2}$.

2. **QMLE 的计算** 由于在 (4.69) 式中 $(e_i - e_{i-1})$ 的占优特性, (4.21) 中的似然函数 $l(v_\varepsilon, \sigma^2)$ 对于 σ^2 相对平缓, 而在 v_ε 方向上达到峰值. 图 4.1 的上图基于参数为 $(\sigma^2, v_\varepsilon) = (1, 1)$ 的 $n = 1000$ 个模拟生成的样本, 给出了似然函数. 因此, 利用 Newton 迭代方法得到的 MLE 对于初始值非常敏感. 在图 4.1 的上图中, 尽管 $\hat{v}_\varepsilon$ 取值接近 1, 但 σ^2 的估计值的取值范围很大. 为了解决这个问题, 研究中可以使用 Hansen 等 (2008) 中建议的 MA(1) 参数化方法:

$$\Delta_i = \eta_i - \theta\eta_{i-1}, \tag{4.70}$$

其中 η_i 独立同分布, 服从 $N(0, \sigma_\eta^2)$ 分布. (4.70) 式的对数似然函数与 $l(v_\varepsilon, \sigma^2)$ 的形式相同, 但在 (4.22) 式中, 三对角矩阵 $\boldsymbol{\Omega}$ 的对角元素和非零非对角元素分别替换成 $\sigma_\eta^2(1 + \theta^2)$ 和 $-\theta\sigma_\eta^2$, 其中

$$\sigma_\eta^2(1 + \theta^2) = 2v_\varepsilon + \frac{\sigma^2}{n}, \quad \theta\sigma_\eta^2 = v_\varepsilon. \tag{4.71}$$

Hansen 等 (2008) 基于上式中的 σ^2 的求解结果, 提出使用 $n\hat{\sigma}_\eta^2(1 + \hat{\theta}^2)$ 作为 σ^2 的估计量, 其中 $(\hat{\sigma}_\eta^2, \hat{\theta})$ 为 MA(1) 模型 (4.70) 的 MLE. 根

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

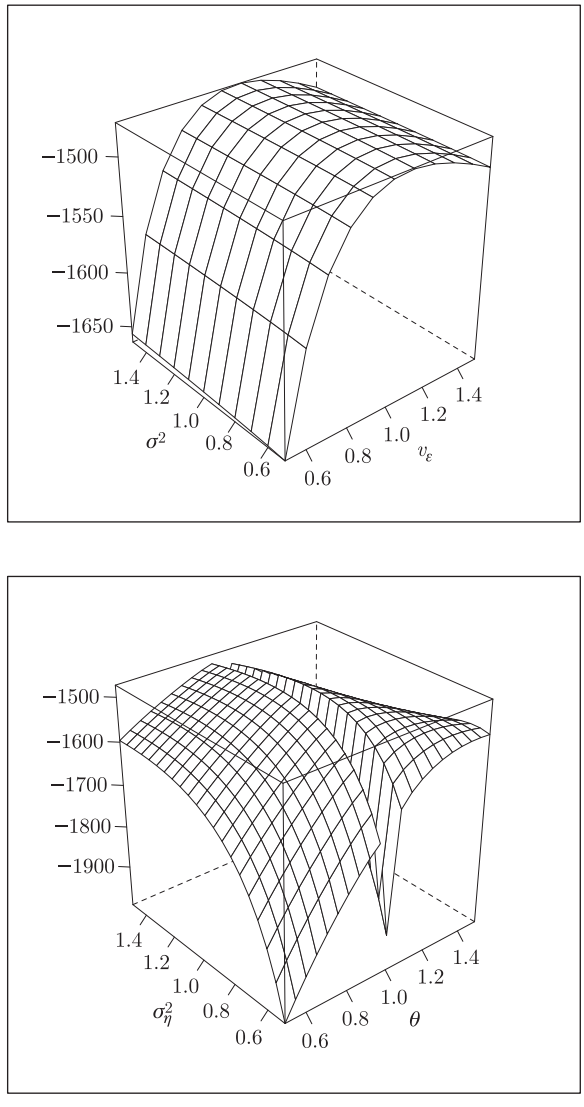


图 4.1　上图为根据模型 (4.21) 生成得到 $n = 1000$ 个样本的似然函数, 等价于模型 (4.69) 的生成样本似然函数图; 下图为根据 MA(1) 模型 (4.70) 的生成样本似然函数图

据 (4.71) 可知

$$\theta = 1 + \frac{\sigma^2}{nv_\varepsilon} - \sqrt{\frac{\sigma^2}{4nv_\varepsilon^2}\left(1 + \frac{\sigma^2}{nv_\varepsilon}\right)},$$

$$\sigma_\eta^2 = \frac{v_\varepsilon}{1 + \frac{\sigma^2}{nv_\varepsilon} - \sqrt{\frac{\sigma^2}{4nv_\varepsilon^2}\left(1 + \frac{\sigma^2}{nv_\varepsilon}\right)}}.$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

因此, θ 的值接近 1, 因此 MA(1) 模型 (4.70) 接近于不可逆, 而 σ_η^2 的值则接近于 v_ε. 在图 4.1 的下图中, 给出了参数为 (σ_η^2, θ) 时的对数似然函数方程. 图中给出了 $\theta = 1$ 附近的最大似然函数特征, 由图可以看出 MLE 估计量 $(\hat{\sigma}_\eta^2, \hat{\theta})$ 难以求解, 具体也可参见 Brockwell 和 Davis (2002, 196 页).

3. **模型 (4.68) 的反卷积和最小最大估计量** Cai 等 (2010) 对于前面提到的 QMLE 计算进行了研究, 并提出了 $(v_\varepsilon, \sigma^2)$ 的最小最大估计量, 该估计量的计算中并不需要进行复杂的迭代过程, 且其与 Xiu 提出的 QMLE 估计量具有相同的边际渐近分布①. 论文在建立估计量时, 对于模型 (4.68) 进行反卷积, 将 $X_{i\delta} = \sigma B_{i\delta}$ 作为信号, 而观测值 $Y_{i\delta}$ 作为信号和微观结构噪声的卷积.

奇异值分解 (SVD) Cai 等 (2010) 利用 $\mathrm{Cov}(B_t, B_s) = s \wedge t$ 来求解 (4.68) 的逆问题, 即有 $\boldsymbol{Y} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{C} + v_\varepsilon \boldsymbol{I})$, 其中 $\boldsymbol{C}$ 的第 (i, j) 个元素为 $\left(\frac{i}{n}\right) \wedge \left(\frac{j}{n}\right)$. $\boldsymbol{C}$ 的奇异值分解 (SVD) 为 $\boldsymbol{C} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$, 其中 $\boldsymbol{U} = (u_{i,j})_{1 \leqslant i,j \leqslant n}$ 是一个正交矩阵, 定义如下:

$$u_{ij} = \sqrt{\frac{4}{2n+1}} \sin\left(\frac{(2j-1)i\pi}{2n+1}\right), \tag{4.72}$$

而 $\boldsymbol{\Lambda}$ 是一个对角矩阵, 第 i 个对角元素为

$$\lambda_i = \left[4n \sin^2\left(\frac{(2i-1)\pi}{4n+2}\right)\right]^{-1}. \tag{4.73}$$

因此, 通过分解式 $\boldsymbol{Z} = \boldsymbol{U}^\top \boldsymbol{Y} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{\Lambda} + v_\varepsilon \boldsymbol{I})$, 可得到 n 个独立变量:

$$Z_i \sim N(0, \sigma^2 \lambda_i + v_\varepsilon), \quad i = 1, \ldots, n, \tag{4.74}$$

从而得到 σ^2 的线性估计量如下:

$$\hat{\sigma}^2 = \frac{1}{b_n} \sum_{i=k+1}^{m} \frac{\lambda_i}{(\tilde{\sigma}^2 \lambda_i + \tilde{v}_\varepsilon)^2} (Z_i^2 - \tilde{v}_\varepsilon), \quad \text{对于 } 1 < k < m < n, \tag{4.75}$$

其中 $\tilde{\sigma}^2$ 和 $\tilde{v}_\varepsilon$ 分别是 σ^2 和 v_ε 的初始估计量, 定义如下:

$$\tilde{v}_\varepsilon = \frac{1}{n-m} \sum_{i=m+1}^{n} Z_i^2, \quad \tilde{\sigma}^2 = \frac{1}{k} \sum_{i=1}^{k} \frac{Z_i^2 - \tilde{v}_\varepsilon}{\lambda_i}. \tag{4.76}$$

① 尽管 v_ε 和 σ^2 的 QMLE 是渐近不相关的, 但在 Cai 等 (2010) 中并没有分析文中的估计量是否符合该性质.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

Cai 等 (2010) 指出, 当 $n \to \infty$, $\dfrac{m}{n} \to 0$, 而 $\dfrac{m}{\sqrt{n}} \to \infty$ 时, 满足 $\sqrt{n}(\tilde{v}_\varepsilon - v_\varepsilon) \xrightarrow{\mathcal{L}} N(0, 2v_\varepsilon^2)$, 且 $\tilde{v}_\varepsilon$ 是渐近最小最大值. 论文中同时指出, 当取值 $k = [n^{1/2-\gamma}]$ 和 $m = [n^{1/2+\gamma}]$, $0 < \gamma < \dfrac{1}{20}$, b_n 取值如下时:

$$b_n = \sum_{i=k+1}^{m} \frac{\lambda_i^2}{(\tilde{\sigma}^2\lambda_i + \tilde{v}_\varepsilon)^2}, \tag{4.77}$$

$n^{1/4}(\tilde{\sigma}^2 - \sigma^2) \xrightarrow{\mathcal{L}} N(0, 8\sqrt{v_\varepsilon}\sigma^3)$, 且 $\hat{\sigma}^2$ 渐近最小最大. 在进行 $\boldsymbol{Z} = \boldsymbol{U}^\top \boldsymbol{Y}$ 变形时, 需要的步骤数仅为 $O(n \log n)$, 而 (4.76)、(4.77) 和 (4.75) 式的计算只需要 $O(n)$ 个步骤, 且并不需要进行迭代计算.

上述方法将 $\boldsymbol{Y}$ 中的相关项进行变形, 进行 $\boldsymbol{Z} = \boldsymbol{U}^\top \boldsymbol{Y}$ 变换, 从而得到不包含相关项的 $\boldsymbol{Z}$. Munk 等 (2010) 利用该方法, 基于 σ_t^2 和 $v_{\varepsilon,t}$ 都是确定平滑函数而不是恒定参数的假设, 对于 $\int_0^T \sigma_t^2 dt$ 和 $\int_0^T v_{\varepsilon,t} dt$ 进行估计. 注意到, 虽然 Xiu (2010) 也使用了同样的符号标记 $dX_t = \sigma_t dB_t$, 但在该论文中 σ_t^2 是一个随机过程 (局部有界且取值为正的半鞅过程). 确定性和平滑的 σ_t^2 的假设来自反卷积相关文献, Hoffmann 等 (2012) 基于该假设, 通过小波分析对于瞬时波动率 σ_t^2 进行建模. Sabel 等 (2015) 在此基础上进行了拓展, 讨论了更一般的情形, 如存在跳跃和舍入误差, 或微观结构误差等.

4. **QML 准则和渐近理论** QML 准则在与数据和参数相关的 "参数模型" 下, 通过最大化似然函数来估计参数向量 $\boldsymbol{\theta}$. 前面章节在对于 QMLE 进行讨论时, 对于 QML 已经进行了一定的介绍, 并指出实际的数据生成机制比假定的模型要复杂得多. 空间统计学就是一个典型的例子, 而更一般的图像重建过程中的时空马尔可夫随机场则更难以处理, 参见 Lai 和 Lim (2015), 该论文中使用 QML 准则, 将随机场进行分块, 假设不同分块相互独立. 在更早的 Besag (1974, 1986) 研究中, 使用了一个复合似然函数, 该函数为多个条件密度方程的乘积, 某一位置的图像的密度方程的值为某一近邻系统下, 在给定近邻图像值时的似然函数值. 尽管在建模时存在过于简单化或潜在的错误设定, 当 QMLE 中能捕获参数向量和数据生成机制的基本关系时, 统计量仍然是一致且渐近正态的, 具体参见 White (1982)、Lee 和 Hansen (1994) 以及 Cox 和 Reid (2004).

5. **做市和负 MA(1) 自相关关系** 在 Roll 模型 (4.1) 建立以前, Mandelbrot (1963) 对于证券价格变化和做市商之间的关系进行了讨论, 论文

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

指出专家可以在一定程度上消除稳定分布模型预测的跳跃. Mandelbrot 和 Taylor (1967) 则建议使用从属随机过程来对于专家的行为进行量化. Niederhoffer 和 Osborne (1966) 将逐笔交易中价格的变动 (负 MA(1) 特性) 归结于专家的薪酬. Roll (1984) 利用 (4.1) 式, 通过模拟买卖反弹对于 MA(1) 特性进行了解释.

6. **负 MA(1) 自相关关系的非交易解释和异步交易** Miller 等 (1994) 提到, 股票的交易并不是连续的, 其可能会在连续的区间中交易, 但不会在每个区间的 "收盘时间" 进行交易, 这个现象称为 "异步交易". 股票也可能在一个区间中并不进行交易, 称为 "非交易". "非交易" 可以看作模型 (4.4) 的特殊情形, 原因在于 t_i 至 t_{i+1} 区间内未进行交易, 且 t_i 并不需要均匀取值. Lo 和 MacKinlay (1990) 引入了一个非交易的随机模型, 并利用该模型对于负 MA(1) 自相关关系进行了解释, 具体细节可参见 Lai 和 Xing (2008, 291 – 292 页).

7. **全天波动率的 RealGARCH 建模** 所有的已实现测度都在交易时间内对于波动率进行量度. RealGARCH 方法与 4.6.6 节中提到的 MEM 和 HEAVY 方法概念上并不相同, 原因在于 RealGARCH 方法利用 (给定 h_t 时的) 同时发生的全天波动率, 对于已实现测度 (记为 x_t) 进行建模, 所使用模型为 (4.63) 中 $x_t = \xi + \varphi h_t + \tau(z_t) + u_t$ (或 (4.62) 中的对数线性模型). RealGARCH 方法得到的全天波动率模型可由以下的示意图表示:

$$h_t \left\{ \begin{array}{l} \text{市场在第 } t-1 \text{ 天收盘} \\ \quad\downarrow \\ \text{市场在第 } t \text{ 天开盘} \\ \quad\downarrow\downarrow \\ \text{市场在第 } t \text{ 天收盘} \end{array} \right\} v_t$$

且有 $x_t = v_t +$ 噪声项, 其中 v_t 为第 t 天交易时间内高频交易收益率的条件方差. 因此, v_t 的模型为 $\xi + \varphi h_t$. 另外, 当 r_t 为收盘至收盘的收益率时, φ 的值应该小于 1. Hansen 等 (2012) 第 5.3 节的实证结果说明: (a) 当 r_t 为开盘至收盘收益率, x_t 为交易所交易基金 SPY 的已实现核波动率估计量时, $\hat{\varphi} \approx 1$. (b) 当 r_t 为收盘至收盘收益率时, $\hat{\varphi} = 0.74 < 1$.

8. **Epps 效应产生的原因** Epps 效应的产生主要归因于高频交易数据的两大特征. 第一个特征是 4.6.4 节中提到的非活跃交易在高抽样频率下的粘

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

性. 在 4.4.2 节中的同步区间的宽度下降时, 两个资产的同步价格变动为 0, 或者非常小 (比如说流动资产的变动只有一个最小交易单位). 另一个特征是异步性, 异步性导致其中一个资产的同步价格大幅度变化时, 另一个资产的同步价格变动却很小或为 0. Zhang (2011) 利用 "前笔交易" 方法对于异步性的影响进行了讨论, 但是由于前面提到的原因, 当同步区间很小时, 用于协变异估计的有用信息非常少. Malliavin 和 Mancino (2002, 2009) 中提到的傅里叶方法在解释异步性影响上更为有效, 原因在于傅里叶方法提供了一个数据驱动窗口的宽域, 从而可以在不忽略数据的情况下对于协变异进行分析.

9. **错位偏误和广义同步方法** 尽管 4.4.2 节中的广义同步方法中的任意选择方法看起来非常不符合直觉, 但是 Aït-Sahalia 等 (2010) 研究指出, 相比于 "前笔交易" 方法, 随机选取 $\tilde{t}_{i,j} \in (\tau_{j-1}, \tau_j]$ 可以消除错位偏误的影响. 论文提到 "在两种情形下, 在需要时, 可以选取资产在前一笔交易的价格作为在抽样时间 τ_j 时的价格. 相反, 此处我们建议对于各个资产, 使用区间内任意时间的价格. 在实际中, 连续交易的订单的记录可能会有错误, 但由于我们建议使用的方法对于交易的选取并没有任何要求, 因此只要错位的数据点在同一个抽样区间内, 求得的估计量对于错位偏误是稳健的." 信息学中各交易所交易数据汇总的问题和潜在的错位偏误将在 8.3.1 节中进行讨论.

10. **协方差的对称正定傅里叶估计量** Liu 和 Ngo (2014) 指出, 已实现方差的傅里叶统计量并没有涉及同步性的问题, 因此 Epps 效应对于傅里叶统计量不适用. 瞬时协方差的傅里叶估计量 (4.45)可能会是一个特征值为负的矩阵. Malliavin 和 Mancino (2009) 提出了瞬时协方差的另一个估计量:

$$\tilde{\Sigma}_t^{(j,k)} = \sum_{l=-N_S}^{N_S} \left(1 - \frac{|l|}{N_S}\right) e^{ilt} \left[\frac{2\pi}{2N_X+1} \sum_{n=-N_X}^{N_X} \hat{c}_n(dp_j)\hat{c}_{l-n}(dp_k)\right], \tag{4.78}$$

其中 $\hat{c}_n(dp_j) = \hat{a}_n(dp_j) - \sqrt{-1} \times \hat{b}_n(dp_j)$, $\hat{a}_n(dp_j)$ 和 $\hat{b}_n(dp_j)$ 可以通过将 (4.33)–(4.35)拓展到 $-n$ 来定义. Akahori 等 (2014) 对于估计量 (4.78) 进行了研究, 将其表达为

$$\tilde{\Sigma}_t^{(j,k)} = \sum_{r=2}^{n_j} \sum_{s=2}^{n_k} \mathcal{G}_{N_S}(t - t_{j,r}) D_{N_X}(t_{j,r} - t_{k,s}) R_{j,r} R_{k,s}, \tag{4.79}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 $R_{j,r}=X_{j,t_{j,r}}-X_{i,t_{j,r-1}}$, $R_{k,s}=X_{k,t_{k,s}}-X_{k,t_{k,s-1}}$, 而 $\{t_{j,r}:r=1,\ldots,n_j\}$ 和 $\{t_{k,s}:s=1,\ldots,n_k\}$ 分别是对数价格 $X_{i,t}$ 和 $X_{j,t}$ 的交易时间. 类似于在 (4.42) 中, 在 (4.79) 中 $D_N(x)$ 为调整后的 Dirichlet 核函数, $\mathcal{G}_M(t)$ 为 Fejér 核函数, 形式如下:

$$\mathcal{G}_M(t)=\sum_{|l|\leqslant M-1}\left(1-\frac{|l|}{M}\right)e^{ilt}=\frac{1}{M}\left(\frac{\sin\dfrac{Mt}{2}}{\sin\dfrac{t}{2}}\right)^2.$$

Akahori 等 (2014) 指出, 由于 $\mathcal{G}_{N_S}(t-t_{j,r})D_{N_X}(t_{j,r}-t_{k,s})$ 对于 j 和 k 不对称, $\tilde{\Sigma}_t^{(j,k)}\neq\tilde{\Sigma}_t^{(k,j)}$. 为了解决不对称性问题, 保证估计量的正定性, 论文中定义了一个更广义的统计量如下:

$$\breve{\Sigma}_t^{(j,k)}=\sum_{l\in\mathbb{L}}f(l)e^{ilt}\left[\frac{2\pi}{|\mathbb{S}(l)|}\sum_{(m,n)\in\mathbb{S}(l)}\hat{c}_m(dp_j)\hat{c}_n(dp_k)\right],\tag{4.80}$$

其中 $|\mathbb{S}(l)|$ 是集合 $\mathbb{S}(l)$ 的势. 论文中同时证明, 当 $\mathbb{L}=\{-N_S,\ldots,-1,0,1,\ldots,N_S\}$, $\mathbb{S}(l)=\{(m,l-m):m=-N_X,\ldots,-1,0,1,\ldots,N_X\}$, 且对于任意 $l\in\mathbb{L}$, $f(l)=\left(1-\dfrac{|l|}{N_S}\right)$ 时, $\breve{\Sigma}_t^{(j,k)}$ 简化为 $\tilde{\Sigma}_t^{(j,k)}$. Akahori 等 (2014) 令 $\mathbb{L}=\{-2M,-2M+1,\ldots,-1,0,1,\ldots,2M-1,2M\}$, 且满足对于任意 $l\in\mathbb{L}$,

$$\mathbb{S}(l)=\begin{cases}\{(-M+l+v,M-v):v=0,1,\ldots,2M-l\}, & \text{当 } l\geqslant 0,\\ \{(M+l-v,-M+v):v=0,1,\ldots,2M+l\}, & \text{当 } l<0,\end{cases}$$

并令 f 为 $\mathbb{L}$ 的正定函数, 证明此时 $\breve{\Sigma}_t^{(j,k)}$ 是正定的. 论文中同时建议对于 f 进行如下取值: (a) $f(l)=\left(1-\dfrac{|l|}{2M}\right)\exp\{-\gamma|l|\}$, (b) $f(l)=\left(1-\dfrac{|l|}{2M}\right)\exp\{-2\pi^2\dfrac{l^2}{C}\}$, 以及 (c) $f(l)=\left(1-\dfrac{|l|}{2M}\right)^2$, 且 $2M+1$ 为质数.

11. **半正定多元已实现核估计量** 将单变量已实现核估计量进行多元推广时, 所得的如下估计量并不是正定的:

$$\tilde{\boldsymbol{K}}(\boldsymbol{Y}_\tau)=\tilde{\boldsymbol{\Gamma}}_0+\sum_{h=1}^{H}\kappa\left(\frac{h-1}{H}\right)(\tilde{\boldsymbol{\Gamma}}_h+\tilde{\boldsymbol{\Gamma}}_{-h}),$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中

$$\tilde{\boldsymbol{\Gamma}}_h = \sum_j \left(\boldsymbol{Y}_{\tau_j} - \boldsymbol{Y}_{\tau_{j-1}}\right)\left(\boldsymbol{Y}_{\tau_{j-h}} - \boldsymbol{Y}_{\tau_{j-h-1}}\right)^\top;$$

参见 Barndorff-Nielsen 等 (2011, 6.1 节). 在 4.4.4 节中列出的条件 (a)–(d) 是保证在 (4.29) 中定义的 MRK 估计量 $\boldsymbol{K}(\boldsymbol{Y}_\tau)$ 正定性的重要条件.

4.8 习题

1. 考虑微观结构噪声模型 (4.4). 假设 $dX_t = \sigma dB_t$, B_t 为标准布朗运动. 假设 ε_i 独立同分布, 服从 $[-a, a]$ 上的均匀分布, 且与 X_t 独立. 令 $\Delta_i = Y_{t_i} - Y_{t_{i-1}}$, $t_i = \dfrac{i}{n}$.

   (a) 证明 $E(\varepsilon_i^2) = \dfrac{a^2}{3}$, $E(\varepsilon_i^4) = \dfrac{a^4}{5}$.

   (b) 定义 $\hat{a}$ 为 a 的 QMLE. $\hat{a}$ 的渐近分布是? (提示: 使用 (4.23)、(4.24) 和 Delta 方法.)

2. 考虑 4.2.1 节中的 Roll 模型. 假设有效价格过程的形式为 $P_t^* - P_0^* = \sigma B_t$, 其中 B_t 为标准布朗运动. 定义已观测的交易价格变化向量为 $\boldsymbol{D} = (D_1, \ldots, D_n)^\top$, 即 $D_1 = P_{t_1}$, 当 $i = 2, \ldots, n$ 时, $D_i = P_{t_i} - P_{t_{i-1}}$ 及 $t_i - t_{i-1} = \dfrac{1}{n}$. c 的 QMLE 的渐近分布是什么?

3. 考虑 (4.32) 式.

   (a) 利用分部积分方法, 证明 $a_k(dX)$ 可以写成如下形式:

   $$a_k(dX) = \frac{X_{2\pi} - X_0}{\pi} + \frac{k}{\pi}\int_0^{2\pi} X_t \sin(kt)dt. \tag{4.81}$$

   (b) 证明 (4.34) 可由下式求得:

   $$\hat{a}_k(dX) = \frac{X_{2\pi} - X_0}{\pi} + \frac{k}{\pi}\sum_{i=0}^{n}\int_{t_{i-1}}^{t_i} \tilde{X}_t \sin(kt)dt,$$

   其中对于任意 $s \in (t_i, t_{i+1}]$, $\tilde{X}_s = X_{t_i}$, 即 $\tilde{X}_s$ 为 X_t 的分段常数近似.

   (c) 利用 (a) 和 (b) 中的方法, 证明 (4.35).

   (d) 讨论是否可以在每个区间 $(t_{j-1}, t_j]$ 中, 利用分段线性近似方法替代分段常数近似方法来估计 X_t.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

4. 4.5.1 节中提到 Σ_t 的傅里叶系数可以更一般地表示为

$$a_0(\Sigma) = \underset{N_x\to\infty}{\text{p-lim}} \frac{\pi}{N_X+1-n_0} \sum_{j=n_0}^{N_X} \left[a_j^2(dX)+b_j^2(dX)\right],$$

$$a_k(\Sigma) = \underset{N_x\to\infty}{\text{p-lim}} \frac{2\pi}{N_X+1-n_0} \sum_{j=n_0}^{N_X} \left[a_j(dX)a_{j+k}(dX)\right], \quad \text{当 } k>0,$$

$$b_k(\Sigma) = \underset{N_x\to\infty}{\text{p-lim}} \frac{2\pi}{N_X+1-n_0} \sum_{j=n_0}^{N_X} \left[a_j(dX)b_{j+k}(dX)\right], \quad \text{当 } k\geqslant 0.$$

Malliavin 和 Mancino (2002, 2009) 建议选取 $n_0=1$ 来消除漂移项的影响. 请对此进行具体阐述. $\left(\text{提示: } a_0(dX)=\hat{a}_0(dX)=\dfrac{X_{2\pi}-X_0}{2\pi}.\right)$

5. 利用文件 `NASDAQ_ITCH.txt` 中的数据, 使用如下子样本抽样方法估计已实现方差.

   (a) 4.3.1 节中的稀疏抽样方法.

   (b) 4.3.3 节中的两种时间尺度抽样方法.

   对于 1 秒、10 秒和 30 秒区间以及 $k=1,5,10$ 进行子抽样. 同时, 在 $H=10$ 时, 估计 4.3.4 节中的已实现核.

6. 文件 `hkex.txt` 包含香港交易所 2014 年 3 月 11 日 HKEx 的交易数据. 使用上午交易的数据估计 (4.4)–(4.5) 中的 $[X]_t$. 步骤如下:

   (a) 求出 $\delta=20,40,60$ 秒情况下, 4.3.6 节提到的 QMLE, 并估计 v_ε;

   (b) 同样在 $\delta=20,40,60$ 秒情况下, 令 $H=5$, 求出 (4.17) 中的已实现核函数;

   (c) 求傅里叶统计量 $\widehat{[X]}_t^{(F)}=2\pi\hat{a}_0(\Sigma)$, 其中 $\hat{a}_0(\Sigma)$ 由 (4.36) 定义, $N_X=20$.

7. 文件 `hsbc.txt` 包含香港交易所 2014 年 3 月 11 日 HSBC 的交易数据. 将该数据与第 6 题中的数据结合, 求解上午交易的 HKEx 和 HSBC 的已实现方差. 步骤如下:

   (a) 在 $\delta=60$ 秒情形下, 求解 4.4.3 节中的 QMLE;

   (b) 在 $H=5$ 时, 类似 (4.28), 求解已实现核函数;

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

(c) 在 $N_X = 20$ 时, 求解傅里叶统计量.

8. 利用第 7 题中的 HSBC 数据对于 4.6.1 节中的 EACD 和 WACD 模型进行拟合, 并比较哪个模型拟合效果更好.

9. 使用第 7 题中的 HSBC 数据拟合 4.6.4 节中的模型, 并检验 $\mathrm{Cov}(\tilde{A}_i, \tilde{A}_{i-1})$ 是否显著为负.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

第五章　限价指令簿: 数据分析和动态模型

第四章主要针对高频交易的价格和时间进行计量经济分析, 而在本章中则主要对于限价指令簿 (LOB, 也称为限价订单簿) 进行研究. 4.1 节对于限价指令簿进行了简要介绍, 此处的 "簿" 并不是指实际上的纸质簿. 实际上, LOB 是一个在全球各交易市场中广泛使用的匹配买方和卖方的交易机制, 交易双方的指令都会被记录在一个匿名的交易簿中. 在任何时间, LOB 都包含买卖双方的多层 (或多级) 信息, 每一层对应于一个不同的价格水平, 相邻的层级之间的价格差为最小价格增量. 对于大多数美国的证券交易所而言, 大多数股票的最小价格增量为 $0.01. 对于远期而言, 最小价格增量等于标的合同的名义价值. 交易员在进行金融资产的买卖时, 可以发出不同的指令. 限价指令和市场指令是最常见的两种指令方式. 限价指令是指在特定的价格下买入或卖出特定份额的股票的指令. 当限价指令发送至交易所的交易管理系统时, 指令中提到的股票便加入特定价格的买卖序列. 市场中的其他交易员在 LOB 进入系统时能看到的唯一变化是特定价格水平下序列长度的变化, 而与指令发生者相关的信息并不会在交易所内公开传播. 市场指令是指在当前市场下买入或卖出特定份额的股票的指令. 在大多数情形下, 一个卖出 1000 股的指令会使交易簿中该股票买方侧最高层级降低 1000 单位的流动性. 当最高层级的流动性低于 1000 单位时, 则该指令会降低下一个层级的流动性, 直到降低的流动性的总量达到 1000 单位为止. 市场指令的优势在于其基本上肯定会被执行. 然而, 当市场流动性不

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

足时, 最终执行指令时会以一定可能性在远低于最佳买入或卖出价格的情况下执行. 因此, 大多数交易所提供了多种方式来保护市场指令, 避免指令在预先设定的价格区间外执行, 或者在最高层级以外交易.

5.1 节和 5.2 节对于交易所公布的指令簿信息的解析和 LOB 的构建进行介绍, 并介绍了 LOB 的典型事实. 5.3 节介绍了对于 LOB 数据进行最大似然拟合的二元 Hawkes 过程, 并利用实际数据给出了拟合实例. 5.4 节介绍了 LOB 数据分析相关的机器学习方法. 5.5 节讨论了分析 LOB 变化的排队模型及其应用. 5.6 节为补充材料和习题.

5.1 从市场数据到限价指令簿 (LOB)

市场交易事件的信息传播 交易所的指令簿中的信息有两种传播方式: 基于订单的传播方式和基于价格水平的传播方式. 有的交易所只公布基于价格水平的数据, 如芝加哥商业交易所 (CME); 而一些其他的交易所同时以两种方式公布数据, 如洲际交易所 (ICE).

在基于价格水平的订单簿中, 交易所在与特定信号相关的交易事件发生时, 对于订单簿进行快照更新. 交易事件可以是在现有订单簿中增加一个订单 (ADD)、取消现有订单 (CXL) 或取消现有订单并替换为新订单 (CXR)[①]. 基于快照信息, 交易所可以构建包含市场化的价格水平和各交易水平的交易量的信息的订单簿. 当现有交易取消时, 交易所会对该信息进行公布, 并更新各价格水平的状态.

基于订单传播的订单簿提供的信息量多于基于价格水平传播的订单簿. 在基于订单的订单簿中, 交易所可以构建一个包含订单数量和各订单更细节化的信息. 基于这些信息, 可以判断订单序列中各个订单分别来自哪个客户. 相比而言, 构建基于订单的订单簿对于资源的需求更高, 且与设备引擎非常相关. 因此, 在构造一个基于订单的订单簿时, 存在一个数据信息量和潜在隐患之间的权衡.

数据解析和订单簿构建 在接收到交易所的信息后, 高频交易平台的订单簿引擎对于信息进行解析, 并根据收到的订单事件的类型对于相应的订单状态进行修改. 举例来说, 对于 ADD 事件, 引擎在给定的价格水平上更新信息, 当该价格水平不存在时, 新建一个价格水平, 然后将该订单作为新的价格水平序

① 在 CME 等交易所中, 订单的执行和取消都当作订单去除事件. 区分两种事件的唯一办法是基于后续的执行信息来判断.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

列的第一项. 订单构建过程的有效执行非常重要, 当订单事件数量很大时, 一个非最优的订单簿引擎会导致订单交易策略的执行时间大大增加.

5.2 LOB 数据的特征

图 5.1 利用一分钟内快照对于限价订单簿发展过程进行了阐述. 图中每一行代表交易执行之前和之后订单簿的状态. 上面的虚线表示最优买入价, 底部的虚线表示最优卖出价. 两条线的距离即为买入–卖出差价. 中间的实线表示市场中间价, 即为最优买入价和最优卖出价的平均值. 订单簿的上半部分, 即图中市场中间价上面的部分, 包含卖出的订单, 而订单簿的下半部分包含买入的订单. 阴影柱状图表示各价格水平上的交易量. 出于可视化需要, 订单簿中不同时间的价格序列都以该时间最大交易量为单位进行缩放. 灰色柱表示交易发生之前订单簿的状态, 而浅灰色柱表示交易发生之后订单簿的状态.

在价格变化时, 可以发现订单簿中买卖两侧的订单序列的形状变化并不相同. 订单簿对于潜在的市场供需不平稳给出相应的信号, 订单簿的变化也传递出投资者对于价格变化的反应信息.

5.2.1 订单簿的价格调整机制

订单簿中的典型价格调整机制包括以下几个关键部分: 市场订单或可销售限价订单[①] 的出现、执行、取消、价格变化以及短期价格平稳. 举例来说, 在 LOB 中, 在一个大批量卖出股票的订单出现后的典型价格变化过程包括:

- 该卖出订单出现在交易所, 并取出订单簿中买入侧一个或多个价格水平的订单 (09:00:22.860);
- 取消订单簿买入侧相同价格水平或低于该价格水平的订单 (09:00:40.625);
- 增大买入–卖出价差 (09:00:43.625);
- 卖出订单到达, 消除买入–卖出价差 (09:00:43.576);
- 合并订单簿中买入或卖出侧的订单, 市场达到一个短期的平稳状态 (09:00:47.922).

① 市场限价订单是价格在市场最低价或略高于最低价时的买入订单, 或者在市场最高价或略低于最高价时的卖出订单. 因此, 市场买入订单是价格极限为无穷时的市场限价买入订单. 类似地, 市场卖出订单是价格极限为 0 时的市场限价卖出订单. 另外, 市场化订单是指市场限价订单或市场订单.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

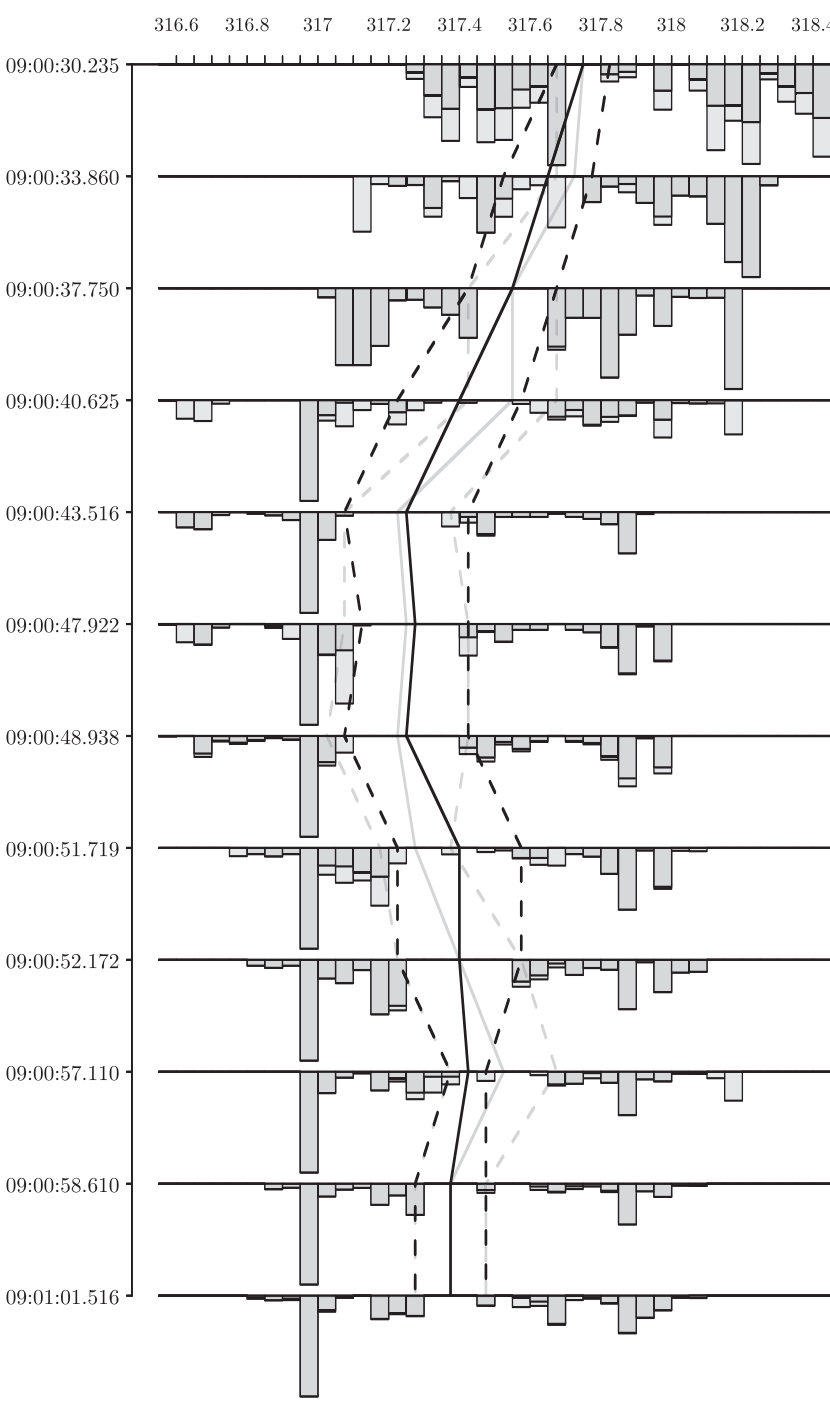


图 5.1 10 级的限价订单簿在英国石油公司 (BP) 的交易发生前 (灰色线) 和发生后 (黑色线) 演变过程的快照. 虚线表示最优买入和卖出价格. 实线表示中间价格的平均值. 在订单簿中, 柱状图按照最大序列长度进行比例缩减, 以深灰色和浅灰色表示, 便于区别订单簿在交易发生前后的变化

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

价格变化通过由占订单簿一定比例的一个大批量的订单或一系列小订单引起. 当交易所发布订单簿某一侧的交易标的显著减少的信息时, 一些市场参与者会认为此时市场上有内部交易者, 市场的公平价格会因此改变, 因而会减持标的资产. 交易标的的消费和订单的取消会首先使得买入–卖出价差变大. 而价差的增大会反过来使得做市商提交价格改进订单, 以资本化买入–卖出价差. 对于买入–卖出价差是否过大的判断是基于价格动态 (如波动率) 和策略本身 (如策略中当前的存货和效用) 的函数.

5.2.2 交易量不平衡和其他指示量

交易量不平衡通常简单定义为买入订单总交易量和卖出订单总交易量的差值, 然而该测度忽略了一个重要的事实, 即订单簿中不同价格水平的订单被执行的概率是显著不同的. 一个合理定义的买入–卖出不平衡的测度应当能够提取短期内交易方向和交易频度的精确信息. 基于该理论, 交易者在对于一个交易价格非最优卖出价或最优买入价的限价订单进行分析时, 可以首先求出交易完成时间的实证累积概率函数方程, 然后分析在特定时间区间内的概率密度函数. 图 5.2 给出了 BP 公司订单在提交后能够在 5 秒内完成的相对频率 (实证概率). 此处设定频率为交易价格与最优价格之间的距离的函数, 且标的实际

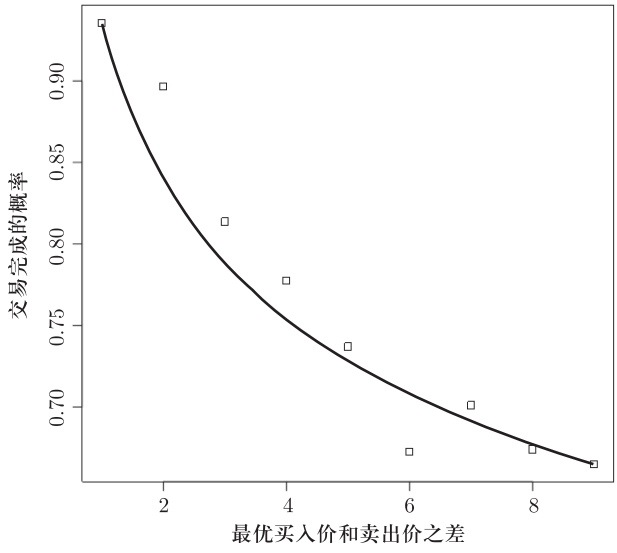


图 5.2 2010 年 6 月 25 日, BP 交易提交之后能在 5 秒内完成的概率. 方块表示基于实证数据的相对频率, 实线则为基于 Bouchaud 等 (2002) 建议的幂函数的拟合曲线

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

价格过程中并不存在漂移项, 这个假设在短期时间框架内是合理假设. 举例来说, 对于一个限价卖出订单而言, 距离为 0 指交易价格即为最优买入价格 (即该订单为有效市场订单), 距离为 1 指交易价格比最优买入价高一个单位. 图中的幂函数方程为非线性回归方程, 形式为 $y = cx^{-\alpha}$.

概率权重交易量 为了对于市场的供给和需求进行量化, 定义如下的概率权重交易量. 定义 $\pi_{l,i,\tau} = P(t_f < t + \tau | l, i)$ 为一个在 t 时刻提交的订单, 在价格水平 l 下, 能够在提交后的 τ 秒内顺利执行 (t_f 为实际交易时间) 的概率, 其中 $i \in \{1, 2\}$, $i = 1$ 代表买入侧, $i = 2$ 代表卖出侧. 定义 $v_{t,l;i}$ 为时刻 t, 价格水平为 l 时, 买入/卖出侧的交易序列的长度. 因此, 概率权重交易量可以定义为

$$\bar{v}(t, \tau, L; i) := \frac{1}{\sum_{i,l} v_{t,l;i}} \sum_{l=0}^{L} v_{t,l;i} \pi_{l,i,\tau}. \tag{5.1}$$

概率 $\pi_{l,i,\tau}$ 通常通过对于 LOB 买入/卖出侧在特定价格水平下的订单交易频率进行实证研究得到. 图 5.3 给出了时刻 $t \in \{t_i : i = 0, 1, \ldots, n\}$ 时, 概率权重累积交易量的买入和卖出侧的交易量之差 $\bar{v}(t, \tau, L; 1) - \bar{v}(t, \tau, L; 2)$ 的时间序列图. 该变量对于以交易概率为权重时的限价订单簿的不平稳特征进行了简要刻画.

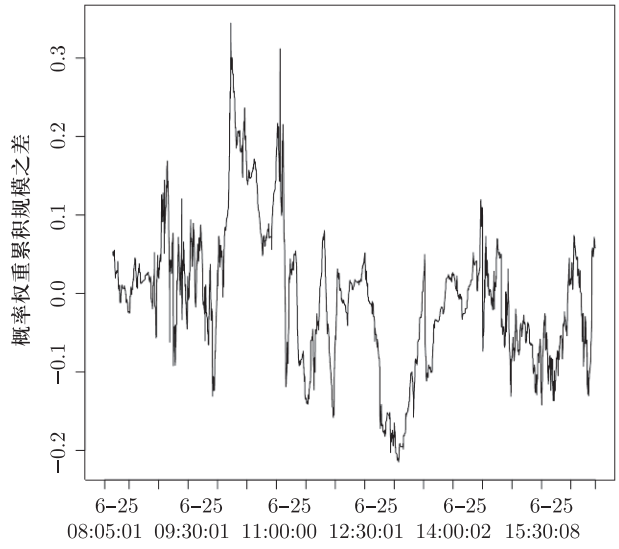


图 5.3 BP 公司在 2010 年 6 月 25 日的概率权重累积交易量之差 $\bar{v}(t, \tau, L; 1) - \bar{v}(t, \tau, L; 2)$ 的时间序列

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

不考虑 EMH 的争议情况下, 当交易所的价格水平为离散值时, 交易发生时资产的公平价值也存在一定的不确定性. 举例来说, 假设 BAC (美国银行) 刚刚的交易价格为 15.89, 而市场的买入报价为 15.88, 卖出报价为 15.89, 此时 BAC 的公平价值仍然存在不确定性, 尽管人们可能认为实际价格应该更接近于 15.89 而不是 15.88. 估计资产的公平价值的一个方法是利用 LOB 中包含的信息进行如下分析.

逆交易量权重公平价值 (IWFV) 价格 定义 p_0^B 和 v_0^B 分别是价格水平为 1 时的买入价和交易量, 类似地, 定义卖出价和交易量,[①] 则逆交易量权重公平价值价格定义为

$$\widehat{P}^{(\mathrm{IWFV})} = \frac{p_0^B v_0^A + p_0^A v_0^B}{v_0^B + v_0^A}.$$

该定义的经济学原理在于, 卖出侧的交易量越高, 资产卖出的压力越大, 因此买入价格的权重应该相应地越大.

阈值流动性公平价值 (TLFV) 价格 对于给定的阈值 θ, 时刻 t 时买入侧和卖出侧资产消费水平 L_t^B 和 L_t^A 分别为

$$L_t^B = \inf\left\{L : \sum_{l=0}^{L} v_{l,t}^B \geqslant \theta\right\}, \quad L_t^A = \inf\left\{L : \sum_{l=0}^{L} v_{l,t}^A \geqslant \theta\right\},$$

阈值流动性公平价值定义为

$$\widehat{P}_t^{(\mathrm{TLFV})} = \frac{1}{2\theta}\left[\theta p_{L_t^B,t}^B + \sum_{l=0}^{L_t^B-1} v_{l,t}^B (p_{l,t}^B - p_{L_t^B,t}^B) + \theta p_{L_t^A,t}^A + \sum_{l=0}^{L_t^A-1} v_{l,t}^A (p_{l,t}^A - p_{L_t^A,t}^A)\right].$$

该定义的经济学原理与 θ 单位的资产的平均买入和卖出价格有关. 相比而言, $\widehat{P}^{(\mathrm{IWFV})}$ 始终是在最优买入和卖出价格之间, 而 $\widehat{P}_t^{(\mathrm{TLFV})}$ 则可能会在价格水平为 1 时的价格水平之外.

5.3 LOB 数据的多元点过程拟合

4.6.2 节对于一元点过程模型进行了介绍, 尤其是 Hawkes 自激励点过程模型, 并利用模型基于 TAQ 数据对日内交易密度进行了建模. 图 5.1 类型的 LOB 数据则包含订单簿中买入侧和卖出侧多个价格水平的数据. Abergel 和

① 价格水平为 1 通常是指最优买入价 (BBO), 即最低卖出价和最高买入价同时出现.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

Jedidi (2015) 提出了一种 Hawkes 过程的多元拓展模型, 以对于多个价格水平下的数据进行建模.

多元 Hawkes 过程 当双随机点过程 $(N_t^1, \dots, N_t^K)^\top$ 的强度过程符合如下条件时:

$$\lambda_t^{(k)} = \mu^{(k)} + \sum_{j=1}^{K} \alpha_{jk} \int_{u<t} e^{-\beta_{jk}(t-u)} dN_u^{(j)}, \quad k = 1, \dots, K, \tag{5.2}$$

该过程即称为*包含指数核的 K-变量 Hawkes 过程*. 此处 "双随机" 是指点过程本身和其强度过程都是随机过程, 附录 C 对于自激励多元点过程给出了更详细的背景介绍.

5.3.1 市场化订单的多元点过程

(5.2) 式中的联合强度模型表明, 市场化订单的到达可以看作订单簿买入侧和卖出侧的自激励事件. 如图 5.4 所示, 在市场买入侧的新订单提交后, 买入侧和卖出侧订单数呈指数衰减. 对于某些类型的资产, 如美国交易所的证券, 其交易往往是以一手为单位进行交易的,① 一手为股票的正常交易单位, 通常包括 100 股或更大交易量的股票. 由于交易量是由交易的发起者所决定的, 交易量也是一个潜在的信息来源. 由于大批量交易相比于噪声交易而言, 其权重更高, 因此大批量交易中包含的信息量也更多. (5.2) 式中的基准强度 $\mu^{(k)}$ 包含了市场订单的平均到达速度的信息. 当订单簿中买方侧的限价订单多于卖方侧的限价订单时, 价格上升的概率增大, 反之则价格上升的概率减小. 在点过程中反映该特征的一个方法是利用概率权重的交易量 (5.1) 来对于强度过程的基准强度 $\mu^{(k)}$ 进行比例缩放. 在加入这些特征和其他外生参数后, 强度过程包含的信息集会更加丰富.

考虑 $K = 2$ 时的情形. 定义 w_{1i} 和 w_{2j} 分别为时刻 t_i 和 t_j 时, 买方侧和卖方侧的订单交易量. 则强度过程可以写成:

$$\begin{aligned}
\lambda_t^{(1)} &= \mu^{(1)} \bar{v}_{2t} + \frac{1}{\bar{w}_1} \sum_{t_i<t} \alpha_{11} w_{1i} e^{-\beta_{11}(t-t_i)} + \frac{1}{\bar{w}_2} \sum_{t_j<t} \alpha_{12} w_{2j} e^{-\beta_{12}(t-t_j)}, \\
\lambda_t^{(2)} &= \mu^{(2)} \bar{v}_{1t} + \frac{1}{\bar{w}_2} \sum_{t_j<t} \alpha_{22} w_{2j} e^{-\beta_{22}(t-t_j)} + \frac{1}{\bar{w}_1} \sum_{t_i<t} \alpha_{21} w_{1i} e^{-\beta_{21}(t-t_i)},
\end{aligned} \tag{5.3}$$

① 零星交易是指交易量低于一手的交易, 这些交易在交易序列中往往并不能占据很好的位置, 且交易所收取的清算费用可能更高. 具体可参见交易所公布的相关条款, 以对零星交易和其他交易的规定和要求有更加深入的了解 (如参见 NYSE 第 124(c) 条规则).

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

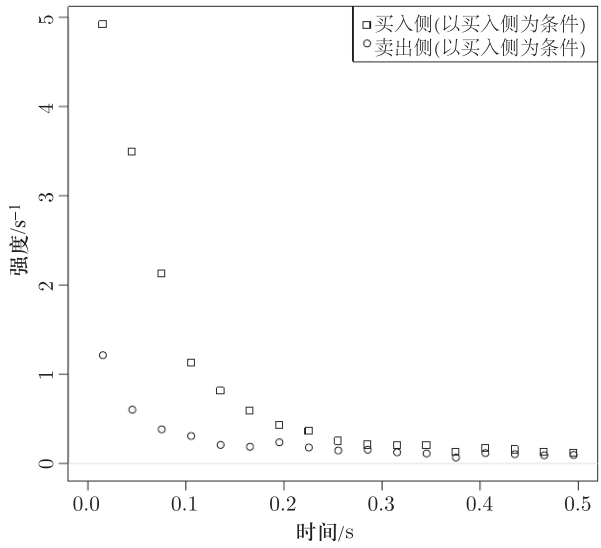


图 5.4 在市场买入侧的新订单提交后, 20 毫秒至 500 毫秒时间范围内买入侧和卖出侧订单数的条件密度图. 数据来源为 2010 年 6 月 25 日 BP 公司的逐笔数据

其中 $\bar{w}_1$ (或 $\bar{w}_2$) 为交易量 w_{1i} (或 w_{2i}) 在 $[0,t)$ 时间内的平均值, $\bar{v}_{1t}$ 和 $\bar{v}_{2t}$ 分别是买入订单和卖出订单的概率权重交易量. 强度过程的对数似然函数可以写成 $\mathcal{L}(\theta_1)+\mathcal{L}(\theta_2)$, 其中 $\theta_1=(\mu^{(1)},\alpha_{11},\alpha_{12},\beta_{11},\beta_{12})$, $\theta_2=(\mu^{(2)},\alpha_{21},\alpha_{22},\beta_{21},\beta_{22})$, 且有

$$\begin{aligned}\mathcal{L}(\theta_1)=&-\mu^{(1)}\sum_{i:t_i<T}\bar{v}_{2t_i}(t_i-t_{i-1})-\frac{\alpha_{11}}{\beta_{11}}\sum_{i:t_i<T}\frac{w_{1i}}{\bar{w}_1}\left(1-e^{-\beta_{11}(T-t_i)}\right)\\&-\frac{\alpha_{12}}{\beta_{12}}\sum_{j:t_j<T}\frac{w_{2j}}{\bar{w}_2}\left(1-e^{-\beta_{12}(T-t_j)}\right)\\&+\sum_{i:t_i<T}\log\left(\mu^{(1)}\bar{v}_{2t_i}+\alpha_{11}R_{11}(i)+\alpha_{12}R_{12}(i)\right),\end{aligned}\tag{5.4}$$

其中 R_{11} 和 R_{12} 由以下递归过程给出:

$$\begin{aligned}R_{11}(i)&=\frac{w_{1i}}{\bar{w}_1}e^{-\beta_{11}(t_i-t_{i-1})}\left[1+R_{11}(i-1)\right],\\R_{12}(i)&=\frac{w_{2i}}{\bar{w}_2}\left\{e^{-\beta_{12}(t_i-t_{i-1})}\left[R_{12}(i-1)\right]+\sum_{j':t_{i-1}\leqslant t_{j'}<t_i}e^{-\beta_{12}(t_i-t_{j'})}\right\},\end{aligned}$$

$\mathcal{L}(\theta_2)$ 类似可得. 附录 C.3 给出了相关的背景和求解过程.

Abergel 和 Jedidi (2015) 使用 K 维向量 $\boldsymbol{a(t)}=(a_1(t),a_2(t),\ldots,a_K(t))$ 和 $\boldsymbol{b(t)}=(b_1(t),b_2(t),\ldots,b_K(t))$ 分别表示时刻 t 时, 订单簿的买方侧和卖方

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

侧变量. 其中 $a_i(t) \geqslant 0$ 表示买方侧与最佳买入价格相差 i 个单位的价格水平上的可供交易的标的数量, 相应地 $b_i(t) < 0$ 表示卖出侧的标的数量. 换句话来说, 该模型包含了一个 $2K$ 维的*移动结构*价格网格, 且最小交易单位即为网格的格点. 价格水平 i 的累积可供交易标的数量为

$$A_i(t) = \sum_{k=1}^{i} a_k(t), \quad B_i(t) = \sum_{k=1}^{i} |b_k(t)|,$$

因此其广义逆函数分别为

$$A_t^{-1}(x) := \inf\left\{p : \sum_{k=1}^{p} a_k(t) > x\right\}, \quad B_t^{-1}(x) := \inf\left\{p : \sum_{k=1}^{p} |b_k(t)| > x\right\},$$

而买入–卖出价差为 $A_t^{-1}(0) - B_t^{-1}(0)$. Abergel 和 Jedidi (2015, 第五节) 使用这些广义逆函数对于最优买入价格和最优卖出价格进行了建模.

5.3.2 实证结果

伦敦证券交易所 (LSE) 的数据集*订单簿重建*包含市场上所有的日内交易信息和交易数据. 我们根据这些数据对于完整的订单簿进行了重建. 数据集包含订单的详情、订单的删除信息和交易信息, 这些记录以最靠近的毫秒作为记录时间. 本节使用 2010 年 6 月 25 日英国石油公司 (BP) 的数据来阐述建模框架和数据拟合过程. 图 5.5 给出了买入价格和买入–卖出价差的时间序列图. 用于订单簿重建的计算机程序首先从*订单详情记录*文件中读取相关数据, 然后与*交易历史记录*数据中的交易和删除信息进行时间匹配. 利用该重建过程, 可以在每一个交易事件发生时, 生成订单簿中的时间序列图. 本书第八章对于市场数据、订单类型、匹配引擎和 LOB 重建进行了更加全面的介绍.

此处选取的样本时间为 2010 年 6 月 25 日全天, 即从 08:05:00.000 至 16:25:00.000. 由于交易日开盘后的五分钟和收盘前的五分钟的数据中往往包含大量的干扰项 (如错误记录的数据), 因而被舍弃. 图 5.1 给出了 BP 股票的限价订单簿的 30 秒变化过程. 市场中主要包括四种交易类型: 限价买入、限价卖出、有效市场买入和有效市场卖出. 有效市场买入类型的订单包括市场买入订单和以最优交易买入价格成交的限价订单. 类似地, 有效市场卖出类型的订单包括市场卖出订单和以最优交易卖出价格成交的限价订单. 这两种类型的订单在提交时都是希望能够立即成交. 对于买方侧的订单强度, 到达时间间隔定义为两个买方侧有效市场买入类型订单之间的时间差, 以毫秒为单位. 相应地, 可定义卖方侧的订单强度. 图 5.6 给出了买方侧和卖方侧订单的实证强度. 此

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

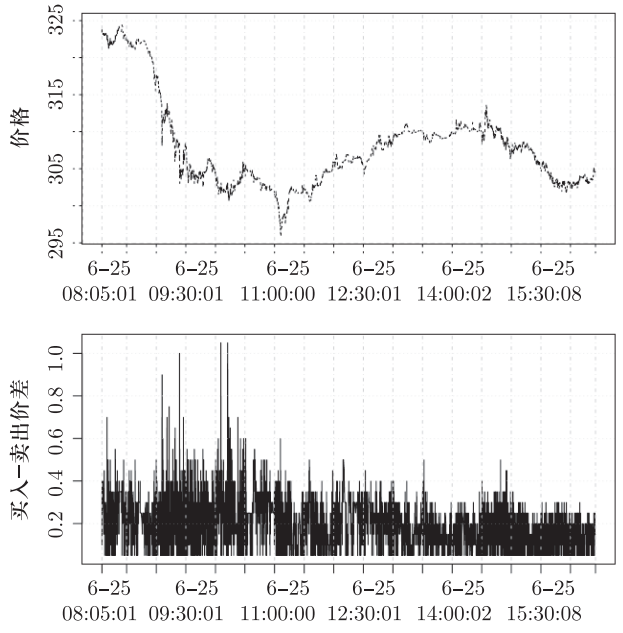


图 5.5　2010 年 6 月 25 日, 订单簿中 BP 股票的最优现行买入价格 (上图) 和买入−卖出价差 (下图)

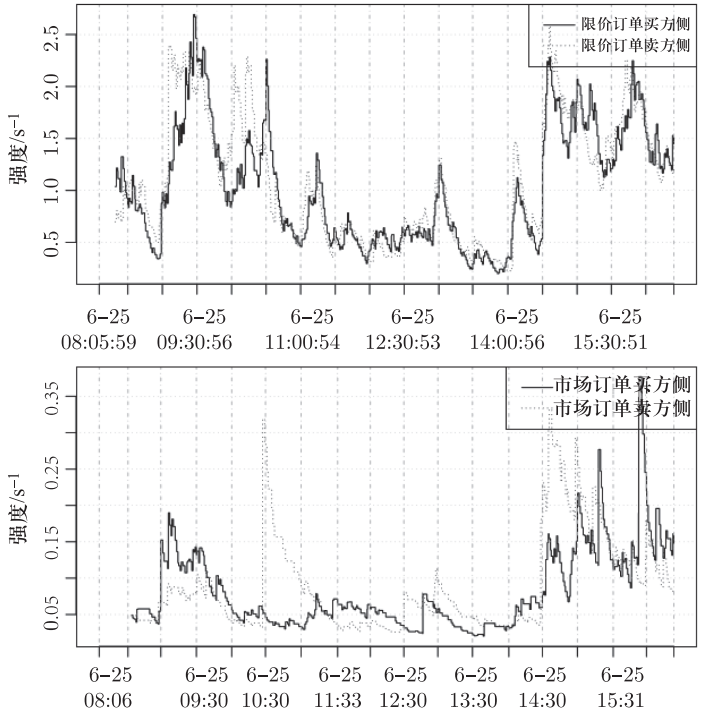


图 5.6　订单簿买方侧和卖方侧有限订单 (上图) 和市场订单 (下图) 的到达时间强度, 数据来源于 2010 年 6 月 25 日 BP 公司数据, 利用一分钟的重叠窗宽进行估计

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

处, 强度是基于 1 分钟为单位的重叠窗宽, 对于市场订单的到达率实证计算所得. 图中的 U 形特征表明价格发现主要集中在交易日的开盘和收盘阶段.

最大似然估计 在本节中利用二元联合过程模型 (5.3)对数据进行拟合, 拟合所使用的四种模型在表 5.1 中给出. 表 5.1 同时给出了各模型下各个参数的最大似然估计量. 根据对数最大似然估计量可知, 模型 1 和模型 2 的拟合效果显著好于模型 3 和模型 4. 在模型 1 和模型 2 中, $\hat{\mu}^{(1)}$ 显著大于 $\hat{\mu}^{(2)}$, 且自激

表 5.1 (5.3) 中四种设定下的 MLE. 模型 1: $\bar{v}_{1t} = \bar{v}_{2t} \equiv 1$, 且 $w_{1i} = w_{2i} \equiv 1$; 模型 2: $w_{1i} = w_{2i} \equiv 1$; 模型 3: $\bar{v}_{1t} = \bar{v}_{2t} \equiv 1$; 模型 4: 参数 $\bar{v}_{1t}, \bar{v}_{2t}, w_{1i}, w_{2i}$ 不受约束. 标准差在括号中给出, † 指该值在 0.05 的置信水平下并不显著不等于 0

	模型 1	模型 2	模型 3	模型 4
$\mu^{(1)}$	0.068	0.193	$0.000^{\dagger}$	$0.000^{\dagger}$
	(0.002)	(0.005)	(0.000)	(0.000)
$\mu^{(2)}$	0.005	0.015	0.005	0.013
	(0.001)	(0.002)	(0.001)	(0.002)
α_{11}	2.726	2.803	2.901	2.900
	(0.114)	(0.116)	(0.129)	(0.129)
α_{22}	2.624	2.631	2.424	2.430
	(0.126)	(0.126)	(0.126)	(0.126)
α_{12}	0.575	0.563	0.004	0.004
	(0.068)	(0.068)	(0.000)	(0.003)
α_{21}	0.002	0.002	0.001	0.001
	(0.000)	(0.000)	(0.000)	(0.000)
β_{11}	5.211	5.418	6.740	6.740
	(0.217)	(0.224)	(0.295)	(0.295)
β_{22}	6.990	7.016	7.659	7.684
	(0.347)	(0.349)	(0.400)	(0.400)
β_{12}	8.422	8.204	0.007	0.007
	(1.076)	(1.072)	(0.001)	(0.001)
β_{21}	0.004	0.004	0.002	0.002
	(0.001)	(0.000)	(0.000)	(0.000)
$\sup_\theta l(\theta)$	−16319	−16208	−18123	−18120

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

励过程的参数 α_{11} 和 α_{22} 显著大于交叉激励过程中的对应参数. 由此可知, 尽管 LOB 中买方侧或卖方侧一个新的订单的提交会对于同一侧和另一侧的订单状态都产生影响, 但对于同一侧的影响要远远大于对另一侧的影响. 交叉激励过程的衰减率 β_{12} 也显著大于 β_{21}, 可知买方侧的市场订单对于卖方侧的影响要大于反向影响.

QQ 图和拟合优度　为了检验这四个模型的拟合优度, 我们使用 QQ 图 (参见 Lai 和 Xing (2008, 第 18 页)) 来比较实证得到的分位数和各模型的理论分位数. 对于强度函数为 λ_t 的点过程, 时变序列 $\Lambda_{t_i} = \int_0^{t_i} \lambda_u du$ $(i = 1, 2, \dots)$ 是一个泊松过程, 因此 $\Lambda_{t_i} - \Lambda_{t_{i-1}}$ 独立同分布, 服从均值为 1 的指数分布. 图 5.7 给出了四种模型对于 2010 年 6 月 25 日 BP 股票数据拟合的 QQ 图. 由图可知, 在模型 3 和模型 4 中, 虽然加入了订单交易量数据, 但拟合结果没有显著改变, 原因可能来自很多交易系统中都采用的分层算法. 这些算法将大批量的订单划分为与交易量中值同阶的各层交易, 以最小化信号作用的影响.

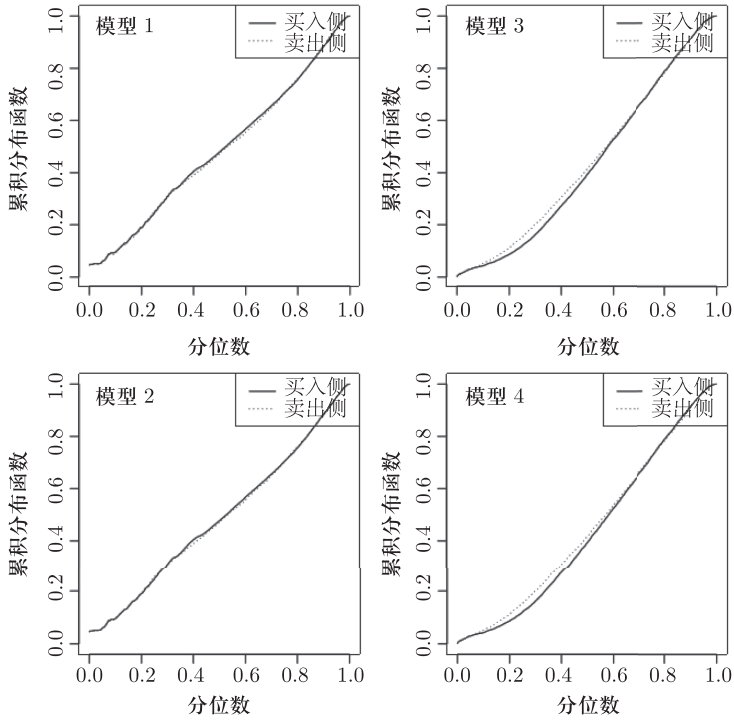


图 5.7　四种模型对于 2010 年 6 月 25 日 BP 股票数据的拟合 QQ 图

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

5.4 基于机器学习的 LOB 数据分析

Fletcher 等 (2010) 讨论了机器学习方法 (特别是支持向量机和多元核学习方法) 在分析 (LOB 中不同价格水平下的) 高维数据中的作用. 利用机器学习方法, 交易者可以在市场事件发生后, 对于订单簿的更新情况进行分析, 从而制定交易策略. Zheng 等 (2013) 也利用机器学习方法 (特别是 Lasso Logistic 回归方法) 对于 CAC40 指数中的 40 只最大的股票的价格跳跃过程进行预测, 其中 CAC40 指数是法国证券市场的基准指数. Kearns 和 Nevmyvaka (2013) (后面简称 KN) 对于机器学习在 LOB 数据分析中的应用进行了回顾, 论文指出:

> 基于历史数据来建立预测模型显然不是数量金融领域的一个新问题: CAPM 中系数估计、Fama-French 因子模型和相关的方法比比皆是. 用于 HFT 的机器学习方法的难点来源于数据的过于细化和对于价格水平较低的订单的行为的理解缺失. 每个订单本身、订单的 (部分) 执行或取消、订单的隐藏流动性中都经常包含微观结构数据, 而低价格水平订单的行为包括获利性购买或卖出、大批量订单的最优执行等. 从机器学习的角度考虑, 尽管 CAPM 模型及其变形已经分析了超额收益率、市盈率等参数的预测建模时相关的变量或 "特性", 但是在 HFT 问题中, 在分析与未来价格变动相关的订单簿中的流动性时, 其先验分布是未知的. 因此选取合适的 "特性"或特性工程是 HFT 的机器学习建模中的一个重要的过程, 这也是本节的重点.

KN 论文第 4 节将机器学习方法应用到 LOB 数据的短期价格变化方向预测问题中. 在该论文的一个案例中, KN 通过数据分析确定了交易的时间和方式, 其中交易时间和方式分别指在给定的状态空间下应当在何种情形下和或在哪个方向 (买入或卖出) 进行交易. 特别地, 论文中提出了以 "生成 α 为目的" 的数据分析的两个主要组成部分如下:

- 研究相关的 "特性", 从而对于 LOB 的特定状态, 能够对于价格变化的方向进行可靠的预测, 此处 "可靠" 是指 "使得获利性的交易的权重大于非获利性的交易", 而不是指大多数时间内预测正确.
- 研究学习算法, 从而以相对较低的成本, 将预测结果或 "α" 用于实际交易的执行.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

KN 对于上述案例以及另一个智能交易选择的机器学习方法 (该方法会在 7.5 – 7.7 节中具体讨论) 进行总结, 认为机器学习对于高频交易中的数据分析和预测提供了一个有效、可行且条理化的框架, 但 "获利并不容易". 正如 2.8 节和第三章提到的, 可盈利的量化策略不仅需要对于信息进行有效收集和分析以进行预测, 也需要开发相应的动态策略以优化前面提到的存在不确定时的业绩指标, 3.4 节补充材料 7 中提到的强化学习和 AlphaGo 即为典型例证.

特征工程和机器学习 特征工程是利用非参数回归方法来构建预测模型的重要技术. 当考虑时间因素时, 这些方法被用于研究移动时间窗问题. 机器学习是计算机学科的一个子学科, 其起源于计算学习理论和人工智能中的特征识别和强化学习. 机器学习在回归和聚类方面与统计学习有所重合 (参见 2.9 节补充材料 9). 在分析 LOB 数据方面, 聚类分析致力于对价格变化方向进行预测, 而回归分析则用于预测价格变化值. 此处回归分析包括线性回归分析、Logistic 回归和非线性回归以及前进法/后退法逐步方法以进行变向选择 (参见 Lai 和 Xing (2008), 第 1 章和第 4 章). 非参数回归 (参见 Lai 和 Xing (2008), 第 7 章) 通常是对于大量的基本函数的线性回归. 机器学习中通常涉及大量特征 (预测变量), 记为 p 个, 问题在于当 p 的值与 n 的阶数相同或大于 n 时, 如何进行数据处理. 利用前进法/后退法逐步特征选取方法和合适的选择准则, 可以选取特定部分的特征以建立模型. 另一个方法是在损失方程中加入正则项, 作为各系数的约束条件, 从而避免过度拟合情形.

线性回归、广义线性回归及其非线性扩展形式 (参见 Lai 和 Xing (2008), 第 1 章和第 4 章) 常常被用于在机器学习和特征工程中建立预测模型. 另一大类预测模型包括*决策树*. 决策树是一个树形图, 各枝叶为关于某一特性的条件 (如 $X_i > C$). 在构建决策树时, 首先构建根节点, 当根节点的条件满足时转向左侧分支, 如果条件不满足时则转向右侧分支. 决策树中的叶节点为样本的估计值. 拟合得到的决策树模型往往方差很大. 减小方差的一个办法是利用加权的训练样本建立足够多的决策树模型, 其中权重由之前得到的决策树的拟合结果确定. 接下来, 利用 Boosting 算法将这些决策树进行测试, 该过程可参见 Hastie 等 (2009) 的第 10 章, 特别是 AdaBoost 算法部分. 另一个常用的决策树模型是*随机森林*, 该算法利用样本的随机抽样子集来对于决策树进行训练. 当 p 很大时, Boosting 决策树方法和随机森林方法都可以给出各个特征的重要性评分, 便于选取合适的特征用于建模.

支持向量机 (SVM) 是另一种广泛使用的机器学习聚类方法, 该方法通过建立一个最大化两类样本的边际量的超平面将训练样本分为两类. 在分析超平

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

面时, 利用核函数来替代 SVM 中的内积, 可以将 SVM 方法扩展到非线性超平面的情况. 相比于高维基本开展方法, 核函数方法在计算时速度也更高, 具体参见 Hastie 等 (2009) 的第 12 章. 最后, 另一个常用的选取基本函数的方法为利用神经网络选取法, 参见 Hastie 等 (2009) 和 Lai 和 Xing (2008). Lai 等 (2016b) 和 Deng 等 (2017) 介绍了一种决策树和神经网络的综合方法. Deng 等 (2017) 同时介绍了统计学习和机器学习在金融科技中的其他应用.

5.5 LOB 动态的排队论建模

订单簿显然可以看作一个排队系统, 系统中的服务器来自于多个价格水平下的买方序列和卖方序列. 对于只考虑价格水平为 1 的数据的情形, 目前研究将排队论模型修正后用于 LOB 数据的建模, 并计算得到 LOB 未来的变化概率的公式. 首先, 建立 LOB 与多类别优先排队模型之间的联系, 特别是包含取消事件的 LOB 模型和包含中途退出的排队模型, 以及市场订单与优先排序中的服务在减少序列长度方面的联系. 同时, 与排队论中申请服务类似, 限价订单也会增加序列长度. 在排队论框架下对 LOB 建模时, 可以对于序列长度设定扩散和流动极限, 从而分析订单簿的价格的等待时间. 5.6 节的补充材料 3 对于 Kruk (2003) 中提出的拍卖情形下的扩散和流动极限以及渐近理论进行了介绍. 在进行缩放后, 最优买入和卖出序列的长度收敛至第一象限中的反射的二维布朗运动模型. Cont 和 de Larrard (2013) 与 Guo 等 (2015) 在该论文的基础上, 对于 LOB 中价格水平为 1 数据, 建立了简化形式排队论模型. Huang 等 (2015) 利用 LOB 中其他价格水平的数据提出了一个替代模型, 该模型会在介绍 Cont 和 de Larrard 与 Guo 等的模型后进行介绍.

5.5.1 价格水平为 1 的 LOB 数据的简化形式模型的扩散极限

实证结果 (如 Biais 等 (1995)) 表明大部分订单流出现在最优买入价格和最优卖出价格上. 受此结果启发, Cont 和 de Larrard (2013) 提出了一个简化形式下的模型, 该模型中的状态变量, 即限价订单簿买方侧的买入价和序列长度分别定义为 s_t^b 和 q_t^b, 卖方侧的卖出价和序列长度分别定义为 s_t^a 和 q_t^a. 为了对模型进行简化, 论文中进一步假设:

(a) 买入–卖出价差为常数 δ, 买入和卖出价是 δ 的倍数;

(b) 当最优买入序列耗尽时, 最优买入价和最优卖出价都下降一个单位;

(c) 当最优卖出序列耗尽时, 最优买入价和最优卖出价都上升一个单位;

(d) 当最优买入 (或卖出) 序列耗尽时, 会抽样生成一个新的最优买入 (或

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

卖出) 序列长度, 抽样过程基于密度函数 $\tilde{f}(\cdot, y)$ (或 $f(x, \cdot)$), 函数只与序列历史中最优买入序列长度的当前值 y (或最优卖出序列长度的当前值 x) 有关.

(e) 限价订单、市场订单和取消的订单的单位都为 1, 且其产生的过程分别服从参数为 λ, μ 和 θ 的独立泊松过程.

假设 (d) 也称为序列耗尽时序列长度的随机重新初始化, 其关键的设定在于基于序列减短时间 $\tau_i = \tau_i^a \wedge \tau_i^b$, 定义了离散状态空间 $\delta\mathbb{Z} \times \mathbb{N}^2$ 上的连续时间马尔可夫链 (s_t^b, q_t^b, q_t^a), 其中 $\tau_i^a = \inf\{t > \tau_{i-1} : q_t^a = 0\} - \tau_{i-1}$, $\tau_i^b = \inf\{t > \tau_{i-1} : q_t^b = 0\} - \tau_{i-1}$. 根据假设 (e) 可知, τ_i 独立同分布. 另外, 由于该模型的更新特性, 对于连续价格变动的分析可以被简化为对于第一次价格变动的分析. 因此在讨论时, 可将 (τ_1^a, τ_1^b) 和 τ_1 分别简化为 (τ_a, τ_b) 和 τ.

在排队论研究中, 在增消链、(τ_a, q_0^a) 和 (τ_b, q_0^b) 的独立性和稳定状态下买入侧和卖出侧的对称性的研究很多. 如在对称性的研究上, 研究表明

$$P(\tau_a > \tau_b | q_0^b = x, q_0^a = x) = P(\tau_b > \tau_a | q_0^b = x, q_0^a = x) = \frac{1}{2}, \tag{5.5}$$

则有

$$\begin{aligned} P(\tau > t | q_0^b = x, q_0^a = x) &= \frac{P(\tau_a > t | q_0^b = x, q_0^a = x)}{2} + \frac{P(\tau_b > t | q_0^b = x, q_0^a = x)}{2} \\ &= P(\tau_a > t | q_0^b = x, q_0^a = x) \\ &= P(\tau_b > t | q_0^b = x, q_0^a = x). \end{aligned} \tag{5.6}$$

(5.6) 式中的第一个等式由 (5.5) 式所得, 而 $P(\tau_a > t | q_0^b = x, q_b^a = x) = P(\tau_a > t | q_0^a = x)$ 则可根据买入序列和卖出序列的独立性得到. 因此 (5.6) 式可以进一步简化为

$$P(\tau > t | q_0^b = x, q_0^a = x) = P(\tau^b > t | q_0^b = x) = P(\tau^a > t | q_0^a = x).$$

此外, 利用增消链可以得到第一次到达状态 0 的时间的 Laplase 变换 $\mathcal{L}(s, x) = E(e^{-s\tau_b} | q_0^b = x)$ 的显式表达, 进而对 Laplace 变换结果求逆, 则可得到公式 (5.7), 具体参见 Cont 和 de Larrard (2013, 第 9–10 页).

给定 (q_0^a, q_0^b) 时, 耗尽时间 τ 的条件分布

$$P(\tau > t | q_0^b = x, q_0^a = y) = \sqrt{\left(\frac{\mu + \theta}{\lambda}\right)^{x+y}} \xi_{x,\lambda,\theta+\mu}(t) \xi_{y,\lambda,\theta+\mu}(t), \tag{5.7}$$

其中, 对于任意整数 ν,

$$\xi_{\nu,\lambda,\theta+\mu}(t) = \int_t^\infty \frac{\nu}{u} I_\nu \left(2\sqrt{\lambda(\theta+\mu)u}\right) \exp\left[-u(\lambda + \theta + \mu)\right] du,$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

I_ν 是第一类改进 Bessel 方程, 定义为

$$I_\nu(x)=\sum_{m=0}^{\infty}\frac{1}{m!\Gamma(\nu+m+1)}\left(\frac{x}{2}\right)^{2m+\nu}.$$

在满足 $\lambda=\mu+\theta$ 条件的特定情形下, 即考虑假设 (e) 时, 限价订单的平均交易量与市场订单与取消的订单之和相等时, LOB 具有*平衡订单流*. 该现象也令人回想起在经典的排队系统中, 当服务率与申请到达率相同时, 会出现*高负荷*情形.

$\lambda=\mu+\theta$ 情形下价格增长概率 $\pi^{\mathrm{up}}(x,y)$ 的计算 在买方侧包含 x 个订单、卖方侧包含 y 个订单时, 下一次价格变动会是价格增长的概率为

$$\pi^{\mathrm{up}}(x,y)=\frac{1}{\pi}\int_0^{\pi}(2-\cos(t)-\sqrt{(2-\cos(t))^2-1})^y\frac{\sin(xt)\cos\left(\dfrac{t}{2}\right)}{\sin\left(\dfrac{t}{2}\right)}dt. \quad (5.8)$$

假设 (e) 中的泊松到达条件, 即 τ_i 独立同分布的条件, 可以被放宽 (Cont 和 de Larrard, 2012, 3.2 和 3.3 节), 然而放宽后的概率计算公式可能会非常复杂, 特别是积分部分计算复杂. (5.8) 式的计算需要利用 ($\lambda=\mu+\theta$ 条件下) 对称的随机游走中, 初始值在平面的上半象限, 游走过程首先到达 x 轴而不是 y 轴的概率计算公式.

累积价格变化过程 s_t 及其扩散极限 与假设 (a) 中假定买入价格 s_t^b 满足 $s_t^a=s_t^b+\delta$ 不同, 此处我们通过分析如下时间 t 时总的价格变化次数来定义累积价格变化过程 s_t:

$$N_t=\max\{n\geqslant 1:\tau_1+\cdots+\tau_n\leqslant t\},\quad s_t=\sum_{i=1}^{N_t}X_i, \quad (5.9)$$

其中 $X_1,X_2,\ldots$ 为时间 τ_i 发生的价格的依次变化, 当 $q_{\tau_i^a}=0$ 时满足 $X_i=\delta$, 而当 $q_{\tau_i^b}=0$ 时, 满足 $X_i=-\delta$. Cont 和 de Larrard (2013, 引理 1 和定理 1) 进一步假设 $\lambda=\mu+\theta$, $\tilde{f}(x,y)=f(y,x)$, 以及

(f) 对于所有 $(x,y)\in\mathbb{N}^2$, $f(x,y)=f(y,x)$,

(g) $\sum_{x=1}^{\infty}\sum_{y=1}^{\infty}xyf(x,y)<\infty$,

定义假设 (g) 中的无限序列为 $D(f)$, 证明当 $n\to\infty$ 时, $P(X_i=\delta)=P(X_i=$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$-\delta)=\dfrac{1}{2}$, 且

$$\frac{(\tau_1+\cdots+\tau_n)}{n\log n}\xrightarrow{P}\frac{D(f)}{\pi\lambda}, \tag{5.10}$$

$$\left\{\frac{S_{\lfloor tn\log n\rfloor}}{\sqrt{n}},t\geqslant 0\right\}\ \text{弱收敛于}\ \left\{\delta\left(\frac{\pi\lambda}{D(f)}\right)^{1/2}B_t,t\geqslant 0\right\}, \tag{5.11}$$

其中 B_t 为标准布朗运动. (5.11) 中的弱收敛结果表明, 订单流通过因子 $\dfrac{1}{\sqrt{D(f)}}$ 来对于价格波动产生影响, 该结果也与之前 Rosu (2009) 的结果一致. Rosu (2009) 中基于 "在价格变化后买方序列和卖方序列的平均长度, 以股数计的 $\sqrt{D(f)}$", 建立了一个均衡理论来解释上述影响.

为了证明 (5.10) 式, Cont 和 de Larrard (2013) 首先利用 (5.7) 式, 证明当 $t\to\infty$ 时,

$$P[\tau>t|q_0^a=x,q_0^b=y]\sim\left(\frac{xy}{\pi\lambda}\right)t^{-1}; \tag{5.12}$$

参见论文中公式 (6). 在 (5.12) 式左右同时乘以 $f(x,y)$, 并对于 x,y 求和, 可知在 $t\to\infty$ 时, $P(\tau>t)\sim\dfrac{D(f)}{\pi\lambda t}$. 利用该结论, 论文中引理 1 证明 Laplace 变换 $\mathcal{L}(s)=Ee^{-s\tau}=1-s\int_0^\infty e^{-st}P(\tau>t)dt$ 满足对于所有的 s,

$$\lim_{n\to\infty}\left[\mathcal{L}\left(\frac{s}{n\log n}\right)\right]^n=\exp\left(\frac{-D(f)}{\pi\lambda}\right),$$

由 $[\mathcal{L}(s/(n\log n))]^n$ 是 $\dfrac{\tau_1+\cdots+\tau_n}{n\log n}$的 Laplace 变换可得 (5.10). 因为 τ 是特性指数 $\alpha=1$ 时稳定分布的吸引域, 即 $P(\tau>t)\sim$ 常数项$\times t^{-1}$, 所以 (5.10) 式也服从独立同分布变量的标准化求和项 $\dfrac{\sum_{i=1}^n\tau_i-b_n}{a_n}$ 的极限稳定分布, 具体参见 Durrett (2005, 第 151 页). (5.11) 式中的弱收敛结果则可根据 (5.10) 式和 Donsker 函数中心极限定理可得, 具体参见附录 D.

5.5.2 订单位置的流动极限

尽管将 LOB 的动态与多类排队系统问题相联系直观上很有吸引力, 且后者有大量的分析工具包可以使用, 但是这两者之间有着根本的区别. 经典的排队理论侧重于整个系统的稳定性. 而在 LOB 中, 对于个体订单的分析与 LOB 的整体状态的分析同样重要. 因此, 由于很多交易所的管理规定了价格–时间优先顺序 (即最优价格优先, 先进先出), 一个更好的订单位置意味着更短的等

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

待时间和更高的订单执行概率. 实际上, 减少交易中的低延迟排队, 获得更好的订单位置是高频交易公司的技术竞争的驱动因素之一. Moallemi 和 Yuan (2015) 最近的实证研究表明, 当订单位置被合理定义时, 其值大概在半个买入–卖出价差的量级. 分析订单位置是算法交易策略的关键问题. 订单位置也是限价订单被执行概率的重要指标. 正如图 5.8 指出, 最优买入序列中的订单位置同时受到市场订单和订单取消的影响, 其相对位置也受到限价订单的影响.

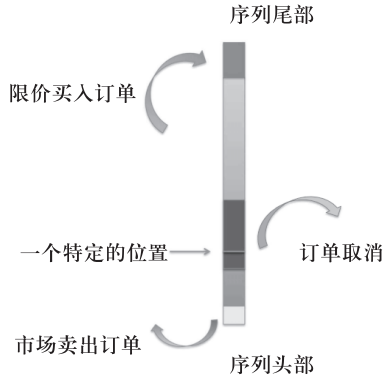


图 5.8 最优买入序列中的订单位置

Guo 等 (2015) 在此基础上将订单位置变化与相关的序列结合进行建模. 在最优买入序列和最优卖出序列中共有六类订单: 最优买入订单 (编号 1)、最优买入价的市场订单 (编号 2)、最优买入价的取消订单 (编号 3)、最优卖出订单 (编号 4)、最优卖出价的市场订单 (编号 5)、最优卖出价的取消订单 (编号 6). Cont 和 de Larrard 论文第 7 页在分析假设 (e) 中的二维变量 (V_i^a, V_i^b) 替换为 $(V_i^1, \ldots, V_i^6)$. 在 Cont 和 de Larrard 的文章中, V_i^a (或 V_i^b) 定义为买方侧 (或卖方侧) 第 i 个订单相关的买入 (或卖出) 序列长度的变化值, 其分布如下:

$$P(V_i^a = 1) = P(V_i^b = 1) = \frac{\lambda}{\lambda + \mu + \theta},$$
$$P(V_i^a = -1) = P(V_i^b = -1) = \frac{\mu + \theta}{\lambda + \mu + \theta}.$$

Guo 等 (2015) 类似地定义 V_i^j 为第 j $(j = 1, \ldots, 6)$ 类排队序列中第 i 个订单的变化. 在 $\boldsymbol{V}_i = (V_i^1, \ldots, V_i^6)$ 中, 只有一项为正, 其他项都为 0. 定义 T_i

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

为第 $i-1$ 个订单与第 i 个订单之间的时间间隔, 定义

$$M_t = \max\left\{m : \sum_{i=1}^{m} T_i \leqslant t\right\}. \tag{5.13}$$

由于 (5.9) 式为衰减时间间隔 τ_i 的更新过程, 因此相应地 (5.13) 式为到达时间间隔 T_i 的更新过程.

定义按比例的净订单流过程 $\boldsymbol{C}_n(t) = [C_n^1(t), \ldots, C_n^6(t)]$ 如下:

$$\boldsymbol{C}_n(t) = \frac{1}{n}\sum_{i=1}^{M_{[nt]}} \boldsymbol{V}_i, \tag{5.14}$$

定义按比例的订单位置为 $Z_n(t)$、按比例的最佳买入和卖出序列订单长度 $Q_n^b(t)$ 和 $Q_n^a(t)$ 如下:

$$\begin{aligned} Q_n^b(t) &= Q_n^b(0) + C_n^1(t) - C_n^2(t) - C_n^3(t), \\ Q_n^a(t) &= Q_n^a(0) + C_n^4(t) - C_n^5(t) - C_n^6(t), \\ dZ_n(t) &= -dC_n^2(t) - \frac{Z_n(t-)}{Q_n^b(t-)} dC_n^3(t). \end{aligned} \tag{5.15}$$

Guo 等 (2015) 指出, (5.15) 式可以从以下角度进行直观理解:

> 买入/卖出序列的长度随着限价订单的到达而增加, 随着市场订单和订单取消而减短, 长度变化与相应事件的过程有关. 当发生订单取消事件或市场订单到达时, 订单位置朝 0 的方向减小, 而新的限价订单的到达则不会改变现有的订单位置. 然而, 限价订单的到达可能会改变现有的订单位置减至 0 的速度 (订单取消事件的发生在每一个序列上是均匀分布的), 因此改变 $\dfrac{Z_n(t-)}{Q_n^b(t-)}$ 因子 (在订单取消时, 与 $-dC_n^3(t)$ 相关).

比例化的净订单流和买入−卖出序列的流动极限 (5.15) 式中的比例化过程可以根据到达时间间隔 T_i 和第 $j \in \{1, \ldots, 6\}$ 类订单序列长度的变化量 V_i^j 的大数定律得到, 论文中进一步假设

(A1) $\dfrac{T_1 + \cdots + T_n}{n} \xrightarrow{P} \lambda^{-1}$;

(A2) 对于某非随机向量 $\boldsymbol{V}_*$, $\dfrac{\boldsymbol{V}_1 + \cdots + \boldsymbol{V}_n}{n} \xrightarrow{P} \boldsymbol{V}_*$.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

实际上, 当 $\boldsymbol{V}_i$ 为稳定遍历过程时, 则有 $\boldsymbol{V}_* = E(\boldsymbol{V}_1)$. 另外, 根据 (A1) 和 (5.14) 可知, $\dfrac{M_{[nt]}}{n} \xrightarrow{P} \lambda t$, 结合 (A2) 和 (5.15) 式, 可知对于任意固定的 t, $\boldsymbol{C}_n(t) \xrightarrow{P} \lambda \boldsymbol{V}_* t$. 根据该结果可以进一步得到

函数弱大数定律 当 $n \to \infty$ 时,

$$\boldsymbol{C}_n(\cdot) \to \lambda \boldsymbol{V}_* \boldsymbol{e}(\cdot) \quad \text{在 } D^6[0,T] \text{ 中}, \tag{5.16}$$

其中 $\boldsymbol{e}(\cdot)$ 为指示函数.

对于 (5.16) 式更具体的背景和假设的介绍可参见 5.6 节补充材料 4. 该节同时对于排队论和流动性极限 (5.15) 的求解进行了介绍. 为简化表达, 论文中同时假设

(A3) 订单的取消事件在每个序列中均匀分布.

作者认为该假设可以进一步放宽至订单取消速度随着序列长度而增加的情况.①

(5.15) 的流动极限 假设存在常数 q^b, q^a 和 z, 使得 $(Q_n^b(0), Q_n^a(0), Z_n(0)) \Rightarrow (q^b, q^a, z)$. 定义

$$v^b = -V_*^1 + V_*^2 + V_*^3, \quad v^a = -V_*^4 + V_*^5 + V_*^6,$$
$$\alpha = \lambda V_*^2, \quad \beta = \frac{q^b}{\lambda V_*^3}, \quad c = -\frac{v^b}{V_*^3},$$

则当 $n \to \infty$ 时,

$$(Q_n^b, Q_n^a, Z_n) \Rightarrow (Q^b, Q^a, Z) \quad \text{在 } D^3[0,T] \text{ 中}, \tag{5.17}$$

其中 $Q^b(\cdot)$, $Q^a(\cdot)$ 和 $Z(\cdot)$ 定义为

$$\begin{aligned} &Q^b(t) = q^b - \lambda v^b \min(t,\tau), \quad Q^a(t) = q^a - \lambda v^a \min(t,\tau), \\ &\frac{dZ(t)}{dt} = -\lambda \left(V_*^2 + V_*^3 \frac{Z(t-)}{Q^b(t-)} \right) \quad \text{对于 } t < \tau,\ Z(0) = z. \end{aligned} \tag{5.18}$$

① 在实际中, 订单取消的原因很多. 研究表明, 订单的取消是市场流动性的主要来源 (如 Cont (2011)). 实际上, 对于订单取消的特性进行实证研究对于监管和交易都有重要作用. CNBC 新闻中报道了一个极端的情形: "有一个神秘的计算机程序每天提交订单, 并随后取消该订单. 该订单占据了美国证券市场 (2012 年 10 月 5 日) 当个星期总交易量的 4%. ····· 该程序可以在 25 毫秒中对于 500 只股票进行下单, 但该算法从来不会实际执行任何一次交易."

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

此处

$$\tau = \min\left\{\frac{q^a}{\lambda v^a}, \frac{q^b}{\lambda v^b}, \tau^z\right\},$$

$$\tau^z = \begin{cases} \left(\dfrac{(1+c)z}{\alpha} + \beta\right)^{c/(c+1)} \beta^{1/(c+1)} c^{-1} - \beta/c, & c \notin \{-1, 0\}, \\ \beta(1 - e^{-\frac{z}{\alpha\beta}}), & c = -1, \\ \beta \log\left(\dfrac{z}{\alpha\beta} + 1\right), & c = 0. \end{cases}$$

5.5.3 基于 LOB 的序列反应模型

Cont 和 de Larrard (2013) 中的排队模型只使用了价格水平为 1 的 LOB 数据, 且对于限价订单/市场订单/取消订单设定了强先验假设 (e). 本节介绍一个 Huang 等 (2015) 提出的更为一般的数据驱动模型, 该模型使用了一个更为灵活的排队系统的定义, 即定义了连续时间马尔可夫链. 附录 B 对于马尔可夫链进行了简要介绍. 令 $\boldsymbol{q}_t = (q_t^{(-K)}, \ldots, q_t^{(-1)}, q_t^{(1)}, \ldots, q_t^{(K)}) \in \mathbb{N}^{2K}$, 其中 $q_t^{(-j)}$ 和 $q_t^{(j)}$, $j = 1, \ldots, K$ 分别为第 j 个最优买入价和最优卖出价对应的序列长度. 假设 $\boldsymbol{q}_t$ 为连续时间马尔可夫跳跃过程, 其生成矩阵 $\boldsymbol{Q} = (Q_{\boldsymbol{x},\boldsymbol{y}})_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{N}^{2K}}$ 定义为:

$$\boldsymbol{Q}_{\boldsymbol{x},\boldsymbol{y}} = \begin{cases} f_i(\boldsymbol{x}), & \text{当 } \boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e}_i, \\ g_i(\boldsymbol{x}), & \text{当 } \boldsymbol{y} = \boldsymbol{x} - \boldsymbol{e}_i, \\ -f_i(\boldsymbol{x}) - g_i(\boldsymbol{x}), & \text{当 } \boldsymbol{y} = \boldsymbol{x}, \\ 0, & \text{其他}, \end{cases} \tag{5.19}$$

其中 $\boldsymbol{e}_i \in \mathbb{N}^{2K}$ 为单位向量, 其第 i 个元素的值为 1 ($i = \pm 1, \ldots, \pm K$), 其他项的值为 0. f_i 和 g_i 满足如下条件:

(i) 存在 $\delta > 0$, 使得对于任意 $i = \pm 1, \ldots, \pm K$ 和任意满足各元素大于某个下界 l 的向量 $\boldsymbol{x}$, 有 $f_i(\boldsymbol{x}) - g_i(\boldsymbol{x}) < -\delta$.

(ii) 存在 $H > 0$, 使得 $\sum_{i=-K}^{K} f_i \leqslant H$.

(i) 和 (ii) 保证内含的马尔可夫链存在稳定分布, 且为遍历过程, 因此大数定律在此情况下仍然适用, 具体可参见 Meyn 和 Tweedie (1993) 与 Lai 和 Xing (2008) 的附录 B. 特别地, 假设 (ii) 使得订单到达速度是有界的, 排队系统不会过于发散. 通过对函数 $f_i(\boldsymbol{q})$ 和 $g_i(\boldsymbol{q})$ 进一步定义, 可以建立一个反映序列长度

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

之间的相关性的模型, 此处序列为价格–时间优先准则下不同的价格水平的稳定状态 (即在稳定分布下的序列).

基于个体序列密度的进一步说明 目前的研究对于包含 $2K$ 个序列的简单 (但是非实际) 情形进行了分析, 并得到了重要的发现. 这 $2K$ 个序列包括限价订单 (L) 的到达、市场订单 (M) 的到达和取消订单 (C), 且这些事件之间相互独立, 因而可得,

$$f_i(\boldsymbol{q}) = \lambda_i^L(q^{(i)}) \quad \text{且 } g_i(\boldsymbol{q}) = \lambda_i^M(q^{(i)}) + \lambda_i^C(q^{(i)}), \tag{5.20}$$

其中 $\lambda_i^L(q^{(i)})$, $\lambda_i^M(q^{(i)})$ 和 $\lambda_i^C(q^{(i)})$ 分别是第 i 个序列中这些事件的概率密度, 且参数满足对称性假设 $\lambda_i^L = \lambda_{-i}^L$, $\lambda_i^M = \lambda_{-i}^M$ 以及 $\lambda_i^C = \lambda_{-i}^C$.

为了对于由价格–时间优先准则得到的 $2K$ 个序列之间的依赖关系进行说明, Huang 等 (2015) 对于 (5.20) 式中的 $g_i(\boldsymbol{q})$ 项进行了如下的简化改进:

$$g_i(\boldsymbol{q}) = \begin{cases} \lambda_i^C(q^{(i)}) + \lambda_{\text{buy}}^M(\boldsymbol{q}) I_{\text{bestask}(\boldsymbol{q})=i}, & \text{当 } i \geqslant 1, \\ \lambda_i^C(q^{(i)}) + \lambda_{\text{sell}}^M(\boldsymbol{q}) I_{\text{bestbid}(\boldsymbol{q})=i}, & \text{当 } i \leqslant -1, \end{cases} \tag{5.21}$$

其中 $\text{bestask}(\boldsymbol{q}) = \min\{i > 0 : q^{(i)} > 0\}$, $\text{bestbid}(\boldsymbol{q})$ 类似可以定义, $\lambda_{\text{buy}}^M(\boldsymbol{q})$ 和 $\lambda_{\text{sell}}^M(\boldsymbol{q})$ 分别是市场买入订单和市场卖出订单的稳定状态密度. 论文指出, 该模型是拟生灭过程 (QBD) 的一个特例. QBD 过程是生灭过程的一个推广, 与生灭过程相同, 状态的下一次变动只能到达最近的一个点, 但发生一次变动的时间间隔服从的分布则比经典的生灭过程要复杂得多, 具体可参见 Motyer 和 Taylor (2006) 与 Latouche (2010). QBD 过程的稳定分布可以利用矩阵几何方法求得. Huang 等 (2015) 论文中 2.4.2 节、2.4.3 节和附录 A.0.3 都利用了该方法, 论文基于遍历假设, 利用大数定律得到相应的样本估计值, 来取代 QBD 中的参数, 并利用法国电信股票的数据对此过程进行了阐述.

包含均值回归参考价格的序列–反应模型 Huang 等 (2015) 论文的 2.4.4 节和 2.4.5 节提出了一种新方法, 对于 (5.20) 式中 $i = \pm 1$ 处进行修改, 从而对于序列之间的相关性进行建模. 论文根据 (5.20) 及其变形的实证研究结果得到了一种新方法, 称为 "序列–反应模型", 其包含一个均值回归参考价格 p_t^{ref}, 从而使得 p_t^{ref} 不变化的时间区间 (5.20) 式依然成立, 但当 p_t^{ref} 变化时, (5.20) 也相应改变. 下面的任意一个事件都会改变中间价格, 并触发 (5.20) 的变化:

- 在 $q_t^{(1)} = 0$ 时, 在买入–卖出差价中加入一个限价买入订单, 或者在 $q_t^{(-1)} = 0$ 时, 在买入–卖出差价中加入一个限价卖出订单;

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

- 在其中一个最优序列中, 取消最近的限价订单;
- 在其中一个最优序列中, 市场订单耗去最近的限价订单;

由于该模型被用于计算 "很多有用的短期预测", 包括 "消极订单的执行概率、价格升高的概率 ……" 以及对于 "复杂交易" 的模拟表现等, Huang 等 (2015) 对于 p_t^{ref} 和前面提到的触发变化的事件, 给出了相关的模拟结果. 具体而言, 定义 δ 为最小交易单元, 假设 p_t^{ref} 变化概率为 θ. 在 t 时刻, 当 p_t^{ref} 变化时, $q_t^{(1)} = 0$ 或 $q_t^{(-1)} = 0$ (即对应于 5.5.1 节中的序列衰减的情况) 由 $p_t^{\text{ref}} - p_{t-}^{\text{ref}}$ 的值是 δ 还是 $-\delta$ 所决定. 论文中指出:

> 当 $q_t^{(i)}$ 的值变为其中一个近邻的值时 (当 p_t^{ref} 增大时, 则变为右侧近邻值, 减小时则变为左侧近邻值) …… 我们对于 LOB 中直到第三个限价 ($i = \pm 3$) 的值进行记录 …… 在分析序列交换过程时, 需要非常细致: 不同序列的平均事件规模并不相同. 因此, 当 $q_{t-}^{(i)}$ 变化为 $q_t^{(j)}$ 时, 新的值需要利用 LOB 中价格水平 i 和 j 中平均事件规模的比率进行标准化.

论文指出, 序列–反应模型的基本思想在于:

> 在封闭的物理系统中, 随机性的唯一来源是不同参与者之间的内生反应, 而市场的变化过程并不是这样. 市场的变化也受到外部信息的影响, 如新闻事件可能会增加价格的波动率. 因此, 在序列–反应模型中引入外生变量显得非常有必要.

在序列–反应模型中, 为引入外生变量 ("外生信息"), 论文假设当 p_t^{ref} 变化时, LOB 的状态围绕新的参考价格, 以概率 $\tilde{\theta}$ 发生变化, 变化大小服从其固有概率分布. Huang 等 (2015) 的 3.1 节利用法国电信股票的数据, 基于中间价收益率的 10 分钟标准差对于 θ 和 $\tilde{\theta}$ 进行了标定. 在标定时, 使用了均值回归比率 $\eta = \dfrac{N_c}{(2N_a)}$, 其中 N_c (或 N_a) 是在研究的区间内 p_t^{ref} 在同一方向 (或不同方向) 连续变动的次数. 在序列–反应模型中加入 p_t^{ref} 项意味着马尔可夫链推广到 $(\boldsymbol{q}_t, p_t^{\text{ref}})$. 考虑到 p_t^{ref} 在序列–反应模型中的重要作用, Huang 等 (2015) 中的 2.2.2 节对于 p_t^{ref} 进行了明确定义: 当买入–卖出价差为奇数时, p_t^{ref} 定义为中间价, 而当价差为偶数时, p_t^{ref} 定义为中间价 $\pm$ 最小交易单位的一半, 其中 $\pm$ 的符号的选取是取决于哪一项可以使参考价接近 p_{t-}^{ref}. Huang 等 (2015) 也给出了参考价格的其他定义, 但这些定义相对来说都要复杂得多.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

5.6 补充材料和习题

1. 图 5.9 给出了基于一元 Hawkes 过程的强度过程 λ_t 的一个模拟样本路径, Hawkes 过程的强度函数为

$$\lambda_t = \mu + \sum_{\tau_i < t} \alpha e^{-\beta(t-\tau_i)},$$

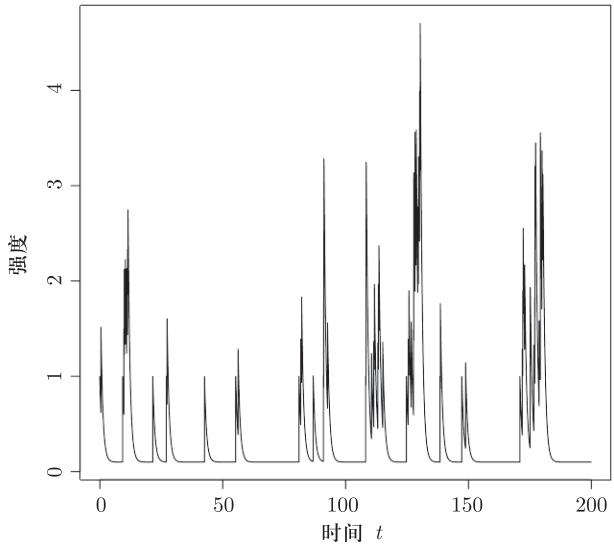


图 5.9 一元 Hawkes 过程的模拟强度

其中 $\mu = 0.3$, $\alpha = 0.6$, $\beta = 1.2$. 图 5.10 给出了基于二元 Hawkes 过程的模拟样本路径, Hawkes 过程的强度函数为

$$\lambda_t^{(1)} = \mu^{(1)} + \sum_{\tau_i < t} \alpha_{11} e^{-\beta_{11}(t-\tau_i)} + \sum_{\tau j < t} \alpha_{12} e^{-\beta_{12}(t-\tau_j)},$$
$$\lambda_t^{(2)} = \mu^{(2)} + \sum_{\tau_j < t} \alpha_{22} e^{-\beta_{22}(t-\tau_j)} + \sum_{\tau i < t} \alpha_{21} e^{-\beta_{21}(t-\tau_i)},$$

其中 $\mu^{(1)} = \mu^{(2)} = 0.5$, $\alpha_{11} = \alpha_{22} = 0.8$, $\alpha_{12} = \alpha_{21} = 0.5$, $\beta_{11} = \beta_{12} = 1.5$, $\beta_{22} = \beta_{21} = 1.5$. 分析两个过程之间的互相激励和在每次跳跃之后的指数衰减, 并阅读附录 C 中的 C.4 节, 研究 Hawkes 过程模拟方法.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

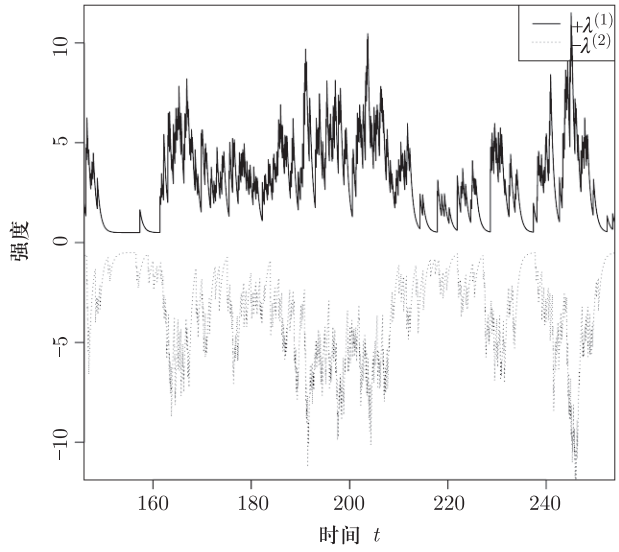


图 5.10 二元 Hawkes 过程的模拟强度, 其中上图中用 $\lambda^{(1)}$、下图中对 $\lambda^{(2)}$ 进行了反置以便查看

2. 使用 NASDAQ ITCH 数据 (可从本书网站下载) 来标定买方侧交易强度的一元点过程, 并对于拟合模型的特性进行评述.

3. **5.5 节提到的 Kruk 模型背后的基本思想** 考虑一个拍卖过程, 记为 n, 包含 N 个价格水平 $P_1^{(n)}, \ldots, P_N^{(n)}$. 定义 $A_i^{(n)}(t)$ 和 $B_i^{(n)}(t)$ 为时刻 t 时, 价格水平 P_i 上的剩余买入和卖出订单数. 显然, 当相同价格水平下的买入和卖出订单互相抵消交易时, $A_i^{(n)}(t)B_i^{(n)}(t) = 0$. Kruk (2003) 假设:

   - 买方的到达服从更新过程, 到达时间间隔为 $u_k^{(n)}$, $k = 1, 2, \ldots$, 其均值为 $\dfrac{1}{\lambda^{(n)}}$, 标准差为 $\alpha^{(n)}$; 卖方的到达也服从更新过程, 到达时间间隔为 $v_k^{(n)}$, $k = 1, 2, \ldots$, 其均值为 $\dfrac{1}{\mu^{(n)}}$, 标准差为 $\beta^{(n)}$. 两个到达过程之间相互独立.
   - 第 k 个买方 (或卖方) 愿意购买 (或卖出) $l_k^{(n)}$ (或 $m_k^{(n)}$) 只股票, 其中 $l_k^{(n)}$ (或 $m_k^{(n)}$), $k = 1, 2, \ldots$ 取值为正, 独立同分布, 其均值为 $\bar{l}^{(n)}$ (或 $\bar{m}^{(n)}$), 标准差为 $\kappa^{(n)}$ (或 $\nu^{(n)}$).
   - 第 k 个买方 (或卖方) 愿意在价格为 $b_k^{(n)}$ (或 $a_k^{(n)}$) 进行交易, 其中 $b_k^{(n)}$ (或 $a_k^{(n)}$), $k = 1, 2, \ldots$ 独立同分布, 满足

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$$P(b_k^{(n)} = i) = P(a_k^{(n)} = i) = p_i, \quad i = 1, 2, \ldots, N.$$

Kruk 的结果引出了 5.1.1 节和 5.1.2 节

- 在 $N = 2$ 时, 选取比例为 $\tilde{B}_k^{(n)}(t) = n^{-1/2} B_k^{(n)}(nt)$ 和 $\tilde{A}_k^{(n)}(t) = n^{-1/2} A_k^{(n)}(nt)$, 则有

$$\left(\tilde{A}_1^{(n)}, \tilde{B}_1^{(n)}, \tilde{B}_2^{(n)}, \tilde{A}_2^{(n)}\right) \Longrightarrow (0, 0, \boldsymbol{Z}),$$

其中 $\boldsymbol{Z}$ 为反射布朗运动半鞅过程.

- 在 $N \geqslant 3$ 时, 选取比例为 $\tilde{B}_k^{(n)}(t) = n^{-1} B_k^{(n)}(nt)$ 和 $\tilde{A}_k^{(n)}(t) = n^{-1} A_k^{(n)}(nt)$, 则有

$$\left(\tilde{A}^{(n)}, \tilde{B}^{(n)}\right) \Longrightarrow (\bar{a}\boldsymbol{e}, \bar{b}\boldsymbol{e}),$$

其中 $\bar{a}$, $\bar{b}$ 是由模型参数确定的两个常数, $\boldsymbol{e}$ 为恒等映射 $\boldsymbol{e}(t) = t$.

4. **序列的扩散和流动极限** 在此处, 我们给出了 5.5.1 节和 5.5.2 节中的结果的背景介绍和额外补充. Cont 和 de Larrard 的结果 (5.10) 是在 $EX_i = 0$ 情况下的扩散极限, 且是函数中心极限定理的结果. Blanchet 和 Chen (2013) 也给出了单笔交易价格过程的跳跃极限和价差过程的跳跃扩散极限. 5.5.2 节中提到的 Guo 等的结果为流动极限, 且由函数强大数定律所得. 扩散极限来自于高负荷的情形, 且与中心极限定理有关. 排队论中的流动性模型来源于对大型容器的液体流动问题的研究, 容器同时连接着注入液体和排出液体的管道. 该问题中的随机扰动项 (即随机输入和输出项) 根据大数定律来达到平衡, 最终得到一个确定性的极限分布, 即为 (5.16). Whitt (2002) 对于流动和扩散极限进行了综述. 随机系统 (5.17) 的弱极限的求解过程更复杂. 除了流动性极限外, Guo 等 (2015) 发现订单位置的波动服从均值回归的高斯过程. 均值回归从本质上来说就是订单位置的流动极限与根据订单簿净流动修正的序列长度有关, 其中净流动定义为限价订单减去市场订单和取消订单. 论文表明, 均值回归的速度与订单到达强度和取消率成比例.

Horst 和 Paulsen (2015) 对于 5.5.1 节中的简化模型进行了研究, 并求出了同时包含价格和交易量的 LOB 数据的流动极限. Horst 和 Kreher (2015) 对该研究进行了拓展, 包含了与 LOB 的状态相关的订单动态特征.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

5. **对于 LOB 的实证研究** Ahn 等 (2001) 利用 1996 年 7 月至 1997 年 6 月香港交易所 (HKSE) 的恒生指数 (HSI) 中包含的 33 只股票的交易记录与买入和卖出记录, 对于 LOB 进行了实证研究. 交易记录包含所有的价格和交易量数据, 且都以秒为单位进行记录, 买入和卖出记录则包含每 30 秒区间内最优的五个买入和卖出价格水平下剩余的限价订单的规模. 这 33 只股票占据了研究区间内整个市场资本化的 70%. 在该区间中, 每天的交易时间为 10:00 至 12:30 和 14:30 至 15:55. Ahn 等 (2001) 将上下午的交易时间进行合并, 并划分为 16 个 15 分钟的交易区间. 论文定义 (短期) 波动率为逐交易收益率的平方, 定义 LOB 快照的深度为最优的五个买入和卖出价格水平下剩余的限价订单的规模, 并将这两个变量对于所有的股票进行了平均. Ahn 等 (2001) 的图 1 给出了每 15 分钟区间内的波动率和深度. 图中波动率项给出了波动率的常见白天特征, 而平均深度的图像则为 U 形. Lee 等 (1993) 利用 NYSE 的数据、Biais 等 (1995) 利用巴黎证券交易所的数据也都给出了类似的深度的特征. Ahn 等 (2001) 基于 LOB 数据对于下述两个假设进行了检验, 这也是实证研究的主要目标.

假设 1: 短期价格波动的增大 (或减小) 通常跟随着限价订单相对于市场订单的下单量的减少, 因此市场深度通常会随之增大 (或减小).

假设 2: 市场深度的增大 (或减小) 通常跟随着短期价格波动率的减小 (或增大).

令 $R_{i,t}$ 为时间区间 t 中第 i 次交易的收益率, $i = 1, \ldots, N_t$, 其中 N_t 为该区间中总交易次数, 论文中考虑了如下的变量:

- $\mathrm{RISK}_t = \sum_{i=1}^{N_t} R_{i,t}^2$ 为时间区间 t 的短期波动率.
- $\mathrm{RISK}_t^+ = \sum_{i=1}^{N_t} R_{i,t}^2 I_{\{R_{it}>0\}}$ 为上行波动率.
- $\mathrm{RISK}_t^- = \sum_{i=1}^{N_t} R_{i,t}^2\ I_{\{R_{it}<0\}}$ 为下行波动率.
- $\mathrm{DEPTH}_t^{\mathrm{bid}}$ = 时间区间 t 结束时, 最优的五个买入价格水平下的总剩余股票数量, $\mathrm{DEPTH}_t^{\mathrm{ask}}$ 可类似定义.
- DEPTH_t =$\mathrm{DEPTH}_t^{\mathrm{bid}}$+ $\mathrm{DEPTH}_t^{\mathrm{ask}}$.
- $\mathrm{DIFF}_t^{\mathrm{buy}}$ = 在时间区间 t 中新下单的限价买入订单和市场买入订单数之差; $\mathrm{DIFF}_t^{\mathrm{sell}}$ 可类似定义.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

- NBUY_t = 在时间区间 t 中市场买入订单的总数量; NSELL_t 可类似定义.
- $\mathrm{NTRADE}_t = N_t$.
- $t = k$ 时, $\mathrm{TIME}_{k,t} = 1$; 其他情况下, $\mathrm{TIME}_{k,t} = 0$.

为检验假设 1, 作者首先进行了截面数据回归:

$$\begin{aligned}\mathrm{DEPTH}_t = {} & \alpha_1 + \beta_1 \mathrm{RISK}_{t-1} + \theta_1 \mathrm{NTRADE}_t \\ & + \sum_{k=1}^{14} \gamma_{k,1} \mathrm{TIME}_{k,t} + \rho_1 \mathrm{DEPTH}_{t-1} + \varepsilon_{t,1}, \end{aligned} \tag{5.22}$$

其中 DEPTH_{t-1} 和二值变量 $\mathrm{TIME}_{k,t}$ 作为日间方差和市场深度的自相关项的控制变量. 在拟合该模型时, 使用的方法为广义矩估计 (GMM, 参见 Lai 和 Xing (2008), 9.7 节), GMM 方法的 t 统计量对于异方差、自相关和 ε_t 与回归变量相关等情形都是稳健的. 作者也分析了 DEPTH_t 的其他两个定义: 只包含最优价格的情形、只包含最优五个价格中次优价格的情形. 在三种定义下, DEPTH_t 都与 NTRADE_t 显著负相关, 该结果也与 Lee 等 (1993) 的结果一致, 该结果表明交易会减少市场的流动性. 另外, Ahn 等 (2001) 中的表 II 表明, β_1 显著为正, 这也与假设 1 一致.

为了分析短期波动性对订单流的影响, Ahn 等 (2001) 使用 GMM 方法对于另一个截面模型进行了拟合:

$$\begin{aligned}\Delta\mathrm{DEPTH}_t = {} & \alpha_2 + \beta_2 \mathrm{RISK}_{t-1} + \rho_2 \Delta\mathrm{DEPTH}_{t-1} \\ & + \sum_{k=1}^{13} \gamma_{k,2} \mathrm{TIME}_{k,t} + \varepsilon_{t,2}, \end{aligned} \tag{5.23}$$

其中 $\Delta\mathrm{DEPTH}_t = \mathrm{DEPTH}_t - \mathrm{DEPTH}_{t-1}$ 为时间区间 t 内, 新下单的限价订单和市场订单规模之和与取消订单的规模之差, 是订单流组成的量度. 由论文中表 III 可知, ρ_2 显著为负, 因此订单流是自我调整的, 即当订单簿很厚时, 交易量更多, 而当订单簿很薄时, 限价订单会增加. 论文中同时指出, 短期波动率对订单流的影响并不强. 事实上, 在计算 DEPTH 时, 只利用最优价格下的序列长度时, 只有 7 只股票的 β_2 显著为正. 为进一步讨论订单组成与买方侧和卖方侧的流动性驱动下的价格波动性的关系, 作者考虑了第三个截面模型如下:

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$$\begin{aligned}\mathrm{DIFF}_t^{\mathrm{buy}} &= \alpha_3 + \beta_3^+ \mathrm{RISK}_{t-1}^+ + \beta_3^- \mathrm{RISK}_{t-1}^- \\ &\quad + \rho_3 \mathrm{DIFF}_{t-1}^{\mathrm{buy}} + \sum_{k=1}^{13} \gamma_{k,3} \mathrm{TIME}_{k,t} + \varepsilon_{t,3}, \qquad (5.24a)\\ \mathrm{DIFF}_t^{\mathrm{sell}} &= \alpha_4 + \beta_4^+ \mathrm{RISK}_{t-1}^+ + \beta_4^- \mathrm{RISK}_{t-1}^- \\ &\quad + \rho_4 \mathrm{DIFF}_{t-1}^{\mathrm{sell}} + \sum_{k=1}^{13} \gamma_{k,4} \mathrm{TIME}_{k,t} + \varepsilon_{t,4}. \qquad (5.24b)\end{aligned}$$

论文中的表 IV 表明, ρ_3 和 ρ_4 都显著为正. 另外, β_3^- 和 β_4^+ 分别对于 17 只股票和 20 只股票为正. 在使用第三种市场深度的定义 (即最优五个价格中的次优价格) 时, 上述数字分别会增加到 30 和 31. 这些结果与流动性提供者更倾向于在收益更高的一侧下限价订单以增加流动性是一致的.

Ahn 等 (2001) 接下来对于市场深度对于波动性的后续影响进行分析, 研究时使用 GMM 来对下面的截面模型进行拟合:

$$\begin{aligned}\mathrm{Risk}_t &= \alpha_5 + \beta_5 \mathrm{DEPTH}_{t-1} + \theta_5 \mathrm{NTRADE}_t \\ &\quad + \sum_{k=1}^{14} \gamma_{k,5} \mathrm{TIME}_{k,t} + \rho_5 \mathrm{RISK}_{t-1} + \varepsilon_{t,5}. \qquad (5.25)\end{aligned}$$

为进一步分析上述影响的方向, 作者将市场深度分为买方侧和卖方侧的深度, 并分别与下行波动率和上行波动率的关系进行分析, 模型如下:

$$\begin{aligned}\mathrm{Risk}_t^+ &= \alpha_6 + \beta_6^- \mathrm{DEPTH}_{t-1}^{\mathrm{bid}} + \beta_6^+ \mathrm{DEPTH}_{t-1}^{\mathrm{ask}} + \theta_6 \mathrm{NBUY}_t \\ &\quad + \sum_{k=1}^{14} \gamma_{k,6} \mathrm{TIME}_{k,t} + \rho_6 \mathrm{RISK}_{t-1} + \varepsilon_{t,6}, \\ \mathrm{Risk}_t^- &= \alpha_7 + \beta_7^- \mathrm{DEPTH}_{t-1}^{\mathrm{bid}} + \beta_7^+ \mathrm{DEPTH}_{t-1}^{\mathrm{ask}} + \theta_7 \mathrm{NSELL}_t \\ &\quad + \sum_{k=1}^{14} \gamma_{k,7} \mathrm{TIME}_{k,t} + \rho_7 \mathrm{RISK}_{t-1} + \varepsilon_{t,7}.\end{aligned}$$

论文中的表 VI 表明, 上行波动率和下行波动率都与最优序列中的市场深度显著相关. 具体而言, 上行波动率只与买方侧市场深度显著负相关, 而下行波动率只与卖方侧市场深度显著负相关.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

6. Ahn 等 (2001) 关于 LOB 动态的一个重要发现在于, 论文对于限价订单流入和高密度市场订单的交替循环出现提供了实证依据, 即当市场上限价订单很少时, 市场暂时的订单不平稳会触发短期波动率的上升, 吸引公开投资者更多地投资限价订单而不是市场订单. 这样的限价订单流入过程会一直持续, 直到短期波动率下降, 进而反过来增加市场订单的密度. 由于上述过程是基于香港证券交易所 (HKSE) 的数据进行分析, 而 HKSE 并没有执行做市机制, 因此 Ahn 等 (2001) 总结认为一个不包含做市商的订单驱动的交易机制是可行和可持续的.

7. **习题: 数据分析** 文件 `hsbc_L1.txt` 中包含 2014 年 3 月 17 日至 3 月 19 日香港证券交易所交易的 HSBC 股票的价格水平为 1 的数据. 利用该文件的数据对于 (5.22) 和 (5.25) 模型进行拟合. 在拟合时, 类似于 Ahn 等 (2001), 先将数据划分为以 15 分钟区间为单位的数据. 需要注意的是, 2014 年 HKSE 的交易时间为 9:30 至 12:00 和 13:00 至 16:00. 然后通过对于时间区间为 30 分钟和 5 分钟情况进行分析, 检验统计结果是否对于时间区间的长度敏感.

8. **习题: 弱收敛** 定义 $\{X_i\}_{i\geqslant 1}$ 为各项独立同分布的序列, 各项的取值范围为 $\{-1,1\}$, 因此可得

$$P(X_1=1)=P(X_2=1)=\frac{1}{2},$$
$$P(X_{i+1}=1|X_i=1)=P(X_{i+1}=-1|X_i=-1)=3/4, \text{ 当 } i>1.$$

   (a) 定义 $S_n(t)\equiv n^{-1/2}\sum_{i=1}^{\lfloor nt\rfloor}X_i$. ϕ-混合稳定序列的不变原理证明, S_n 弱收敛于比例化的布朗运动 $\sqrt{3}B(t)$, $t\geqslant 0$.

   (b) 定义 SDE 序列 $dY_n(t)=Y_n(t)dS_n(t)$, $Y_n(0)=1$. 证明当 $n\to 0$ 时, Y_n 弱收敛于极限过程 $\exp\left\{\sqrt{3}B(t)-\dfrac{t}{2}\right\}$, $t\geqslant 0$.

   附注: SDE 过程 $dY(t)=Y(t)d(\sqrt{3}B(t))$, $Y_n(0)=1$ 的解为 $Y(t)=\exp\left\{\sqrt{3}B(t)-\dfrac{t}{2}\right\}$.

9. **LOB 的多主体模型: 博弈论、经济物理和计算学习理论** 在 1.7 节的补充材料 1–3 中, 我们简要介绍了包含多个买方和卖方的双向拍卖市场中证券交易基础结构背后的经济学原理, 主要包括: 市场参与者获得的信息

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

的异质性、市场机制设计、逆向选择、博弈论和纳什均衡[①]. 此处, 我们侧重于 5.3 节和 5.5 节中讨论的 LOB 模型, 因此接下来主要介绍其他相关的理论. 首先, 5.3 节和 5.5 节中的模型主要是金融数学、运筹学和统计学相关的研究者提出的. 然而, 也有很多其他学科的研究者对于模型的发展做出了贡献. Gould 等 (2013) 的第 3 节讨论了以下几个方面的发展: (a) 经济学研究: 侧重介绍了完全理性的投资者的行为, 理性投资者在信息驱动的市场中, 会选择能够使得期望收益最大化的交易; (b) 经济物理学研究: "非智能" 方法假设投资者 (更精确地说, 投资者的计算机) 盲目按照一系列的准则进行交易, 这些准则认为订单流 "服从特定的随机过程". 论文的 3.10 节对于 LOB 数据的经济物理学分析中的统计问题进行了讨论, 介绍了大量实证研究中提到的市场中的统计规律性. 然而, Gould 等 (2013, 第 1717–1718 页) 同时指出, 这些实证研究 "难以进行, 原因在于研究中通常会假设序列的独立性和平稳性, 从而保证结果 (统计上) 的一致性", LOB 数据很少能满足这些假设. 另外, "很多时候会利用次优估计量 ····· 次优估计量通常方差很大或者有很大的偏差", 作者同时指出 "不同的实证研究的结论往往是相互矛盾的", 原因在于 "不同的交易匹配算法的结果是不同的 ····· 且不同的研究者所使用的数据的质量也参差不齐".

这些实证研究的一个主要发现是长记忆过程的普遍性, 因而指出了对于自相关性的统计预测的困难, 参见 Gould 等 (2013) 的 3.10.2 节. 从概念上来说, 困难之处在于长记忆性对于市场影响研究和高频交易策略设计中常用的均值回归特性提出了挑战. 正如 Lamoureux 和 Lastrapes (1990), Hillebrand (2005) 以及 Lai 和 Xing (2013) 指出, 在对于包含两种时间尺度的时间序列数据进行拟合时, 如果并不加入参数的不确定性, 则一定会产生虚假长记忆性. 此处的两种时间尺度中, 一种尺度为用于收集数据的短尺度, 另一种为包含参数跳跃的长时间尺度. Gould 等 (2013, 第 6 节) 指出, 对于非平稳行为的分析是 LOB 研究中的一个长期未解决的问题. "基本上至今所有的 LOB 模型都集中在建立均衡", 但是现有的

① 此处的博弈论指包含 n 个参与者的非合作博弈, 参与者 i 的策略集为 S_i. 定义其他参与者 j 的行动为 $x_j \in S_j$ 时, 参与者 i 的收益为 $f_i(\boldsymbol{x}), \boldsymbol{x} = (x_1, \ldots, x_n)$, $\boldsymbol{x}$ 称为策略式表述. 定义 $\boldsymbol{x}_{-i}$ 为除 i 外其他 $n-1$ 个参与者的策略式表述. 当策略 $\boldsymbol{x}^*$ 满足对于任意 $1 \leqslant i \leqslant n$ 和任意 $x_1 \in S_1, \ldots, x_n \in S_n$, $f_i(x_i^*, \boldsymbol{x}_{-i}^*) \geqslant f_i(x_i, \boldsymbol{x}_{-i}^*)$ 时, 称 $\boldsymbol{x}^*$ 为纳什均衡. 因此, 在纳什均衡中, 任意参与者偏离现有策略都不能获得更高的收益. 1994 年诺贝尔经济学奖获得者纳什 (John Nash) 证明了纳什均衡在有限 S_i、且允许混合策略情况下的存在性.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

实证证据都表明, LOB 的订单流都会受到频繁的冲击, 导致订单流出现非平稳状态, 因而并不能达到平衡.

Gould 等 (2013) 的第 5 节对于经济学和物理学研究者对 LOB 建模的兴趣和发展进行了评论, 5.3 节侧重于对于基于代理的模型 (ABM) 进行了评论. 但是, 论文指出 "两个学科的研究者之间总体来说是相互独立的". 经济学研究者提出的 LOB 模型一般会将订单流看成是静态的, 且 "以交易者为中心, 在一定的市场情形下分析使用完全理性的框架建立最优交易策略". 相反地, 物理学研究者提出的模型则将订单流看成是动态的 (Farmer 等 (2005)). ABM 模型中 "包含大量的潜在异质代理", 因此对 ABM 模型的研究模型是两种极端, 物理学研究者使用零智能模型进行研究, 而经济学研究者使用完全理性模型进行研究.

哪种模型对于多代理系统的作用更大这个问题, 在经济学和经济物理学研究者中一直存在着争议. 一方面, Lo 和 Mueller (2010) 就物理学研究者对于经济学/金融学模型中的 "觊觎" 进行了反驳, 指出 "经济学和金融学研究者长期以来都努力建立与物理学模型同样具有预测性的经济学系统模型, 特别是金融市场模型. 这样的量化研究已经取得了一系列的重要突破, 然而物理学研究在一些情况下则产生了错误的数学准确性的直觉." 另一方面, 另一位与统计物理学家合作的经济学家 Hommes (2002) 则指出 "在过去的二十年中, 经济学中产生了重要的范式转变: 从关于理性代表性代理的理论分析模型到关于部分理性、异质代理的计算演化模型. 这样的转变至少与三个方面有关: (i) 从代表性代理到异质代理系统的转变; (ii) 从完全理性到部分理性的转变; (iii) 从理论分析模型到计算导向模型的转变."

计算导向的多代理模型在信息学和计算机学科中也有广泛应用, 如参见 Wooldridge (2009), Bloembergen 等 (2015) 以及 Helbing 和 Kirman (2013) 在经济学理论背景下的研究. 计算学习理论作为计算机学科的一个子领域, 旨在研究补充材料中提到的计算复杂性、机器学习和人工智能等问题.

10. **平均场近似和 LOB 模型中的纳什均衡** 平均场论最初起源于物理学中 Pierre Curie 的研究和 Pierre Weiss (1907) 的后续研究. 该理论提出了一个包含大量相互作用的个体的复杂系统, 从而将多体问题近似为一个单体问题, 系统中其他个体对于单个个体的影响都汇总为一个平均影响,

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

等价于将外部场中其他所有粒子间的作用都替换为一个任意的粒子, 参见 Kadanoff (2009). Smith 等 (2003, 3.6 和 3.7 节) 指出了平均场理论在 (多代理) 双向拍卖系统预测中的实用性.

Lions 和 Lasry (2007)、Lasry 和 Lions (2007) 以及 Huang 等 (2007) 分别独立地对于包含大量参考者的策略决策的平均场博弈问题进行了数学推导. 考虑一个包含 n 个参与者的差分博弈过程, 第 i 个参与者的状态过程为

$$dX_t^i = b(t, X_t^i, \alpha_t^i, \bar{\mu}_t)dt + \sigma dB_t^i, \tag{5.26}$$

其中 B^i $(i = 1, \ldots, n)$ 为独立的布朗运动, α_i^t 为可行策略, $\bar{\mu}_t = \sum_{i=1}^n \delta_{X_t^i}$ 是 $X_t^1, \ldots, X_t^n$ 的实证分布. 第 i 个参与者的目标是最大化

$$J_t^i = E\left[\int_0^T f(t, X_t^i, \alpha_t^i, \bar{\mu}_t)dt + g(X_T^i)\right]. \tag{5.27}$$

此处 b 和 f 是在 $[0,T] \times \mathbb{R} \times \mathcal{A} \times \mathbb{R}$ 上的可测实值函数. g 是一个 $\mathbb{R}$ 到 $\mathbb{R}$ 的可测函数. 注意到在博弈中, 各参与者序贯求解同样的控制问题. 由于博弈过程的对称性, 各参与者的最优策略应该是相同的, 反过来各参与者的最优控制过程和状态过程也应该服从同样的分布. 另外, 注意到各参与者与其他参与者之间只通过汇总项 $\bar{\mu}_t$ 进行相互作用. 因此, 当参与者数量很大时, 单个参与者对其他参与者的影响基本是可以忽略的. 在纳什均衡中, 代表性参与者会针对其他参与者的已知最优策略选择相应的最优策略. 当博弈中参与者的数量趋于无穷时, $\bar{\mu}_t$ 的实证分布收敛至 μ_t, 即代表性参与者最优状态的实际分布. 该模型称为平均场博弈 (MFG), 在该模型中, 各参与者求解如下的最优控制问题:

$$\begin{aligned} &\sup_{\alpha} E\left[\int_0^T f(t, X_t, \alpha_t, \mu_t)dt + g(X_T)\right] \\ &\text{满足 } dX_t = b(t, X_t, \alpha_t, \mu_t)dt + \sigma dB_t. \end{aligned} \tag{5.28}$$

此处 μ_t 为包含 n 个参与者的差分博弈中相互作用项的极限, 其刻画了代表性参与者在 MFG 中面对的外部环境, 因此 μ_t 对于最优控制问题 (5.28) 来说是外生的. 另一方面, 由纳什均衡的特性和包含 n 个参与者的差分博弈的对称性可知, μ_t 服从当 (5.28) 式中的 $\bar{\mu}_t$ 替代为 μ_t 时的最优状态过程 X_μ^t. 因此, μ_t 是一个固定点, 可以通过如下的迭代方法求得: 首先给定 μ_t 的分布, 然后代入到 (5.28) 中, 并求解最优控制问题, 得到最优状态过程 X_t^μ 和最优控制 α_t^μ. 接下来, 迭代更新 μ_t 直到满足

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$\mu_t = \mathcal{L}_{X_t^\mu}$. 在 MFG 中, $\mu_t \in \mathbb{R}$, 该问题也可以拓展到高维 $\mathbb{R}^d$ 情形, 具体可参见 Carmona 和 Delarue (2013).

Lions 和 Lasry (2007) 与 Lasry 和 Lions (2007) 利用动态规划准则, 求得了给定 μ_t 时, 对应于 (5.28) 中的 α 的价格函数的 HJB 方程, 并利用统计力学中的 Fokker-Planck 方程 (在随机过程中, 称为 Kolmogorov 向前方程) 求出了 μ_t 的值. 因此, 上述论文求得了一个耦合 PDE 系统, 对于 $n \to \infty$ 情形下 n 个参与者的差分博弈中的纳什均衡进行了刻画. 论文作者与 Guéant, Buera 和 Moll 等合作完成了 Guéant 等 (2011) 和 Achdou 等 (2014) 等论文, 将 MFG 应用于双向拍卖市场和宏观经济学中. Carmona 和 Delarue (2013) 与 Carmona 和 Lacker (2015) 利用随机最大化准则提出了一种替代性方法, 求得固定 μ 的向前向后随机微分方程 (FBSDE) 系统, 其中 μ 定义为固定点 $\mu_t = \mathcal{L}_{X_t^\mu}$. 随机最大化准则最初由 Kushner (1972) 提出, 并由 Peng (1990) 等进一步发展为利用 FBSDE 而不是 HJB 来求解最优控制问题的一个重要的方法. Ren (2016) 对于该方法进一步研究, 并提出了一种有效的数值方法, 以求解基于代理的 LOB 建模的纳什均衡.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

第六章　最优执行与配置

总体而言, 算法交易基于两个不同的时间维度: 一个是以天或周为单位; 另一个是以秒为单位 (不高于 100 秒). Kirilenko 等 (2015) 最近的实证研究表明, 这两种时间维度的交易分别对应于两种不同类型的交易者: 前者对应于基本面交易员; 后者是高频交易员. 这两类交易者的行为在统计上相差很大, 且在 2010 年 5 月的股市崩盘中起到不同的作用. 这两种维度的交易同时也对应于交易订单的分类和配置的两个阶段. 在第一阶段中, 交易员将大份订单以天或周为单位优化拆分为小份订单 (第三章对于其中涉及的交易成本进行了一定的介绍). 在第二阶段中, 交易员将订单进一步拆分到以秒为单位的订单. 在算法交易文献中, 通常将第一阶段称为资产的*最优执行*, 而将第二阶段称为交易所中的*最优配置*或者不同交易所之间的*智能下单* (Smart Oder Routing, SOR).

在本章中, 我们对于同一个交易所中的最优执行和最优配置中的建模和分析方法进行讨论, 交易所之间的智能下单问题则会在下一章中进行讨论. 具体而言, 在 6.1 节中, 我们首先讨论单个资产的最优执行问题, 并介绍该领域的两篇开拓性文章 Bertsimas 和 Lo (1998) 与 Almgren 和 Chriss (2001). 我们同时介绍了两篇文章不断增加的影响及后续的研究. 6.1.3 节对于两篇文章中提到的最优策略中的确定性特性进行了介绍. 为了得到非确定性特性, 6.2 节分析了一个价格的乘数影响所引起的奇异随机控制问题, Guo 和 Zervos (2015) 对该问题进行了求解. 6.3 节考虑了基于限价订单簿数据来推断交易中资产的供求情况下的最优执行问题, 并介绍了该领域中两篇开拓性文章 Obizhaeva 和 Wang (2013) 与 Alfonsi 等 (2010). 6.4 节对于投资组合的最优执行问题进行

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

了回顾. 在 6.5 节中, 我们分析了与最优配置相关的一系列随机优化问题. 6.6 节给出了相关的补充材料和习题, 并介绍了高频交易中的执行算法中常用的 VWAP, TWAP 和 POV 等收益量度.

6.1 单资产最优执行问题

最优执行问题, 特别是在包含短时间内的大量买入或卖出交易情况时, 主要的建模前提是建模的策略应对于股票价格产生影响. 因此, 最优执行即为找到一个能够平衡市场风险和价格影响这两种对立的风险的策略. 正如第五章中指出, 一个大额的卖出订单会使得股票价格下降, 且降低潜在的收益, 而大量的小额订单的交易则需要大量的时间, 因而会增加市场风险. 本质上来说, 最优执行的价格影响模型与经典的对冲和定价模型的不同之处在于, 前者主要对于交易的速度和数量进行控制, 以最小化价格影响.

最优执行问题的研究最初起源于 Bertsimas 和 Lo (1998) 与 Almgren 和 Chriss (1999, 2001). 在模型中, 价格影响是可加的, 且最优策略是一个平均交易量类型的确定性交易策略. 随着研究更加深入, 价格影响主要被分成三类: 永久性、暂时性和短暂性. 永久性影响模型假设每一笔交易都影响未来的股票标的价格, 暂时性影响模型假设每一笔交易只会在交易发生时对于股票产生影响, 而短暂性影响模型则假设交易对于价格的影响是随时间衰减的. Hasbrouck (1991) 与 Potters 和 Bouchaud (2003) 对于交易对股票价格的影响进行了实证研究. 这些基于公开数据的研究发现交易对价格的影响是关于交易额的递增凹函数. 然而, 其他使用私有独特数据的研究, 如 Almgren 等 (2005) 与 Obizhaeva 和 Wang (2013) 并不能验证该结论.

Bertsimas 和 Lo 模型: 在线性价格影响假设下对期望成本的最小化 假设交易者在时刻 $1,\dots,T$ 进行交易, 总计买入 $\tilde{N}$ 股股票. 假设交易者在时刻 t $(1\leqslant t\leqslant T)$ 买入 N_t 股, 成交价为每股 P_t. 假设交易对价格的影响是线性的, 且假设在没有价格影响时, 股票价格服从如下随机游走模型:

$$P_t = P_{t-1} + \epsilon_t + \theta N_t, \tag{6.1}$$

其中 θ 为正常数, ϵ_t 独立同分布并且 $E[\epsilon_t|N_t, P_{t-1}] = 0$; θN_t 是价格影响函数. Bertsimas 和 Lo (1998) 研究了如下的期望成本最小化模型:

$$\min_{N_1,\dots,N_T} E\left[\sum_{t=1}^{T} P_t N_t\right] \quad \text{满足} \ \sum_{t=1}^{T} N_t = \tilde{N}. \tag{6.2}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

6.1.1 节给出了该问题的解, 即 $N_1^* = \cdots = N_t^* = \cdots = N_T^* = \dfrac{\tilde{N}}{T}$.

Almgren 和 Chriss 框架: 永久性和暂时性价格影响模型 论文研究了在固定时间 T 内, 卖出固定数量的股份的一个清算问题. 在论文框架下, 价格的影响分为两类: 永久性影响和暂时性影响. 具体而言, 假设常数 δ 为两次连续交易之间的间隔时长, 且 $K = \dfrac{T}{\delta}$ 为整数. 假设 $k = 0, \ldots, K$ 时的股票价格为 P_k. 股票价格服从如下随机过程:

$$P_k = P_{k-1} + \sigma\sqrt{\delta}\epsilon_k - \delta g\left(\frac{N_k}{\delta}\right), \tag{6.3}$$

其中 ϵ_k 独立同分布, 均值为 0, 方差为 $\sigma^2 > 0$, 且 g 是一个给定的函数, $g\left(\dfrac{N_k}{\delta}\right)$ 表示当卖出 N_k 股标的时, 交易对于股票价格的永久性影响. 另外, 卖出 N_k 股标的同时产生一个短暂性影响 $h\left(\dfrac{N_k}{\delta}\right)$, 且只对时刻 k 时的股票价格 $\tilde{P}_k$ 产生影响如下:

$$\tilde{P}_k = P_{k-1} - h\left(\frac{N_k}{\delta}\right). \tag{6.4}$$

给定任意的交易策略 $(N_1, N_2, \ldots, N_K)$, 在时间 T 内卖出 $\tilde{N} = N_1 + \cdots + N_K$ 股标的, 则总收益为

$$R(N_1, \ldots, N_K) = \sum_{k=1}^{K} \tilde{P}_k N_k. \tag{6.5}$$

是否存在短暂性价格影响下的总收益之差 $C(N_1, \ldots, N_K)$ 称为清算成本. 最优执行问题即为选取最佳的 $N_1, \ldots, N_K$ 以最小化清算成本与交易收益的方差和风险偏好系数 λ 之积, 即为

$$\min_{N_1,\ldots,N_K} \{E\left[C(N_1, \ldots, N_K)\right] + \lambda \mathrm{Var}\left[R(N_1, \ldots, N_K)\right]\}$$
$$\text{满足 } \sum_{k=1}^{K} N_k = \tilde{N},\ N_k \geqslant 0,\ 1 \leqslant k \leqslant K. \tag{6.6}$$

由于 δ 的值通常很小, 我们可以将 (6.6) 简化为一个连续时间模型, 6.1.2 节给出了该问题的解, 该解也是确定性的.

6.1.1 问题 (6.2) 的动态规划求解

令 Y_t 为尚未购买的标的数, 则有

$$Y_t = Y_{t-1} - N_{t-1}, \tag{6.7}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

且 $Y_1 = \tilde{N}$, $Y_{T+1} = 0$. 论文发现, 最优解 $(N_1^*, \ldots, N_T^*)$ 可以用类似于 3.4 节的补充材料 4 中的逆向推断法求得. 令给定时间 t 的信息集 (σ-域) $\mathcal{F}_t$ 下的期望函数为 E_t. 注意到, N_t 是 $\mathcal{F}_t$ 可测的. 令

$$V_t(P_{t-1}, Y_t) = \min_{N_t,\ldots,N_T} E_{t-1}\left[\sum_{k=t}^{T} P_k N_k\right], \tag{6.8}$$

其中 $N_t + \cdots + N_T = \tilde{N} - N_1^* - \cdots - N_{t-1}^*$, 另外, 注意到 $N_T^* = Y_T$, $V_T(P_{T-1}, Y_T) = Y_T(P_{T-1} + \theta Y_T)$. 对于 $t = T-1$,

$$\begin{aligned} V_{T-1}(P_{T-2}, Y_{T-1}) &= \min_{N_{T-1}} E_{T-2}[(P_{T-2} + \theta N_{T-1} + \epsilon_{T-1})N_{T-1} \\ &\quad + V_T(P_{T-2} + \theta N_{T-1} + \epsilon_{T-1}, Y_{T-1} - N_{T-1})] \\ &= \min_{N_{T-1}} [(P_{T-2} + \theta N_{T-1})N_{T-1} \\ &\quad + [P_{T-2} + \theta N_{T-1} + \theta(Y_{T-1} - N_{T-1})](Y_{T-1} - N_{T-1})] \\ &= \min_{N_{T-1}} (\theta N_{T-1}^2 - \theta Y_{T-1} N_{T-1} + \theta Y_{T-1}^2 + P_{T-2} Y_{T-1}). \end{aligned}$$

该问题即为对于 N_{T-1} 的一个二次方程的最小化问题, 因此

$$N_{T-1}^* = \frac{Y_{T-1}}{2} \quad \text{且} \quad V_{T-1}(P_{T-2}, Y_{T-1}) = \frac{3}{4}\theta Y_{T-1}^2 + P_{T-2} Y_{T-1}.$$

按照该思路继续分析, 则可得到 $t < T$ 时的 Bellman 方程如下:

$$V_t(P_{t-1}, Y_t) = \min_{N_t} E_{t-1}\left[P_t N_t + V_{t+1}(P_t, Y_t - N_t)\right], \tag{6.9}$$

其中 $Y_{t+1} = Y_t - N_t$. 则最优价值方程 $V_{T-k}(P_{T-k-1}, Y_{T-k})$ 可以递归求得如下:

$$N_{T-k}^* = \frac{Y_{T-k}}{k+1}, \tag{6.10}$$

$$V_{T-k}(P_{T-k-1}, Y_{T-k}) = Y_{T-k}\left(P_{T-k-1} + \frac{k+2}{2(k+1)}\theta Y_{T-k}\right). \tag{6.11}$$

特别地, 当 $k = T-1$ 时,

$$N_1^* = \frac{Y_1}{T} = \frac{\tilde{N}}{T}, \tag{6.12}$$

$$V_1(P_0, Y_1) = Y_1\left(P_0 + \frac{T+1}{2T}\theta Y_1\right) = P_0\tilde{N} + \frac{\theta\tilde{N}}{2}\left(1 + \frac{1}{T}\right). \tag{6.13}$$

因此 $N_1^* = \cdots = N_t^* = \cdots = N_T^* = \dfrac{Y_1}{T} = \dfrac{\tilde{N}}{T}$.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

6.1.2 连续时间模型和变分法

上述模型也有相对应的连续时间模型, 在求解连续时间模型时, 可以利用变分法 (CoV) 直接计算, 而不是利用 3.4 节补充材料 5 中提到的连续时间动态规划方法.

变分法和 Euler-Lagrange 方程 给定方程 $f = f(x, y, y')$, 其中 $y = y(x)$, $f' = dy/dx$, 令

$$I = \int_{x_1}^{x_2} f(x, y, y')dx, \tag{6.14}$$

则 CoV 的基本问题为找到函数 $y(x)$ 来最大化或最小化 I, 其边界条件为

$$y(x_1) = y_1, y(x_2) = y_2. \tag{6.15}$$

使得 f 能取得 (6.14) 式的极值的一个必要条件是 $y = y(x)$ 满足如下的 Euler-Lagrange 方程:

$$\frac{\partial f}{\partial y} - \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y'}\right) = 0. \tag{6.16}$$

在特殊情况下, 当 f 仅仅是 y 和 y' 的函数时, Euler-Lagrange 方程变成 Beltrami 形式

$$y'\frac{\partial f}{\partial y'} - f = C, \tag{6.17}$$

其中 C 是一个常数.

CoV 的一个有名的例子是寻找两个点 $A = (x_1, y_1)$ 和 $B = (x_2, y_2)$ 之间的最短路径. 假设 $x_1 \neq x_2$, 且假设 A 和 B 之间的路径可以表示为 $(x, y(x))$, 其中 $y_1 = y(x_1)$, $y_2 = y(x_2)$. 则两点之间路径的长度取决于 $y(x)$; 令该路径长度为 $I(y)$, 根据 Pythagoras 定理, 可知

$$I(y) = \int_{x_1}^{x_2} \sqrt{1 + (y')^2}dx,$$

因此 $f(x, y, y') = \sqrt{1 + (y')^2}$. 由于 $\dfrac{\partial f}{\partial y'} = \dfrac{y'}{\sqrt{1 + (y')^2}}$, 最短路径 $y(x)$ 满足 Euler-Lagrange 方程 $\dfrac{y''}{[1 + (y')^2]^{\frac{3}{2}}} = 0$. 因此, $y'' = 0$, 则最短路径是 y_1 和 y_2 之间的直线. 对于一般的 f, Euler-Lagrange 方程可能有多个解, 因此需要进一步的检验以确定极值点是最大值还是最小值. 另外, Euler-Lagrange 方程只

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

是在 y 两阶可微的条件下适用, 但是 (6.14) 式中的优化函数不一定二阶可微. Gelfand 和 Fomin (2000) 对于该情形下的变分法方法进行了详细的分析.

Almgren 和 Chriss 模型 在 (6.4) 式中, 令 $\delta \to 0$, 且 $h(x) = \eta x$, $g \equiv 0$, 则该问题可以转化为如下的连续时间模型. 首先, 在 (6.3) 式中, $\sigma\sqrt{\delta}\epsilon_k$ 变成增量 σdB_t, 其中 B_t 是标准布朗运动过程. 接下来, 股票执行价格 $\tilde{P}_t$ 为

$$\tilde{P}_t = P_t - \eta\dot{Y}_t = P_t + \eta U_t, \tag{6.18}$$

其中 Y_t 是时刻 t 时的股票数, $\dot{Y}_t = -U_t$ 是该时刻卖出股票的速度. 因此, 期望清算成本函数可以由下式得到:

$$C = E\left[\int_0^T \tilde{P}_t U_t dt\right] - E\left[\int_0^T P_t U_t dt\right] = \eta E\left[\int_0^T U_t^2 dt\right]. \tag{6.19}$$

交易成本的方差为

$$\mathrm{Var}\left[\int_0^T Y_t dP_t\right] = \sigma^2 E\left[\int_0^T Y_t^2 dt\right]. \tag{6.20}$$

结合 (6.19) 式和 (6.20) 式, 由 $U_t = \dot{Y}_t$ 可得到 (6.6) 式的连续时间形式如下:

$$\inf_{Y_t} E\left[\eta\int_0^T \dot{Y}_t^2 dt + \lambda\sigma^2\int_0^T Y_t^2 dt\right]. \tag{6.21}$$

在 CoV 框架下, 与 (6.21) 式相关的 Euler-Lagrange 方程为

$$\eta\frac{d^2Y_t}{dt^2} - \lambda\sigma^2 Y_t = 0. \tag{6.22}$$

上面的二阶 ODE 方程的解是 e^{At}, $A^2 = \dfrac{\lambda\sigma^2}{\eta}$ 的线性函数. 由边界条件 $Y_0 = \tilde{N}$ 和 $Y_T = 0$, 可得如下的唯一解:

$$Y_t^* = \tilde{N}\frac{\sinh(A(T-t))}{\sinh(AT)},$$

即可得到最优交易策略 Y_t^*, 该最优解是确定性解.

一般化模型 (6.21) 式的一个一般化模型拓展是

$$\inf_{Y_t} E\left[\eta\int_0^T G(Y_t, \dot{Y}_t)dt\right]. \tag{6.23}$$

特别地, 令 $G(y,u) = C_1(y,u) + \lambda C_2(y,u)$, 且

$$C_1(Y_t, U_t) = Y_t g(U_t) + U_t h(U_t), \quad C_2(Y_t, U_t) = \sigma^2 Y_t^2 + U_t^2 l^2(U_t),$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 g, h 和 l 是给定的函数并且 $U_t = -\dot{Y}_t$. $l = 0$ 的情形对应于 (6.3) 式和 (6.4) 式中永久性和暂时性价格影响方程. 相对应的 Euler-Lagrange 方程为

$$0 = f_y(y, -\dot{y}) + \frac{d}{dt} f_u(y, -\dot{y}) = f_y(y, -\dot{y}) + \dot{y} f_{yu}(y, -\dot{y}) - \ddot{y} f_{uu}(y, -\dot{y}), \quad (6.24)$$

其中 $f(y, u) = yg(u) + uh(u) + \lambda\sigma^2 y^2 + \lambda u^2 l^2(u)$. 在一定的边界条件 Y_0 和 Y_T 下, 该二阶微分方程可以进行求解.

Bertsimas 和 Lo 模型与连续时间形式 类似地, Bertsimas 和 Lo 模型可以通过将 (6.1) 式中的随机游走的增量 ε_t 替换为 σdB_t, 将 θN_t 替换为 $\theta \dot{Y}_t$ 得到. 因此, (6.2) 式的连续时间形式为

$$\min_{Y_t \in \mathcal{A}} E\left[\theta \int_0^T U_t dB_t + U_t^2 dt\right], \quad (6.25)$$

其中 $U_t = \dot{Y}_t$ 是购买股票的速率, $\mathcal{A}$ 是可行策略集 (满足 $Y_0 = 0$ 和 $Y_T = \tilde{N}$ 条件). 利用 CoV, 可知最优执行策略与离散情形的最优解一致: 最优解为 $Y_T^* = \frac{\tilde{N}}{T}$, $Y_t^* = \tilde{N}\left[1 - \left(\frac{t}{T}\right)\right]$, 其中 $Y_0 = 0$, $Y_T = \tilde{N}$.

6.1.3 最优确定性策略

在求解最优执行问题时, 常用的模型框架是将股票价格利用传统的随机游走模型或布朗运动模型建模, 并假设大额股票交易影响股票价格. 该框架的一个巧妙之处是, 最优策略是一个确定性策略. 该策略会产生可预测的交易行为, 因此可能会产生市场操纵行为, 如捕食交易. 参见 Schied 和 Schoneborn (2009) 与 Moallemi 等 (2009) 论文中研究的博弈模型.

从数学上来说, 随机控制模型通常会通过求解 HJB 方程来求解, 而不是通过 CoV, 参见 3.4 节的补充材料 5. Predoiu 等 (2011) 利用分布积分和 P_t 的鞅结构, 巧妙地解释了为什么最优执行问题是一个特例. 定义最初头寸为 $Y_0 = 0$, 末期头寸为 $Y_T = \tilde{N}$. 假定成本为 $\int_0^T P_t dY_t$, 其中 P_t 是一个鞅过程, 则有

$$E\left[\int_0^T P_t dY_t\right] = E\left[P_T Y_T - P_0 Y_0 - \int_0^T Y_t dP_t\right] = P_0 \tilde{N}.$$

因此, 优化问题中不包含 P_t 的随机项. 在目标函数添加了风险项时, 只要该风险项与标的价格过程独立, 该论述依然成立, 类似的方法可以用于求解离散时间框架下的最优确定性策略, 参见 6.3.1 节.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

6.2 乘数价格影响模型

Guo 和 Zervos (2015) 在几何布朗运动 (GBM) 模型下, 假设存在一个乘数价格影响模型, 研究了该情形下的最优执行问题. 求解可得一个随机马尔可夫最优策略, 即最优策略取决于与有效价格过程相关的状态变量. 此处的控制方程是奇异型的. 在接下来, 我们讨论了该模型的基本思想和结果. 3.4 节的补充材料 6 对于奇异随机控制模型给出了相关的背景介绍, 乘数影响模型的具体介绍可以参见 Becherer 等 (2015).

6.2.1 模型设定和随机控制问题

假设有效价格过程 X_t^0 与交易摩擦和影响无关, 服从如下 GBM 过程:

$$dX_t^0 = X_t^0(\mu dt + \sigma dB_t), \tag{6.26}$$

其中 B_t 是一个 $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ 空间上的 $\mathcal{F}_t$ 适应的布朗运动过程, 其中 $\{\mathcal{F}, t \geqslant 0\}$ 是一个信息集. 定义 ξ_t^{s} (或 ξ_t^{b}) 为在时间 t 时, 投资者所购买 (或卖出) 的股票数目, 则 ξ^{s} 和 ξ^{b} 都是 $\mathcal{F}_t$ 适应的 càdlàg 过程, 且 $\xi_0^{\mathrm{s}} = \xi_0^{\mathrm{b}} = 0$. 定义 Y_t 为时间 t 时, 投资者所持有股数, $Y_0 = y \geqslant 0$, 满足

$$Y_t = y - \xi_t^{\mathrm{s}} + \xi_t^{\mathrm{b}}$$

方差有限, 该条件避免了无限的买入或卖出行为. 对于 $T \in (0, \infty]$, 在时间 T 内完成所有资产的清算即为施加如下的限制条件:

$$Y_{T+} = 0 \quad \text{当 } T < \infty, \quad \lim_{T \to \infty} Y_T = 0 \quad \text{当 } T = \infty.$$

此处交易价格过程 X_t 与有效价格过程 X_t^0 之间的关系为

$$X_t = X_t^0 \exp(-\lambda \xi_t^{\mathrm{s}} + \lambda \xi_t^{\mathrm{b}}),$$

其中 $\lambda > 0$ 是买入或卖出行为的价格影响因子, 令 $\delta \geqslant 0$ 为折现因子, $C_{\mathrm{b}}, C_{\mathrm{s}} \geqslant 0$ 为买入或卖出一股的利差, 令

$$X_t \circ_s d\xi_t^{\mathrm{s}} = X_t d(\xi^{\mathrm{s}})_t^{\mathrm{c}} + X_t \int_0^{\Delta \xi_t^{\mathrm{s}}} e^{-\lambda u} du, \tag{6.27}$$

$$X_t \circ_b d\xi_t^{\mathrm{b}} = X_t d(\xi^{\mathrm{b}})_t^{\mathrm{c}} + X_t \int_0^{\Delta \xi_t^{\mathrm{b}}} e^{\lambda u} du, \tag{6.28}$$

其中 $d\xi_t^c$ 是 $d\xi_t$ 的连续时间部分, 附录 A 中给出了鞅过程的背景介绍. 为解释 (6.27) 和 (6.28), 我们假设投资者进行的一小笔交易对于股票价格的影响与交

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

易额线性相关. 特别地, 当投资者在时刻 t 卖出 (或买入) 一个小额 $\varepsilon > 0$ 股股票时, 股票价格呈现出一个跳跃, 幅度为

$$\Delta X_t = X_{t+} - X_t = -\lambda\varepsilon X_t \quad (\text{或 } \Delta X_t = X_{t+} - X_t = \lambda\varepsilon X_t),$$

其中 $\lambda > 0$ 为常数, 等价地,

$$X_{t+} = (1-\lambda\varepsilon)X_t \simeq e^{-\lambda\varepsilon}X_t \quad \left(\text{或 } X_{t+} = (1+\lambda\varepsilon)X_t \simeq e^{\lambda\varepsilon}X_t\right).$$

考虑到在时间 $[0,t]$ 内卖出 $\Delta\xi_t^{\mathrm{s}}$ 股等价于卖出 N 次股票, 每次卖出 $\varepsilon = \Delta\xi_t^{\mathrm{s}}/N$, 卖出的时间为时刻 t 前的 N 个均匀分布的时间段. 这样的卖出交易对于股票价格的影响可以产生的收益为

$$\sum_{j=0}^{N-1} e^{-\lambda j\varepsilon}X_t\varepsilon \simeq \lambda^{-1}X_t(1-e^{-\lambda\Delta\xi_t^{\mathrm{s}}}) = \int_0^{\Delta\xi_t^{\mathrm{s}}} X_t e^{-\lambda u}du.$$

注意到与连续时间且常规的控制设定条件下的附加价格影响模型相比, 此处的控制 ξ^{s} 和 ξ^{b} 是奇异型的. 另外, $\lambda X_t d(\xi^{\mathrm{s}})_t^{\mathrm{c}}$ 是非线性的. 实际上, $X_t d(\xi^{\mathrm{s}})_t^{\mathrm{c}} + X_t\int_0^{\Delta\xi_t^{\mathrm{s}}} e^{-\lambda u}du$ 是一个非线性的价格影响因子, 同时包含了暂时性和永久性影响, 对于永久型影响而言, 其影响随着交易额呈指数递减. 为最大化投资者在 $[0,T]$ 时间内买入和卖出交易的期望收益, 最优执行相关的随机控制问题的价值函数为

$$v(T,x,y) = \sup_{\xi^{\mathrm{s}},\xi^{\mathrm{b}}} E\left[\int_{[0,T]} e^{-\delta t}\left(X_t \circ_s d\xi_t^{\mathrm{s}} - X_t \circ_b d\xi_t^{\mathrm{b}} - C_{\mathrm{s}}d\xi_t^{\mathrm{s}} - C_{\mathrm{b}}d\xi_t^{\mathrm{b}}\right)\right].$$

6.2.2 有限时间情形下的 HJB 方程

正如 3.4 节的补充材料 5 和 6 中指出, 求解前面的随机控制问题的一个标准的方法是将价值函数与 HJB 方程结合起来. 由于该方程可能有多个解, 且可能这些解都不能识别价值方程, 因此该方法需要进行 “检验” 来确定所求得的解是合适的解. 对于前面提到的 $T < \infty$ 下的问题,HJB 方程的形式为

$$\begin{aligned}\max\{-w_t(t,x,y) + \mathcal{L}w(t,x,y), -\lambda x w_x(t,x,y) - w_y(t,x,y) + x - C_{\mathrm{s}},\\ \lambda x w_x(t,x,y) + w_y(t,x,y) - x - C_{\mathrm{b}}\} = 0,\end{aligned} \quad (6.29)$$

其边界条件为

$$w(0,x,y) = \frac{x}{\lambda}\left(1-e^{-\lambda y}\right) - C_s y, \quad (6.30)$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

infoway.io

实时金融行情数据接口

低延迟 • 高可用 • 全覆盖 **专为量化交易设计的行情API**

全市场覆盖 一个接口查所有

股票API
涵盖A股、港股、美股等多个市场。

外汇API
包含56组主要货币对，高频更新。

加密货币API
超过100个币种，全天24小时实时报价。

商品期货API
贵金属、能源天然气、期货指数等

Infoway API是一款专门为量化交易设计的行情接口。支持逐笔成交/订单簿快照等颗粒度，兼容RESTful API和WebSocket，满足高频量化、资产管理系统、行情终端等场景需求。

全域市场一网打尽

每日处理4300万条市场数据流，开发者无需对接多个数据源即可构建多资产策略系统。

简单易接入

我们的API设计注重易用性，支持多种编程语言，确保无论您使用什么开发环境，都能快速将实时行情数据集成到您的应用中。

实时行情接口：www.infoway.io

免费试用！

注册后自动获得试用资格，无需实名认证。（**本活动长期有效**）

联系我们

官网地址：https://infoway.io
客服邮箱：hi@infoway.io



实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 $\mathcal{L}w(t,x,y)=\sigma^2x^2w_{xx}(t,x,y)/2+\mu xw_x(t,x,y)-\delta w(t,x,y)$.

为求解上述 HJB 方程, 考虑如下方法. 在给定的时刻 $t>0$ 每股价格为 $x>0$, 股数为 $y>0$, 此时有三种选择: (i) 等待时间 Δt, 然后继续进行最优选择; (ii) 卖出小额 $\varepsilon>0$ 股份, 然后继续进行最优选择; (iii) 买入小额 $\varepsilon>0$ 股份, 然后继续进行最优选择. 若选择等待, 此时不一定是最优选择, 可以得到如下的不等式:

$$v(t,x,y)\geqslant E\left[e^{-\delta\Delta t}v(t-\Delta t,X_{\Delta t},y)\right].$$

利用 Itô 公式, 除以 Δt, 然后令 $\Delta t\downarrow 0$, 则可知

$$-v_t(t,x,y)+\frac{1}{2}\sigma^2x^2v_{xx}(t,x,y)+\mu xv_x(t,x,y)-\delta v(x,y)\leqslant 0. \tag{6.31}$$

第二种卖出股份的选择则对应于如下的不等式:

$$v(t,x,y)\geqslant v(t,x-\lambda x\varepsilon,y-\varepsilon)+(x-C_{\mathrm{s}})\varepsilon.$$

将公式重新整理, 并令 $\varepsilon\downarrow 0$,

$$-\lambda xv_x(t,x,y)-v_y(t,x,y)+x-C_{\mathrm{s}}\leqslant 0. \tag{6.32}$$

第三种买入的选择对应于如下的不等式:

$$v(t,x,y)\geqslant v(t,x+\lambda x\varepsilon,y+\varepsilon)-(x+C_{\mathrm{b}})\varepsilon.$$

取极限 $\varepsilon\downarrow 0$,

$$\lambda xv_x(t,x,y)+v_y(t,x,y)-x-C_{\mathrm{b}}\leqslant 0. \tag{6.33}$$

由于三种选择中总有一种是最优选择, 因此 (6.31)–(6.33) 中有一个不等式可以取等号, 则可证得 (6.29). 由于该模型假设价格影响是永久性的, 边界条件 (6.30) 对应于时刻 0 时, y 股的价值. Guo 和 Zervos (2015, 命题 4) 证明得到奇异随机控制问题中与价值函数相关的 HJB 方程 (6.29)–(6.30). 相对应地, 论文定义了如下的等待、卖出、买入区域:

$$\begin{aligned}
\mathcal{W}&=\left[(t,x,y)\in[0,T]\times\mathbb{R}_+^*\times\mathbb{R}_+:\mathcal{L}w(t,x,y)=0\right],\\
\mathcal{S}&=\left[(t,x,y)\in[0,T]\times\mathbb{R}_+^*\times\mathbb{R}_+:\lambda xw_x(t,x,y)+w_y(t,x,y)-x+C_{\mathrm{s}}=0\right],\\
\mathcal{B}&=\left[(t,x,y)\in[0,T]\times\mathbb{R}_+^*\times\mathbb{R}_+:\lambda xw_x(t,x,y)+w_y(t,x,y)-x-C_{\mathrm{b}}=0\right],
\end{aligned}$$

其中 $\mathbb{R}_+=[0,\infty)$、$\mathbb{R}_+^*=(0,\infty)$. 论文指出, 当 $\delta\geqslant\mu$ 时, 最优策略中不包含任何的购买交易, 即集合 $\mathcal{B}$ 在此情况下可以直接忽略. 另外, 在此情况下, 当

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$C_{\mathrm{s}}=0$ 时, 最优决策是在时刻 0 时卖出所有的股份. 而当 $\mu=\delta$, $C_{\mathrm{s}}>0$ 时, 在时刻 T 卖出所有的股份可得一系列的 ϵ 最优策略 . 另一方面来说, 如果 $\delta<\mu$, 则有 $v(T,x,y)=\infty$, 那么可以得到无限高的收益.

6.2.3 无限时间情形 $T=\infty$, 且 $0\leqslant\mu<\delta$, $C_{\mathrm{s}}>0$

正如上一节指出, 当满足 $0\leqslant\mu<\delta$ 时, 最优策略中不包含任何的买入行为, 且当 $C_{\mathrm{s}}>0$ 时, 并不需要在时刻 0 时卖出所有的股份. 另外, 在无限时间 $T=\infty$ 情形下, 并不需要在无限时间内对所有的资产进行清算. 因此, 与 (6.30) 不同, 此处我们使用如下的边界条件:

$$w(x,0)=0 \quad \text{对于所有 } x>0, \tag{6.34}$$

HJB 方程可以简化为

$$\max\{\mathcal{L}w(x,y),-\lambda x w_x(x,y)-w_y(x,y)+x-C_{\mathrm{s}}\}=0, \tag{6.35}$$

由于无限时间情形下的价值函数与最优策略是不随时间变化的, 因此在等待区间 $\mathcal{W}$ 内, w 需要满足如下 (关于 x) 的 ODE 方程

$$\frac{1}{2}\sigma^2x^2w_{xx}(x,y)+\mu x w_x(x,y)-\delta w(x,y)=0,$$

其中唯一的边界条件是, 在 $x\downarrow 0$ 时, $w(x,y)=A(y)x^n$, 其中 n 是二次方程 $\sigma^2n(n-1)/2+\mu n-\delta=0$ 的正数解. 由边界条件 (6.34) 可知 $A(0)=0$.

卖出区间 $\mathcal{S}$ 的边界为 $x=F(y)$, 因此当 $x\geqslant F(y)$ 时, $(x,y)\in\mathcal{S}$, 而当 $x<F(y)$ 时, $(x,y)\in\mathcal{W}$. 由于

$$-\lambda x w_x(x,y)-w_y(x,y)+x-C_{\mathrm{s}}=0 \quad \text{当 } x\geqslant F(y), \tag{6.36}$$

对 x 微分可得

$$-\lambda x w_{xx}(x,y)-\lambda w_x(x,y)-w_{yx}(x,y)+1=0 \quad \text{当 } x\geqslant F(y). \tag{6.37}$$

Guo 和 Zervos (2015, 命题 5) 使用了平滑拟合准则 (参见 3.4 节补充材料 5), 令 w 满足 $C^{2,1}$ (即对于 x 二阶连续可导, 对于 y 一阶连续可导), 结合 (6.36), (6.37) 和 $w(x,y)=A(y)x^n$, 可得到关于 $F(y)$, $A(y)$ 和 $\dot{A}(y)$ 的两个联立方程, 其解如下:

$$F(y)=\frac{nC_{\mathrm{s}}}{n-1}:=F_0, \quad \dot{A}(y)F_0^n=-\lambda nA(y)F_0^n+F_0-C_{\mathrm{s}}. \tag{6.38}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

由于 $A(0)=0$, 则 (6.38) 式的 ODE 的解为

$$A(y)=\frac{1}{\lambda n^2}\left(\frac{n-1}{nC_{\mathrm{s}}}\right)^{n-1}(1-e^{-\lambda n y}).$$

在卖出区域 $\mathcal{S}$ 中, 存在两个可行的行为: 一次性卖出所有的股份, 或者卖出一部分, 然后继续进行优化. 令 $\mathbb{Y}(x)=\lambda^{-1}\ln\left(\dfrac{x}{F_0}\right)$. 则 $F_0-x=-x(1-e^{-\lambda\mathbb{Y}(x)})$, $y-\mathbb{Y}(x)>0\Leftrightarrow x<F_0e^{\lambda y}$. 将卖出区域分成以下两部分:

$$\mathcal{S}_1=\{(x,y)\in\mathbb{R}_+^*\times\mathbb{R}_+^*:x\geqslant F_0,y\leqslant\mathbb{Y}(x)\}$$
$$\text{和}\quad\mathcal{S}_2=\{(x,y)\in\mathbb{R}_+^*\times\mathbb{R}_+^*:x\geqslant F_0,y>\mathbb{Y}(x)\},$$

Guo 和 Zervos (2015) 指出, $\mathcal{S}_1$ 对应于在时刻 0 卖出所有股份的区域, 其价值函数为

$$w(x,y)=\frac{1}{\lambda}x(1-e^{-\lambda y})-C_{\mathrm{s}}y\quad\text{当 }(x,y)\in\mathcal{S}_1,$$

而 $\mathcal{S}_2$ 对应于在时刻 0 时卖出 $\mathbb{Y}(x)$ 股, 价值函数为

$$w(x,y)=w(x,y-\mathbb{Y}(x))+\frac{1}{\lambda}x(1-e^{-\lambda\mathbb{Y}(x)})-C_{\mathrm{s}}\mathbb{Y}(x)\quad\text{当 }(x,y)\in\mathcal{S}_2.\ (6.39)$$

如图 6.1 所示, 当股票价格处于等待区域 $\mathcal{W}$ 时, 最优决策是不采取交易; 而当时刻 0 时的股票价格在 $\mathcal{S}_1$ 区域时, 最优决策是立即卖出所有的股份; 而当时刻 0 时的股票价格在 $\mathcal{S}_2$ 区域时, 最优决策是卖出一部分的股票, 使得此

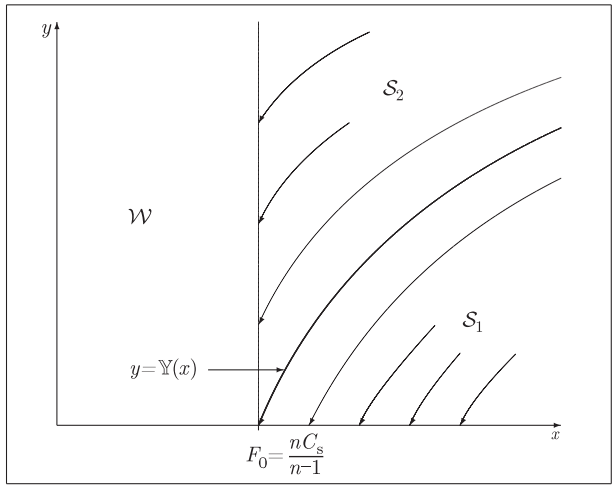


图 6.1 时刻 0 时最优策略中的 “等待区域” “全部卖出” 和 “部分卖出” 区域

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

时并不对于股票产生很大的影响, 使其落到 F_0 区域. Guo 和 Zervos (2015, 命题 5) 证明得到, 价格函数是 HJB 方程的一个 $C^{2,1}$ 形式解, 如下所示:

$$w(x,y)=\begin{cases}0, & \text{当 } y=0, x>0,\\ A(y)x^n, & \text{当 } y>0, x\leqslant F_0,\\ A\left(y-\mathbb{Y}(x)\right)F_0^n+\dfrac{x-F_0}{\lambda}-C_{\mathrm{s}}\mathbb{Y}(x), & \text{当 } y>0, F_0<x<F_0e^{\lambda y},\\ \dfrac{1}{\lambda}x\left(1-e^{-\lambda y}\right)-C_{\mathrm{s}}y, & \text{当 } y>0, F_0e^{\lambda y}\leqslant x.\end{cases}$$

根据 (6.36) 式及 (6.38) 式中 ODE 函数的解, (6.39) 可以表示为 $A(y-\mathbb{Y}(x))F_0^n+\lambda^{-1}(x-F_0)-C_{\mathrm{s}}\mathbb{Y}(x)$.

6.2.4 价格操纵与传导性价格影响

最优执行理论中的价格操纵行为与收益是否可以由零和交易产生相关, 零和交易即在某一个给定的 (有限) 时间内, 买入与卖出的股数净值为 0. 数学上来说, 时间范围 $T>0$ 内的可行零和交易是指一对 $\mathcal{F}_t$-适应的 càdlàg 的交易 $(\zeta^{\mathrm{s}},\zeta^{\mathrm{b}})$, 满足 $\zeta_0^{\mathrm{s}}=\zeta_0^{\mathrm{b}}=0$, $\zeta_{T+}^{\mathrm{s}}=\zeta_{T+}^{\mathrm{b}}$ 且 $\sup\limits_{t\in[0,T]}(\zeta_{t+}^{\mathrm{s}}-\zeta_{t+}^{\mathrm{b}})\leqslant\Gamma$, 其中 $\Gamma>0$ 是一个与交易策略本身相关的常数. 对于附加价格影响模型中的价格操纵存在性的研究很多. Huberman 和 Stanzl (2004) 指出, 对于 $h=0$ 时的 Almgren-Chriss 模型而言, 为保证对于所有的 $T>0$ 都不存在价格操纵, $g(\cdot)$ 应该是一个线性函数, 具体也可参见 Gatheral (2010) 与 Alfonsi 和 Blanc (2014). 对于存在乘数价格影响的 GBM 模型, Guo 和 Zervos (2015) 证明得到当满足 $\delta\geqslant\mu$ 时, 当零和交易的收益非负时, 价格操纵是不存在的.

为了求解传导性价格影响模型, Guo 和 Zervos (2015, 注释 2) 提出此时应该引入一个新的状态变量 $Z_t=\lambda\int_0^t G(t-s)dY_s$, 其中 G 是一个核函数, 刻画价格随时间衰减的过程. 特别地, 在满足 $G(u)=e^{-\gamma u}$ 时, Z_t 服从 $dZ_t=-\gamma Z_tdt-\lambda d\xi_t^{\mathrm{s}}+\lambda d\xi_t^{\mathrm{b}}$ 过程, 且价格函数是 t,x,y 和 z 的函数.

6.3 基于 LOB 数据的最优执行问题

本章前面讨论的最优执行模型都是首先对于标的资产价格过程和特定的交易对价格函数的影响进行假设, 然后引入目标函数. 特别地, 当交易对价格的影响只与交易额相关, 而不涉及市场供给受到一系列交易的影响时, Obizhaeva 和 Wang (2013) 提出, 此时在分析最优执行策略时, 应考虑资产的借给与需求

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

的跨期特性. 论文提出了一个包含 LOB 中获取的供给和需求信息的最优执行模型. 该模型最开始在 2005 年发表工作论文, 受到了很多的关注, 且被多方面扩展. Alfonsi 等 (2010) 对于 Obizhaeva 和 Wang (2013) 的基于 LOB 数据的市场影响模型进行了拓展, 从而满足不同的 LOB 形式. 由于第五章提到的块状 LOB 数据并不能利用 Obizhaeva 和 Wang (2013) 模型进行分析, 此处我们使用 Alfonsi 等 (2010) 模型来进行研究, 并使用该论文的结论来求得 Obizhaeva 和 Wang 假设的块状 LOB 数据的特殊形式下的最优执行策略的显式解.

对于 LOB 对于大笔买入交易的反应随时间衰减的建模 考虑一个在一定的时间范围 $[0,T]$ 内买入 $x_0 > 0$ 股股票的交易者. 由于该问题是关于买入交易, 我们集中分析 LOB 的上半部分的数据, 其包含不同买入价格下的股票数据. 股票的最低买入价格称为*最优买入价格*, 在时刻 0 时, 该价格定义为 A_0. 假设大笔交易在初始时期并不活跃, 因此 LOB 的变化过程仅由噪声交易者的行为所决定. 我们假设相应的未受影响的最佳买入价格 A_t^0 是一个鞅过程, 且满足 $A_0^0 = A_0$. 该假设包含了 Obizhaeva 和 Wang 论文中的特殊情形, 即 $A_t^0 = A_0 + \sigma B_t$. Alfonsi 等 (2010) 指出 "我们可以使用任意的鞅过程, 如几何布朗运动过程", 且 "我们可以允许 A_t^0 中存在任意的跳跃过程, 以对于市场上的大额交易员的行为进行建模". 论文同时指出, "由于此处假设了鞅过程, 最优交易策略最后也会是确定性的."

对于上述最佳买入价格 A_t^0, 论文对于 LOB 中可进行交易的股份, 假设了一个连续的买入价格分布函数: 在价格 $A_t^0 + x$ 上提供的股票数为 $f(x)dx$, 其服从连续函数 $f : \mathbb{R} \to (0,\infty)$, 称为 LOB 的*形状函数*. 注意到 f 并不是一个概率分布函数, 其积分反映的是可进行交易的股票的数量. 常数形状的函数对应于 Obizhaeva 和 Wang 提出的块状 LOB 数据形式. 形状函数决定了大笔交易对于市场交易行为的影响. 假设大笔交易员在时刻 $t = 0$ 时, 买入了 $x_0 > 0$ 股股票, 该笔交易会直接买走价格范围在 A_0 和 $A_0 + D_{0+}^A$ 之间的所有股票, 其中 D_{0+}^A 由 $\int_0^{D_{0+}^A} f(x)dx = x_0$ 所决定. 因此, 最优买入价格则会从 A_0 转移到

$$A_{0+} = A_0 + D_{0+}^A.$$

定义 A_t 为时刻 t 时的实际最优买入价格, 即考虑了之前的大笔买入交易的价格影响之后的最优买入价格. 定义

$$D_t^A = A_t - A_t^0$$

为由大笔交易引起的*额外利差*. 另一个买入 $x_t > 0$ 股的交易则会买走价格在

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

A_t 和

$$A_{t+} = A_t + D^A_{t+} - D^A_t = A^0_t + D^A_{t+}$$

之间的所有股票, 其中 D^A_{t+} 由如下条件决定:

$$\int_{D^A_t}^{D^A_{t+}} f(x)dx = x_t. \tag{6.40}$$

因此, D^A_t 反映了市场买入交易对于现在的最佳买入价格的影响. 除了 f 在 D^A_t 和 D^A_{t+} 之间是常数的情况, $D^A_{t+} - D^A_t$ 函数是交易额 x 的一个非线性函数. 因此, 该模型包含了 6.1 节中提到的 Almgren 和 Chriss 提出的 "非线性影响函数". 该情形实际上属于暂时性影响, Alfonsi 等 (2010, 第 144 页) 进一步解释了为什么增加经典的永久性影响 "在我们的模型中会看起来有些不自然", 以及 "不会改变最佳策略".

模型中另一个重要的变量是 t 时期 "已经消耗" 的股票数量

$$E^A_t = \int_0^{D^A_t} f(x)dx. \tag{6.41}$$

该变量刻画了大额交易员对于 LOB 的 "容量" 的影响. 令

$$F(z) = \int_0^z f(x)dx, \tag{6.42}$$

(6.41) 式可以写成

$$E^A_t = F(D^A_t), \quad D^A_t = F^{-1}(E^A_t), \tag{6.43}$$

此处假设 f 严格为正, 以得到上述第二项. 注意到公式 (6.40) 等价于

$$E^A_{t+} = E^A_t + x_t. \tag{6.44}$$

实证研究表明, 限价订单簿对于大额买入订单对价格的影响有一定的弹性, 即最佳买入价格在受到大额订单影响后, 会反弹到最初的位置. 参见 Biais 等 (1995), Potters 和 Bouchaud (2003), Bouchaud 等 (2004) 以及 Weber 和 Rosenow (2005). Alfonsi 等 (2010) 提出了关于市场影响的弹性的两种模型, 弹性速率为 $\rho > 0$:

模型 1 限价订单簿的回复的速度是指数形式的, 即当最初的时间区间 $[t, t+s)$ 并没有大额的订单交易时, E 服从如下分布:

$$E^A_{t+s} = e^{-\rho s} E^A_t. \tag{6.45}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

模型 2 超额利差 D_t^A 以指数形式衰减, 即在最初的时间区间 $[t, t+s)$ 没有大额的订单交易时,

$$D_{t+s}^A = e^{-\rho s} D_t^A. \tag{6.46}$$

E^A 和 D^A 过程可以分别由模型 1 和模型 2 来表示. Alfonsi 等 (2010) 利用了一个类似的方法, 对于大额交易对 LOB 的低价部分的影响进行了建模, 其中低价部分是指价格低于最佳买入价的股票的交易订单. 对于卖出价, 论文则区分了未受大额交易影响的卖出价 B_t^0 和实际的最佳买入价 B_t, 即考虑了初期的大额卖出订单影响情况下的价格. 低于 B_t^0 部分的卖出价的分布, 可以通过将函数 f 限制在 $(-\infty, 0]$ 范围内来求得. 具体而言, 当 $x < 0$ 时, 在价格 $B_t^0 + x$ 上进行交易的数量为 $f(x)dx$. 差值 $D_t^B := B_t - B_t^0$ 通常为负, 称为买入价分布的*超额价差*. 在时刻 t 时, 卖出 $x_t < 0$ 数量的股票, 则会交易掉价格在 B_t 和

$$B_{t+} := B_t + D_{t+}^B - D_t^B = B_t^0 + D_{t+}^B$$

之间的所有股票. 其中, D_{t+}^B 可以类似于 (6.43), 由如下条件求得:

$$x_t = \int_{D_t^B}^{D_{t+}^B} f(x)dx = F(D_{t+}^B) - F(D_t^B) = E_{t+}^B - E_t^B.$$

注意到 F 的负值部分也是由 (6.42) 式来定义的. 当在一定的时间区间 $[t, t+s)$ 内, 没有产生大笔交易时, D^B 和 E^B 的过程与 D^A 和 E^A 是一致的, 即满足模型 1 中 $E_{t+s}^B = e^{-\rho s} E_t^B$ 和模型 2 中 $D_{t+s}^B = e^{-\rho s} D_t^B$.

6.3.1 成本最小化

对于在时刻 t 时, 单次买入 $x_t \geqslant 0$ 份额股票的交易, 大额交易交易员会在价格 $A_t^0 + x$ 时买入 $f(x)dx$ 股, 其中 x 在 D_t^A 至 D_{t+}^A 区间内. 因此, 买入订单的总成本为

$$\pi_t(x_t) = \int_{D_t^A}^{D_{t+}^A} (A_t^0 + x) f(x)dx = A_t^0 x_t + \int_{D_t^A}^{D_{t+}^A} x f(x)dx. \tag{6.47}$$

在实际中, 大笔订单通常会被拆分为多笔连续的小额订单以降低其对市场价格的影响. 因此, 一个问题是该如何对大笔订单进行拆分以降低成本. 假设大笔交易的交易员要在时间 T 内买入总计 $x_0 > 0$ 股的股票, 并在时间点 $t_n = n\tau$ 上进行 $N+1$ 次卖出, 其中 $n = 0, \ldots, N$, $\tau = \dfrac{T}{N}$. Alfonsi 等 (2010) 定义可行策略集为一系列的随机变量 $\boldsymbol{\xi} = (\xi_0, \xi_1, \ldots, \xi_N)$, 满足

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

- $\sum_{n=0}^{N}\xi_n = x_0$,
- 每个 ξ_n 在 $\mathcal{F}_{t_n}$ 上可测,
- 每个 ξ_n 存在下界,

其中 ξ_n 对应于时刻 t_n 时的交易量 (股数), 且可取值为负 (对应于卖出订单), 但其存在一定的下界.

可行策略集 $\boldsymbol{\xi}$ 的平均成本 $\mathcal{C}(\boldsymbol{\xi})$ 指如下的连续订单的总成本的平均值:

$$\mathcal{C}(\boldsymbol{\xi}) = E\left[\sum_{n=0}^{N}\pi_{t_n}(\xi_n)\right]. \tag{6.48}$$

本节中讨论的优化问题的目标是在所有的可行策略集中, 找到能够最小化平均成本的可行策略. 注意到 $\mathcal{C}(\boldsymbol{\xi})$ 的值取决于此处使用的是模型 1 还是模型 2, 因此最优策略集的数量化特征是由模型决定的. 为方便描述, 我们令 F 为无界的, 即满足

$$\lim_{x\uparrow\infty}F(x) = \infty \quad 和 \quad \lim_{x\downarrow-\infty}F(x) = -\infty. \tag{6.49}$$

因此可知, 限价订单簿具有无限深度, 对于最优策略的发展没有影响.

简化的成本函数与确定性策略的简化 Alfonsi 等 (2010, 附录 A) 提出, 最小化成本函数 (6.48) 可以简化为各个成本函数 $C^{(i)}:\mathbb{R}^{N+1}\to\mathbb{R}$, 其中 $i=1,2$ 代指分析的模型. 论文通过将买卖价差函数合并, 提出了一个简化的模型. 进一步具体来说, 论文对于任意的可行策略 $\boldsymbol{\xi}$, 论文提出了两个随机过程 D 和 E 分别对应于卖出和买入过程, 其满足如下条件:

- $E_0 = D_0 = 0$ 且

$$E_t = F(D_t) \quad 和 \quad D_t = F^{-1}(E_t).$$

- 对于 $n=0,\dots,N$, 无论 ξ_n 符号如何, 都满足

$$E_{t_n+} = E_{t_n} + \xi_n \quad 和 \quad D_{t_n+} = F^{-1}(\xi_n + F(D_{t_n})).$$

- 在模型 1 中, $k=0,\dots,N-1$, $E_{t_{k+1}} = e^{-\rho\tau}E_{t_k+}$, 在模型 2 中 $D_{t_{k+1}} = e^{-\rho\tau}D_{t_k+}$.

在简化过程中, 并不需要 E_t 和 D_t 在 $t\notin\{t_0,\dots,t_N\}$ 时的值. 注意到当策略 $\boldsymbol{\xi}$ 中只考虑买入订单时, 满足 $E=E^A$ 和 $D=D^A$, 而当策略 $\boldsymbol{\xi}$ 中只考虑

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

卖出订单时, 满足 $E = E^B$ 和 $D = D^B$. 论文中引入了时间 t_n 时的简化价格, ξ_n 的价格定义如下:

$$\bar{\pi}_{t_k}(\xi_n) := A^0_{t_n}\xi_n + \int_{D_{t_n}}^{D_{t_n+}} xf(x)dx,$$

与 ξ_n 的方向无关, 论文同时指出

$$\bar{\pi}_{t_n}(\xi_n) \leqslant \pi_{t_n}(\xi_n), \text{ 等式在 } \xi_k \geqslant 0 \text{ 时对于任意 } k \leqslant n \text{ 时成立.}$$

简化的价格函数定义为

$$\bar{\mathcal{C}}(\boldsymbol{\xi}) = E\left[\sum_{n=0}^{N} \bar{\pi}_{t_n}(\xi_n)\right].$$

论文接下来将最小化 $\bar{\mathcal{C}}$ 简化为最小化确定性策略中的函数 $C^{(i)}$. 为此, 论文定义:

$$X_t := x_0 - \sum_{t_k<t} \xi_k \quad \text{对于 } t \leqslant T,\ X_{t_{N+1}} := 0.$$

可行策略 $\boldsymbol{\xi}$ 的累积简化价格为

$$\sum_{n=0}^{N} \bar{\pi}_{t_k}(\xi_n) = \sum_{n=0}^{N} A^0_{t_n}\xi_n + \sum_{n=0}^{N}\int_{D_{t_n}}^{D_{t_n+}} xf(x)dx.$$

利用分部积分法, 可得

$$\sum_{n=0}^{N} A^0_{t_n}\xi_n = -\sum_{n=0}^{N} A^0_{t_n}(X_{t_{n+1}} - X_{t_n}) = x_0A_0 + \sum_{n=1}^{N} X_{t_n}(A^0_{t_n} - A^0_{t_{n-1}}). \quad (6.50)$$

由于 $\boldsymbol{\xi}$ 是可行策略, X_t 是一个有界的可预测过程. 因此, 考虑到未受影响的最优买入价格过程 A^0_t 的鞅性, (6.50) 式的期望为 x_0A_0.

接下来, 注意到在模型 $i = 1, 2$ 中, 在给定 $\xi_0, \xi_1, \ldots, \xi_N$ 的值时, 简化的超额价差过程 D 是一个确定性的过程. 因此, 存在一个确定性的方程 $C^{(i)}$: $\mathbb{R}^{N+1} \to \mathbb{R}$, 满足

$$\sum_{n=0}^{N}\int_{D_{t_n}}^{D_{t_n+}} xf(x)dx = C^{(i)}(\xi_0, \ldots, \xi_N). \quad (6.51)$$

该方程满足

$$\bar{\mathcal{C}}(\boldsymbol{\xi}) = A_0x_0 + E\left[C^{(i)}(\xi_0, \ldots, \xi_N)\right].$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

6.3.2 节和 6.3.3 节指出, $C^{(i)}$, $i=1,2$ 存在唯一的最小值, 且该值属于

$$\Xi=\left\{(z_0,\dots,z_N)\in\mathbb{R}^{N+1}:\sum_{n=0}^{N}z_n=x_0\right\}.$$

该最小值对应了模型 i $(i=1,2)$ 最优确定性策略 $\boldsymbol{\xi}^{(i)}$.

6.3.2 模型 1 的最优策略

函数 $C^{(1)}$ 的最小值 假设函数 $h_1:\mathbb{R}\to\mathbb{R}_+$:

$$h_1(y):=F^{-1}(y)-e^{-\rho\boldsymbol{\tau}}F^{-1}(e^{-\rho\boldsymbol{\tau}}y),$$

是一个一一映射, 则存在一个唯一的最优策略 $\boldsymbol{\xi}^{(1)}=(\xi_0^{(1)},\dots,\xi_N^{(1)})$. 初始的市场订单 $\xi_0^{(1)}$ 是如下方程的唯一最优解:

$$F^{-1}\left(x_0-N\xi_0^{(1)}(1-e^{-\rho\boldsymbol{\tau}})\right)=\frac{h_1(\xi_0^{(1)})}{1-e^{-\rho\boldsymbol{\tau}}}. \tag{6.52}$$

中间的订单为

$$\xi_1^{(1)}=\cdots=\xi_{N-1}^{(1)}=\xi_0^{(1)}(1-e^{-\rho\boldsymbol{\tau}}), \tag{6.53}$$

而最后的订单为

$$\xi_N^{(1)}=x_0-\xi_0^{(1)}-(N-1)\xi_0^{(1)}(1-e^{-\rho\boldsymbol{\tau}}).$$

最优策略是确定性策略, 是函数 $C^{(1)}$ 的最小值. 另外, 最优策略只包含买入策略, 即对于所有 n, 满足 $\xi_n^{(1)}>0$.

对于该结果, 有几个评论. 首先, 最优策略 $\boldsymbol{\xi}^{(1)}$ 只包含买入订单, 因此卖出价格过程是不受影响的, 因此 $E_t^B\equiv 0\equiv D_t^B$. 另外, 当 $\xi_0^{(1)}$ 由公式 (6.52) 确定时, 最优策略包含一系列的订单, 在反弹效应作用下, 每个订单的份额刚好等于 LOB 在之前的市场中收到的份额. 在最终时刻 $t_N=T$, 剩下的所有份额都会被买入. Alfonsi 等 (2010, 附录 B) 给出了 $C^{(1)}$ 的一个显式解. 令 $a=e^{-\rho\boldsymbol{\tau}}$, 定义 $\tilde{F}(z)=\int_0^z xf(x)dx$, $G(y)=\tilde{F}(F^{-1}(y))$. 对于任意的 $\boldsymbol{z}=(z_0,\dots,z_N)\in\Xi$, $C^{(1)}(z_0,\dots,z_N)$ 等于

$$G(z_0)-G(0)+G(az_0+z_1)-G(az_0)+G(a^2z_0+az_1+z_2)-G(a^2z_0+az_1)$$
$$+\cdots+G\left(\sum_{k=0}^{N}a^kz_{N-k}\right)-G\left(\sum_{k=1}^{N}a^kz_{N-k}\right),$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

且其二阶可微, 令

$$\hat{h}_1(y) = h_1(y) - (1-a)F^{-1}(x_0 - Ny(1-a)).$$

Alfonsi 等 (2010, 引理 B.2) 表明, 当 h_1 是一一映射时, $\hat{h}_1$ 最多有一个零值. 将该结果与在 $\|\boldsymbol{z}\| \to \infty$ 时, $C^{(1)}(\boldsymbol{z}) \to \infty$ 的结果相结合 (且 $C^{(1)}$ 在 Ξ 中最多有一个最小值). 论文同时指出, 该最小值是唯一的, 最小值为 $\boldsymbol{\xi}^{(1)}$.

h_1 在什么情况下是一一映射? 函数 h_1 是一个连续函数, 满足 $h_1(0) = 0$, 且 $y > 0$ 时 $h_1(y) > 0$. 因此, h_1 当且仅当 h_1 是一个严格递增过程时是一个一一映射. 具体而言,

$$h_1'(y) = \frac{1}{f(F^{-1}(y))} - \frac{e^{-2\rho\boldsymbol{\tau}}}{f(F^{-1}(e^{-\rho\boldsymbol{\tau}}y))} > 0$$

等价于

$$l(y) := f(F^{-1}(e^{-\rho\boldsymbol{\tau}}y)) - e^{-2\rho\boldsymbol{\tau}}f(F^{-1}(y)) > 0.$$

因此, 当 f 对于 $y > 0$ 递减, 而对于 $y < 0$ 递增时, h_1 是一个一一映射. 实际上, 在很多实证研究中, LOB 通常会在最优报价或者其附近达到最大值, 并随后随着与最优报价的距离增加而递减, 该结论也与原假设相符合.

6.3.3 模型 2 的最优策略

类似地, Alfonsi 等 (2010, 附录 C) 求得了 $C^{(2)}$, 并提出当满足如下条件时, 存在唯一的最小值.

函数 $C^{(2)}$ 的最小值 假设函数 $h_1 : \mathbb{R} \to \mathbb{R}$ 定义为

$$h_2(x) := x\frac{f(x) - e^{-2\rho\boldsymbol{\tau}}f(e^{-\rho\boldsymbol{\tau}}x)}{f(x) - e^{-\rho\boldsymbol{\tau}}f(e^{-\rho\boldsymbol{\tau}}x)}$$

且函数是一个一一映射, 其形状函数满足

$$\lim_{|x|\to\infty} x^2 \inf_{z\in[e^{-\rho\tau}x,x]} f(z) = \infty.$$

则存在唯一的最优策略 $\boldsymbol{\xi}^{(2)} = (\xi_0^{(2)}, \ldots, \xi_N^{(2)})$. 最初的市场订单 $\xi_0^{(2)}$ 是如下方程的唯一解:

$$F^{-1}\left(x_0 - N\left[\xi_0^{(2)} - F\left(e^{-\rho\boldsymbol{\tau}}F^{-1}(\xi_0^{(2)})\right)\right]\right) = h_2\left(F^{-1}(\xi_0^{(2)})\right),$$

中间订单为

$$\xi_1^{(2)} = \cdots = \xi_{N-1}^{(2)} = \xi_0^{(2)} - F\left(e^{-\rho\boldsymbol{\tau}}F^{-1}(\xi_0^{(2)})\right),$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

而最终订单由下式决定:

$$\xi_N^{(2)} = x_0 - N\xi_0^{(2)} - (N-1)F\left(e^{-\rho\boldsymbol{\tau}}F^{-1}(\xi_0^{(2)})\right).$$

最优策略是一个确定性策略, 是函数 $C^{(2)}$ 的最小值. 另外, 最优策略只包含买入订单, 即对于任意 n, 满足 $\xi_n^{(2)} > 0$.

6.3.4 块状 LOB 的解析解

Alfonsi 等 (2010, 第 6 节) 利用一个对应于恒定的形状函数 $f(x) \equiv q$ ($q > 0$) 的块状 LOB 来阐述了他们的结果. 论文指出: "在该情形下, 模型 1 和模型 2 并没有区别", 存在唯一的最优解 $\boldsymbol{\xi}^*$ 如下:

$$\xi_0^* = \xi_N^* = \frac{x_0}{(N-1)(1-e^{-\rho\tau})+2},$$
$$\xi_1^* = \cdots = \xi_{N-1}^* = \frac{x_0 - 2\xi_0^*}{N-1}.$$

该结果 "给出 Obizhaeva 和 Wang 得到的递归模型的解析解". 递归模型是通过求解满足如下对于 D_t^A 的假设的动态规划问题得到的:

$$D_t^A = \lambda\sum_{k:t_k<t}\xi_k + \sum_{k:t_k<t}\left(\frac{1}{q}-\lambda\right)e^{-\rho(t-t_k)}\xi_k,$$

其中 $\lambda < 1/q$ 为永久性影响, $1/q - \lambda$ 为暂时性影响.

6.4 投资组合的最优执行

在分析了单个资产的最优执行问题后, 我们自然地将该问题拓展到投资组合上. 具体而言, 当投资人需要在时间 T 内对于 $m \geqslant 1$ 只股票进行买入, 该如何选择才能找到最优的执行策略, 从而最小化期望执行成本? 目前对于该问题的研究较少, 且基本上并没有计量经济学方向的文章对于 4.4 节中提到的异步性和同步性的问题进行研究.

Bertsimas 和 Lo 的线性价格影响模型 Bertsimas 和 Lo 模型已被拓展到多元情况, 在多元模型中, 标的价格模型是一个 GBM 过程, 而价格影响模型仍然是一个线性可加模型. 定义 $\boldsymbol{N}_t = (N_{1t}, N_{2t}, \ldots, N_{mt})^\top$, N_{it} 表示在时刻 t 以 P_t^i 的价格所购买的股票 i 的数量. 定义 $\boldsymbol{P}_t = (P_t^1, P_t^2, \ldots, P_t^m)^\top$, $t = 1, \ldots, T$. 令执行价格向量 $\boldsymbol{P}_t$ 为如下两项之和:

$$\boldsymbol{P}_t = \tilde{\boldsymbol{P}}_t + \boldsymbol{\delta}_t, \tag{6.54}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 $\tilde{\boldsymbol{P}}_t = (\tilde{P}_t^1, \ldots, \tilde{P}_t^m)^\top$ 是在不存在价格影响时的有效价格向量, 用如下的多元 GBM 过程表示:

$$\tilde{\boldsymbol{P}}_t = \operatorname{diag}\left[\exp(z_t^1), \ldots, \exp(z_t^m)\right] \tilde{\boldsymbol{P}}_{t-1}, \tag{6.55}$$

其中 $\boldsymbol{z}_t = (z_t^1, \ldots, z_t^m)^\top$ 独立同分布, 服从 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. 价格影响项 $\boldsymbol{\delta}_t$ 定义为

$$\boldsymbol{\delta}_t = \tilde{\boldsymbol{p}}_t(\boldsymbol{A}\boldsymbol{N}_t + \boldsymbol{B}\boldsymbol{X}_t), \tag{6.56}$$

其中 $\boldsymbol{A}$ 是一个正定 $m \times m$ 矩阵, $\boldsymbol{X}_t$ 是一个包含 n 个信息变量的向量, $\boldsymbol{B}$ 是任意的 $m \times n$ 矩阵, $\tilde{\boldsymbol{p}}_t = \operatorname{diag}(\tilde{P}_t^1, \ldots, \tilde{P}_t^m)$, 且有

$$\boldsymbol{X}_t = \boldsymbol{C}\boldsymbol{X}_{t-1} + \boldsymbol{\eta}_t, \tag{6.57}$$

其中 $\boldsymbol{C}$ 是一个 $n \times n$ 矩阵, 其特征值的模小于 1, $\boldsymbol{\eta}_t$ 独立同分布, 均值为 $\boldsymbol{0}$, 协方差矩阵为 $\boldsymbol{V}$. 注意到价格影响作为有效价格 $\tilde{\boldsymbol{P}}_t$ 的一部分, 是 m 个股票的交易额和其他状态变量 $\boldsymbol{X}_t$ 的线性函数. 定义 $\boldsymbol{Y}_t$ 为时刻 t 时剩下的股票数量, 则有

$$\boldsymbol{Y}_1 = \tilde{\boldsymbol{N}}, \boldsymbol{Y}_{T+1} = \boldsymbol{0}, \quad \text{且} \quad \boldsymbol{Y}_t = \boldsymbol{Y}_{t-1} - \boldsymbol{N}_{t-1}. \tag{6.58}$$

正如 (6.2) 式指出, 最优执行问题为

$$\min_{\boldsymbol{N}_1, \ldots, \boldsymbol{N}_T} E\left[\sum_{t=1}^{T} \boldsymbol{P}_t^\top \boldsymbol{N}_t\right] \quad \text{满足} \ \sum_{t=1}^{T} \boldsymbol{N}_t = \tilde{\boldsymbol{N}}. \tag{6.59}$$

Bertsimas 等 (1999) 通过逆推法求解动态规划问题, 对于 (6.10) 和 (6.11) 式进行了如下一般化:

$$\boldsymbol{N}_{T-k}^* = \left(\boldsymbol{I} - \frac{1}{2}\boldsymbol{A}_{k-1}^{-1}\boldsymbol{A}^\top\right)\boldsymbol{Y}_{T-k} + \frac{1}{2}\boldsymbol{A}_{k-1}^{-1}\boldsymbol{B}_{k-1}^\top \boldsymbol{C}\boldsymbol{X}_{T-k}, \tag{6.60}$$

$$\begin{aligned} V_{T-k} = {} & \boldsymbol{P}_{T-k-1}^\top \boldsymbol{Y}_{T-k} + \boldsymbol{Y}_{T-k}^\top \boldsymbol{A}_k \boldsymbol{Y}_{T-k} \\ & + \boldsymbol{X}_{T-k}^\top \boldsymbol{B}_k \boldsymbol{Y}_{T-k} + \boldsymbol{X}_{T-k}^\top \boldsymbol{C}_k \boldsymbol{X}_{T-k} + d_k, \end{aligned} \tag{6.61}$$

其中 $k = 0, \ldots, T-1$,

$$\begin{aligned} \boldsymbol{A}_k &= \boldsymbol{A} - \frac{1}{4}\boldsymbol{A}\boldsymbol{A}_{k-1}^{-1}\boldsymbol{A}^\top, & \boldsymbol{A}_0 &= \boldsymbol{A}, \\ \boldsymbol{B}_k &= \frac{1}{2}\boldsymbol{C}^\top \boldsymbol{B}_{k-1}(\boldsymbol{A}_{k-1}^\top)^{-1}\boldsymbol{A}^\top + \boldsymbol{B}^\top, & \boldsymbol{B}_0 &= \boldsymbol{B}^\top, \\ \boldsymbol{C}_k &= \boldsymbol{C}^\top \boldsymbol{C}_{k-1}\boldsymbol{C} - \frac{1}{4}\boldsymbol{C}^\top \boldsymbol{B}_{k-1}(\boldsymbol{A}_{k-1}^\top)^{-1}\boldsymbol{B}_{k-1}\boldsymbol{C}, & \boldsymbol{C}_0 &= \boldsymbol{0}, \\ \mathrm{d}_k &= d_{k-1} + E[\boldsymbol{\eta}_{T-k}^\top \boldsymbol{C}_{k-1}\boldsymbol{\eta}_{T-k}], & d_0 &= 0. \end{aligned}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

Obizhaeva 和 Wang 的 LOB 模型的投资组合模型的拓展 Tsoukalas 等 (2014) 指出, 目前单个资产的最优执行问题的研究主要有两个方向. 第一个方向 "基于实证数据研究价格影响函数", 比如 Bertsimas 和 Lo 与 Almgren 和 Chriss. 第二个方向, 以 Obizhaeva 和 Wang (2013) 为例, "重点分析市场微观结构基础", 如 LOB 中反映出来的流动性的供给和需求. 为将第二个方向的研究拓展到投资组合模型上, 论文提出了一个多资产 LOB 模型, 模型中包含相关性风险和共轭买入/卖出动态过程, 即 "任何一个买入或卖出的订单都会影响到买入和卖出两侧的现有限价订单簿的存货和未来订单". 考虑到可行性问题, 论文引入了 Obizhaeva 和 Wang 关于块状 LOB 的假设以及对于 $\tilde{\boldsymbol{P}}_t$ 的随机游走模型. 具体而言, LOB 中资产 i 在时刻 t 时在买入和卖出两侧的形状函数为

$$f_{i,t}^A = q_i^A I_{\{x \geqslant a_{i,t}\}}, \quad f_{i,t}^B = q_i^B I_{\{x \leqslant b_{i,t}\}} \tag{6.62}$$

且 $\tilde{\boldsymbol{P}}_t - \tilde{\boldsymbol{P}}_{t-1}$ 独立同分布, 服从 $N(\boldsymbol{0}, \tau\boldsymbol{\Sigma})$, 其中正如 6.3.1 节中定义的, $\tau = T/N$. 另外, $a_{i,t}, b_{i,t}$ 分别是资产 i 在时刻 t 时的最优卖出价和最优买入价.

对于资产 i, 在时刻 t 时, 买入或卖出 $N_{i,t}^+$ ($N_{i,t}^-$) 订单的永久价格影响为 $\lambda_{ii}(N_{i,t}^+ - N_{i,t}^-)$. 考虑到资产 j 的价格受到其他资产的交叉影响 λ_{ij}, Tsoukalas 等 (2014) 定义 $\boldsymbol{\Lambda} = (\lambda_{ij})_{1 \leqslant i,j \leqslant m}$ 和向量 $\boldsymbol{N}_t^+ = (N_{1,t}^+, \ldots, N_{m,t}^+)^\top$, $\boldsymbol{N}_t^- = (N_{1,t}^-, \ldots, N_{m,t}^-)^\top$, $\boldsymbol{N}_t = ((\boldsymbol{N}_t^+)^\top, (\boldsymbol{N}_t^-)^\top)^\top$. 交易者的总净交易资产之和为 $\boldsymbol{N}$, $\sum_{t=0}^T (\boldsymbol{N}_t^+ - \boldsymbol{N}_t^-) = \sum_{t=0}^T \boldsymbol{\delta}\boldsymbol{N}_t = \boldsymbol{N}$, 其中 $\boldsymbol{\delta}$ 是一个 $m \times (2m)$ 矩阵 $(\boldsymbol{I}, -\boldsymbol{I})$, $\boldsymbol{I}$ 为 $m \times m$ 单位矩阵. 类似于公式 (6.58), 定义 $\boldsymbol{Y}_0 = \boldsymbol{N}$, $\boldsymbol{Y}_t = \boldsymbol{Y}_{t-1} - \boldsymbol{\delta}\boldsymbol{N}_{t-1}$. 同时定义暂时性价格影响矩阵

$$\boldsymbol{Q}^A = \operatorname{diag}\left(\frac{1}{2q_1^A}, \ldots, \frac{1}{2q_m^A}\right), \quad \boldsymbol{Q}^B = \operatorname{diag}\left(\frac{1}{2q_1^B}, \ldots, \frac{1}{2q_m^B}\right)$$

并定义 LOB 的买入和卖出侧的补充过程为

$$\begin{aligned}
\boldsymbol{d}_t^A &= \operatorname{diag}(e^{-\tau\rho_1^A}, \ldots, e^{-\tau\rho_m^A}) \left\{\boldsymbol{d}_{t-1}^A + 2\boldsymbol{Q}^A \boldsymbol{N}_{t-1}^+ - \boldsymbol{\Lambda}\boldsymbol{\delta}\boldsymbol{N}_{t-1}\right\}, \\
\boldsymbol{d}_t^B &= \operatorname{diag}(e^{-\tau\rho_1^B}, \ldots, e^{-\tau\rho_m^B}) \left\{\boldsymbol{d}_{t-1}^B + 2\boldsymbol{Q}^B \boldsymbol{N}_{t-1}^- - \boldsymbol{\Lambda}\boldsymbol{\delta}\boldsymbol{N}_{t-1}\right\}.
\end{aligned}$$

定义

$$\begin{aligned}
\boldsymbol{a}_t &= \tilde{\boldsymbol{P}}_t + \frac{1}{2}\boldsymbol{s}_0 + \boldsymbol{\Lambda}\sum_{k=0}^{t-1} \boldsymbol{N}_k + \boldsymbol{d}_t^A, \\
\boldsymbol{b}_t &= \tilde{\boldsymbol{P}}_t - \frac{1}{2}\boldsymbol{s}_0 + \boldsymbol{\Lambda}\sum_{k=0}^{t-1} \boldsymbol{N}_k + \boldsymbol{d}_t^B,
\end{aligned}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 $\boldsymbol{s}_0$ 是稳定状态的买入卖出价差, 论文发现交易者在时刻 t 时的收益为 $\pi_t = (\boldsymbol{b}_t - \boldsymbol{Q}^B \boldsymbol{N}_t^-)^\top \boldsymbol{N}_t^- - (\boldsymbol{a}_t - \boldsymbol{Q}^A \boldsymbol{N}_t^+)^\top \boldsymbol{N}_t^+$, 即为收益 (通过卖出股票) 和成本 (通过买入股票) 之差. 交易者的最终财富为 $W_N = \sum_{t=0}^N \pi_t$. 论文分析了最大化最终财富的指数化效用问题, 优化问题的约束为 $\sum_{t=0}^T \boldsymbol{\delta} \boldsymbol{N}_t = \boldsymbol{N}$. 论文得到了一个最优执行策略 $(\boldsymbol{N}_t^*,\ 0 \leqslant t \leqslant T)$, 该策略是一个确定性策略, 即与 $\tilde{\boldsymbol{P}}_t$ 无关.

6.5 最优配置

在给定的约束条件下, 将大额订单 (父订单) 拆分成小额订单 (子订单) 以最小化市场影响后, 接下来的步骤是资产配置. 在交易平台中进行资产配置时, 交易员需要考虑三个问题: (a) 是否使用市场订单, 是否使用限价订单, 或同时使用这两者; (b) 在不同的价格水平下, 应该配置的订单数目; (c) 在给定的时间框架下, 应该执行的订单序列. 当使用限价订单时, 交易者并不一定要支付成本, 反而可能会得到 "折扣", 但是由于限价订单并不一定得到最终执行, 所以该折扣也往往伴随着执行风险. 另一方面, 当使用市场订单时, 交易者需要同时支付限价订单与市场订单之间的价差, 以及确保订单执行的费用. 本质上来说, 交易员需要对两者进行平衡, 一方面是执行市场订单时需要支付的价差和费用, 另一方面是执行限价订单的执行/存货风险. 由于在不同的交易所中限价订单的折扣率是不同的, 因此对于不同交易所, 优化问题可能是不一样的. 在一些交易所中, 成功地执行限价订单会得到一个折扣, 折扣值为执行价格的一个比例, 而在一些其他的交易所中, 折扣会是一个固定的数额. 在本章中, 我们考虑在同一个交易所中交易的情况, 此时费用是固定的.

理论上来说, 最优配置问题可以如下表示. 考虑一个在时间 $T > 0$ 内 ($T \leqslant 10$ 秒) 买入 N 股股票的交易. 交易者可以将这 N 股拆分为 $(N_{0,t}, N_{1,t}, \dots)$, 其中 $N_{0,t} \geqslant 0$ 是时刻 t 时交易的市场订单的股数, $N_{1,t}$ 是时刻 t 时以最优买入价买入的股数, 而 $N_{2,t}$ 是时刻 t 时以次优价格交易的股数, 依次类推 $t = 0, 1, \dots, T$. 当在时间 T 内没有进行限价股的交易时, 交易者需要在时刻 T 以市场价购买非执行的限价股. 而在当时间区间内执行了限价股时, 市场需要给予交易员折扣 $r > 0$, 而当执行市场订单时, 则需要支付费用 $f > 0$. 在交易过程中, 虽然不允许有卖出的行为, 但是交易员可以在中途取消未执行的订单, 并在以后重新执行该订单. 在给定 N 和 T 时, 最优执行问题可以理解为寻找最优策略 $(N_{0,t}, N_{1,t}, \dots, N_{k,t})_{t=0,1,\dots,T}$ 以最小化总期望成本. 分析该问题的关

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

键在于分析在特定的价格水平下, 限价订单被执行的概率. 概率的计算往往是非常复杂的, 其复杂程度取决于模型的复杂程度. 接下来, 我们提出了两个模型, 模型中订单配置不受价格影响, 此时该问题可以简化为 $N = 1$ 的情形.

6.5.1 包含均值回归的马尔可夫随机游走模型

Guo 等 (2013) 研究了 $N = 1$ 时的最优配置问题, 假设

1. 最优卖出价与最优买入价的差为 1 个最小报价单位 (Tick).
2. 在每个时刻 $t = 0, 1, \ldots, T$, 最优卖出价格增加或减小一个单位.

定义 A_t 为时刻 t 时的最优卖出价格, 则

$$A_t = \sum_{i=1}^{t} X_i, \quad A_0 = 0, \tag{6.63}$$

其中 X_i $(1 \leqslant i \leqslant T)$ 是在 $\{-1, 1\}$ 上的马尔可夫链, 且有

$$P(X_1 = 1) = \bar{p} = 1 - P(X_1 = -1), \qquad \bar{p} \in [0, 1], \tag{6.64}$$

$$\begin{aligned} P(X_{i+1}=1|X_i = 1) &= P(X_{i+1} = -1|X_i = -1) \\ &= p < \frac{1}{2} \qquad \text{对于 } i = 1, \ldots, T-1. \end{aligned} \tag{6.65}$$

令 $p < 1/2$, 则价格过程是 “均值回归” 的, 该特征在高频交易中是非常常见的. Cont 和 de Larrard (2013) 发现在文献上常见的价格变化服从的负一阶自回归过程 (参见 2.1.2 和 4.2 节) 实际上与 $p < 1/2$ 是等价的. 论文同时对于 CITI 等流动性股票进行研究, 得到 $p = 0.35$. Chen 和 Hall (2013) 进一步对于包含均值回归的高频交易过程进行了统计分析. 除了价格动态过程外, Guo 等 (2013) 也对于给定 A_t 时, 限价订单在时刻 T 之前被执行的条件概率做了如下假设:

- 如果对于某些 $t \leqslant T$, $A_t \leqslant -k$, 则限价订单在 $-k$ 价格上被执行的概率为 1;
- 如果对于所有的 $t \leqslant T$, 满足 $A_t > -k + 1$, 则限价订单在 $-k$ 价格上被执行的概率为 0;
- 如果 $A_t = -k + 1$ 且 $A_{t+1} = -k + 2$, 则限价订单在 t 和 $t + 1$ 的时间区间内, 以 $-k$ 价格被执行的概率为 q.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

静态情形下的最优配置策略 假设交易者仅可以在时刻 $t=0$ 时决定配置策略. 当订单是一个限价订单, 且没有在任意的 $t<T$ 时刻被执行时, 在时刻 $t=T$ 时必须执行一个市场订单. 显然, 在此情况下求解最优配置问题等价于比较期望成本. 期望成本取决于 LOB 中以价格 $-k$ (即比初始最优卖出价低 k 个单位) 成交的订单的参数 $f,r,p,q,\bar{p}$, 其中 k 的值可变, r 是限价订单的折扣, f 是市场订单的费率. 我们此处集中讨论 q 和 $\bar{p}$. 定义 $C(k,q,T,\bar{p})$ 为以 k 个单位低于初始最优卖出订单价格成交的限价订单的期望成本. 该方程的主要结果参见下面的 (R1)–(R3), 且同时讨论了最优配置策略的计算过程. 注意到 $C(0,q,T,\bar{p})$ 为市场订单的期望成本. 根据 (A2), 价格低于 $-T-1$ 的订单在时刻 T 之前不会被执行, 且此时会在时刻 T 时执行市场订单, 因此其期望成本都是一样的, 此处记为 $C(T+1,q,T,\bar{p})$. 因此优化问题可以表示成

$$\min_{0\leqslant k\leqslant T+1} C(k,q,T,\bar{p}) = \min_{0\leqslant k<\infty} C(k,q,T,\bar{p}).$$

对于最优执行策略, 利用 A_t 的均值回归特性和最小值的单调性 (参见 6.6 节的补充材料 9):

(R1) 当 $q=0$ 时, $C(1,0,T,\bar{p})<C(2,0,T,\bar{p})<\cdots<C(T+1,0,T,\bar{p})$. 最优卖出价格优于任意价格的限价订单.

(R2) 当 $q\neq 0$ 时, $C(2,q,T,\bar{p})<C(3,q,T,\bar{p})<\cdots<C(T+1,q,T,\bar{p})$. 最优配置策略只取决于比较下面三个价格水平下的期望成本: 市场订单、最优买入订单和次优买入订单.

(R3) 当 $\bar{p}\geqslant 1-p$ 时, $C(1,q,T,\bar{p})<C(2,q,T,\bar{p})$. 最优配置策略只包含市场订单和最优买入订单.

该结果显著地减少了最优配置问题的计算成本. 与计算 LOB 每个价格水平的期望成本不同, 最优配置策略的分析只包含计算三个成本函数: 市场订单的成本 $C(0,q,T,\bar{p})$, 最优买入价格对应的成本 $C(1,q,T,\bar{p})$, 次优买入价格对应的成本 $C(2,q,T,\bar{p})$.

动态情形下的最优配置策略 假设交易者需要在时刻 T 之前购买一股股票, 且允许其在时刻 t 时以任意价格来购买, 且在后续时间内可以取消订单或者将限价订单改成市场订单. 在时刻 T 时, 尚未执行的限价订单会被自动替换成市场订单. 在该动态设定下, 在任意时刻 t, 只需要考虑两个价格水平下的成本: 最优买入价和市场订单价格, 或者并没有下单. 在每个时间点, 价格的变化

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

都是一个单位, 因此以比最优执行价格低的价格执行订单等价于不下订单的情形, 因此此时订单并不会在下一个时间段被执行. 考虑到 (A_t, X_t) 的马尔可夫特性, 最优配置问题是一个马尔可夫决策模型, 每一个时间段的期望成本可以用递归法来求得. 实际上, 我们可以利用价值函数的齐次性来得到最优解的解析形式. 最优策略是一个包含阈值的策略.

在时间 $t \leqslant T-1$ 的最优配制策略 存在两个整数 t_1^*, t_2^* 满足 $t_1^* < t_2^*$, 且满足

- 如果 $X_t = 1$, 则最优策略是在时间 $t < t_1^*$ 不下单, 并在时间 $t \geqslant t_1^*$ 以最优买入价下单.
- 如果 $X_t = -1$, 则在时间 $t < t_2^*$ 以最优买入价下单, 并在时间 $t \geqslant t_1^*$ 以市场订单下单.

具体可参见图 6.2 和图 6.3.

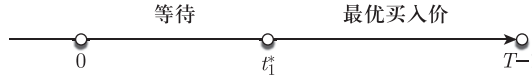


图 6.2 当 $X_t = 1$ 时对于 $1 < t \leqslant T-1$ 的最优配制策略

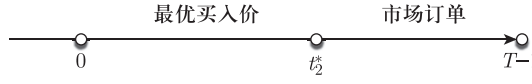


图 6.3 当 $X_t = -1$ 时对于 $1 < t \leqslant T-1$ 的最优配制策略

t_1^*, t_2^* 和 $V_t^*(1)$, $V_t^*(-1)$ 的解析表达式如下, 价值函数是在 $X_t = \pm 1$ 处求得的:

$$t_1^* = T - \left\lceil \frac{1}{\log(p-pq)} \cdot \log \frac{q(1-2p)}{((1-p+pq)(f+2+r)-1)(1-q)(p^2q+1+p^2-2p)} \right\rceil - 1,$$

$$t_2^* = T - \left\lceil \frac{1}{\log(p-pq)} \cdot \log \frac{(f+r+1)(1-p+pq)-(1-p)(1-q)}{(1-p)(1-q)((f+2+r)(1-p+pq)-1)} \right\rceil.$$

$$V_t^*(-1) = \begin{cases} f, & \text{当 } t \geqslant t_2^*, \\ (p+(1-p)q)(-1-r) + (1-p)(1-q)(1+V_{t+1}^*(1)), & \text{当 } t < t_2^*. \end{cases}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$$V_t^*(1)=\begin{cases}(p-pq)^{T-t}\left(f+2+r-\dfrac{1}{1-p+pq}\right)-2-r+\dfrac{1}{1-p+pq}\\ \qquad \text{当 } t_1^*\leqslant t\leqslant T,\\ A\left(\dfrac{a_1+\sqrt{a_1^2+4a_2}}{2}\right)^{t_1^*-t+1}+B\left(\dfrac{a_1-\sqrt{a_1^2+4a_2}}{2}\right)^{t_1^*-t+1}+a_4\\ \qquad \text{当 } t<t_1^*,\end{cases}$$

其中 $a_1=p$, $a_2=(1-p)(1-p-q+pq)$, $a_3=2p-1-(1+r)(1-p)(p+q-pq)+(1-p)(1-p-q+pq)$, $a_4=a_3/(1-a_1-a_2)$, 且 A 和 B 的系数由如下的线性方程确定:

$$A+B=V_{t_1^*+1}^*(1)-a_4,$$

$$A\frac{a_1+\sqrt{a_1^2+4a_2}}{2}+B\frac{a_1-\sqrt{a_1^2+4a_2}}{2}=V_{t_1^*}^*(1)-a_4.$$

由于价格的均值回归特性, 最优配置策略对于剩余的交易时间与市场的动量敏感. 当时间 t 更接近 T 时, 策略会变得更加保守: 当 $X_t=1$ 时, 价格在下一个时间段更可能上升, 因此投资者会希望此时以最优卖出价格下单或者购买市场订单, 该决策取决于 $T-t$, 且在不购买任何其他订单、收益更低的情况下; 当 $X_t=-1$ 时, 价格会在下一个时间段更可能下降, 因此投资者会更加激进, 或者直接不下单, 或者以最优买入价买进, 该策略也取决于 $T-t$. 根据上述公式, t_1^* 和 t_2^* 也可能会是负的.

6.5.2 连续时间马尔可夫链模型

对于最优配置问题, Hult 和 Kiessling (2010) 提出了一个对于 LOB 的连续时间马尔可夫模型. 论文中考虑了无限时间情形 $T=\infty$. 利用 LOB 的马尔可夫排队模型, 论文利用马尔可夫链中的势论 (参见附录 B) 提出了一种计算订单被执行的概率的方法. 论文证明了最优配置策略的存在性, 并提出了一个价值迭代方法来计算执行概率.

Hult 和 Kiessling 排队论模型 定义 $\pi^1<\pi^2<\cdots<\pi^d$ 为 LOB 中的所有价格水平, 定义 $X_t=(X_t^1,\ldots,X_t^d)$ 为 $t>0$ 时买入订单的量 (取值为负时) 或卖出订单的量 (取值为正时), 其中 $X_t^j\in\mathbb{Z}$. 定义 $\mathbb{S}\subset\mathbb{Z}^d$ 为马尔可夫链的状态空间. 对于任意 $x\in\mathbb{S}$, 定义 $j_B(x)=\max\{j:x^j<0\}$ 为最高买入价, $j_A(x)=\min\{j:x^j>0\}$ 为最低卖出价. 则对于 $j_B<j<j_A$, 满足 $x^j=0$, 其中 j_A-j_B 为价差. 定义 $\boldsymbol{e}^j=(0,0,\ldots,1,0,\ldots,0)$ 为 $\mathbb{Z}^d$ 中的一个向量, 其中

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

第 j 个元素为 1. 简记 $j_B(x)$ 为 j_B, 类似地简记 j_A, 不同订单的可能状态变化列表如下:

- 在价格 j 上规模为 k 的买入限价订单: $x \mapsto x - k\boldsymbol{e}^j$, $j \leqslant j_A$.
- 在价格 j 上规模为 k 的卖出限价订单: $x \mapsto x + k\boldsymbol{e}^j$, $j \geqslant j_B$.
- 在价格高于 j_A 上规模为 k 的买入市场订单: $x \mapsto x - k\boldsymbol{e}^{j_A}$, 注意到当 $k > x^{j_A}$ 时, 卖出订单不会被执行.
- 在价格高于 j_B 上规模为 k 的卖出市场订单: $x \mapsto x + k\boldsymbol{e}^{j_B}$, 注意到当 $k > |x^{j_B}|$ 时, 买入订单不会被执行.
- 在价格 j 上规模为 k 的取消的买入订单: $x \mapsto x + k\boldsymbol{e}^j$, $j \leqslant j_B$, $1 \leqslant k \leqslant |x^j|$.
- 在价格 j 上规模为 k 的取消的卖出订单: $x \mapsto x - k\boldsymbol{e}^j$, $j \geqslant j_A$, $1 \leqslant k \leqslant x^j$.

由于转换概率矩阵非常大, 因此很难将马尔可夫链应用于 LOB 上面. Hult 和 Kiessling 使用了一个可行的参数化方法, 从而对数据进行校正, 并做了如下的假设:

(C1) 对于价格与最优卖出 (买入) 价相差 i 个单位的限价买入 (卖出) 订单, 其到达强度为 $\lambda_L^B(i)$ ($\lambda_L^A(i)$).

(C2) 市场买入 (卖出) 订单的到达强度为 λ_M^B (λ_M^A).

(C3) 限价订单和市场订单规模的分布函数 $(p_k)_{k\geqslant 1}$ 和 $(q_k)_{k\geqslant 1}$ 如下:

$$p_k = (e^{\alpha} - 1)e^{-\alpha k}, \quad q_k = (e^{\beta} - 1)e^{-\beta k}.$$

(C4) 假设取消的订单的规模为 1, 且与最优卖出 (买入) 价相差 i 个单位的单位规模的买入 (卖出) 订单的取消强度为 $\lambda_C^B(i)$ ($\lambda_C^A(i)$). 总体来说, 与最优卖出 (买入) 价相差 i 个单位的单位规模的买入 (卖出) 订单的取消强度与 $\lambda_C^B(i)|x^{j_A-i}|$ ($\lambda_C^A(i)|x^{j_B+i}|$) 成比例.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

根据上述假设, 生成矩阵 $\boldsymbol{Q}$ 由如下公式确定:

- 限价订单:

$$\begin{aligned} x &\mapsto x+\boldsymbol{e}^j k, \quad \text{当 } j>j_B, \quad \text{比率为 } p_k\lambda_L^A(j-j_B),\\ x &\mapsto x-\boldsymbol{e}^j k, \quad \text{当 } j<j_A, \quad \text{比率为 } p_k\lambda_L^B(j_A-j). \end{aligned}$$

- 规模为 $k\geqslant 2$ 的市场订单:

$$\begin{aligned} x &\mapsto x+k\boldsymbol{e}^{j_B}, \quad \text{比率为 } q_k\lambda_M^A,\\ x &\mapsto x-k\boldsymbol{e}^{j_A}, \quad \text{比率为 } q_k\lambda_M^B. \end{aligned}$$

- 取消订单:

$$\begin{aligned} x &\mapsto x-\boldsymbol{e}^j, \quad \text{当 } j>j_B, \quad \text{比率为 } \lambda_C^A(j-j_B)|x^j|,\\ x &\mapsto x+\boldsymbol{e}^j, \quad \text{当 } j<j_A, \quad \text{比率为 } \lambda_C^B(j-j_A)|x^j|. \end{aligned}$$

根据 X_t 的生成矩阵 $\boldsymbol{Q}=(Q_{xy})$ (其中 Q_{xy} 是由状态 $x=(x^1,\dots,x^d)$ 到状态 $y=(y^1,\dots,y^d)$ 的转移强度), 我们可以得到转移矩阵 $\boldsymbol{P}=(P_{xy})$, 其中 $P_{xx}=0$, 对于 $y\neq x$, $P_{xy}=\dfrac{Q_{xy}}{\sum_{z\neq x}Q_{xz}}$, 具体可参见附录 B. Hult 和 Kiessling 使用内在链计算了订单的执行概率以及最优策略, 如下所示:

订单执行概率 在研究一个在时刻 0 时的订单以价格 π^j 成交的概率时, 我们需要对于该订单在排队中的位置和整个 LOB 的状态进行分析. 因此, 在分析时需要将 LOB 的马尔可夫链性质和订单的位置相结合. 考虑 X_n 对于 LOB 的内在链. 定义 Y_n 为时刻 n 之前在价格水平 j_0 上包含交易者的订单之内的成交的买入订单的数量, 为了方便, 我们对于买入订单加上负号. 假设交易者会在价格水平 j_0 上交易一个限价买入订单. 初始条件为 $Y_0=x_0^{j_0}-1$, 且只能在订单当交易者的订单被执行或取消之前价格向上趋于 0. 注意到 $(X_n,Y_n)_{n\geqslant 0}$ 也是一个马尔可夫链, 其状态空间为 $\Omega=\mathbb{S}\times\{0,-1,\dots\}$, 转移矩阵为 $\boldsymbol{\Pi}$, 其中 $\Pi_{(x,y),(x',y')}=P_{xx'}p_{yy'}$ 且 $(p_{yy'})_{y,y'\in\{0,-1,\dots\}}$ 是马尔可夫链 $(Y_n)_{n\geqslant 0}$ 的转移矩阵.

为计算一个订单在价格水平变成 J_1 时而被取消之前, 以最优卖出价执行的概率, 定义 $\tau=\inf\{n:(X_n,Y_n)\in\partial D\}$, 其中 $\partial D=\{(x,y)\in\Omega: y=0$ 或 $j_A(x)=J_1\}$ 是 $D=\{(x,y)\in\Omega: y<0$ 且 $j_A(x)<J_1\}$ 的边界, 定义矩阵 $\boldsymbol{V}$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

和 $\boldsymbol{\Phi}$, 对于 $(x,y)\in\Omega$, (x,y) 项定义为

$$V(x,y)=\begin{cases}1, & \text{当 } y=0,\\ 0, & \text{其他,}\end{cases}$$

$$\Phi(x,y)=E\left[V(X_\tau,Y_\tau)I_{\{\tau<\infty\}}|(X_0,Y_0)=(x,y)\right].$$

根据势论 (参见附录 B), Φ 满足如下的线性方程:

$$\boldsymbol{\Phi}=\begin{cases}\boldsymbol{\Pi\Phi}, & \text{在 } D \text{ 中,}\\ \boldsymbol{V}, & \text{在 } \partial D \text{ 上.}\end{cases}$$

在无限时间区间内购买一股的最优策略 在无限时间 $T=\infty$ 内, 最优策略是一个稳定策略. 假设订单簿的初始空间为 x_0. 交易者希望买入一股的股票, 且在价格 π^j 上下单. 在 LOB 中的 n 个订单后, 定义 $j_n=j$ 为此时限价订单的价格水平, Y_n 为此时在价格水平 j_n 上包含交易者的订单在内的剩余的订单数量. 此时可以定义离散的马尔可夫链 $(\tilde{X}_n,Y_n,j_n)$. Hult 和 Kiessling (2010) 假设存在最优价格水平 J_0 和最差价格水平 J_1, 满足 $J_0<j_A(x_0)<J_1$. 交易者并不会在价格水平低于 J_0 时进行交易. 而 J_1 是最差的买入或止损的价格水平. 在满足 $j_A(\tilde{X}_n)=J_1$ 时, 交易者会立即取消限价买入订单, 而在价格水平为 J_1 时交易市场订单. 此时的状态空间为 $\Omega=\mathbb{S}\times\{\ldots,-2,-1,0\}\times\{J_0,\ldots,J_1\}$. 交易者的可行集取决于当时的状态 $\boldsymbol{s}=(x,y,j)$, 交易者在 $y<0$ 时有三种选择:

1. 不进行交易, 等待市场状态的变化.
2. 取消限价订单, 并在价格水平为 $j_A(x)$ 时交易市场订单.
3. 取消当前的限价买入订单, 并在价格水平为满足 $J_0\leqslant j'<j_A(x)$ 的任意价格水平 j' 时, 下单一个新的限价买入订单. 该行为会导致市场状态变化为 $j_n=j'$, $\tilde{X}_n=x+\boldsymbol{e}^j-\boldsymbol{e}^{j'}$ 和 $Y_n=x^{j'}-1$.

在一个给定的状态 $\boldsymbol{s}=(x,y,j)$, 其中 $y<0$ 和 $j_A(x)<J$, 行为集定义为 $C(x,y,j)=\{0,J_0,\ldots,j_A(x)-1\}$. 因此 $a=0$ 表示交易者选择等待, 而 $a=j'$, $J_0\leqslant j'<j_A(x)$ 表示取消当前的订单, 并在价格水平为 j' 时下一个新的订单. 继续当前行为的成本为 0. 当满足 $y=0$ 或 $j_A(x)=J_1$ 时, 则有 $C(\boldsymbol{s})=\varnothing$, 此时只能停止当前的交易. 定义停止交易的集合为 $\boldsymbol{T}=\{-2,-1\}$. 当满足 $y<0$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

时, 唯一可行的停止交易的情况是 -1. 当满足 $y=0$ 时, 由于限价订单会被执行, 因此马尔可夫链总会终止, 我们将该情况记为 -2. 此时的 Bellman 方程 (参见附录 B) 可以简化如下:

$$V_\infty(\boldsymbol{s})=\min\left(\min_{a\in C(\boldsymbol{s})}\sum_{\boldsymbol{s}'\in\Omega}P_{\boldsymbol{s}\boldsymbol{s}'}(a)V_\infty(\boldsymbol{s}'),\ \min_{a\in\boldsymbol{T}(\boldsymbol{s})}v_{\boldsymbol{T}}(\boldsymbol{s},a)\right),\tag{6.66}$$

其中 $j_A(x)<J_1$, $y<0$, $v_{\boldsymbol{T}}(\boldsymbol{s},-2)=\pi^{j_0}$, $v_{\boldsymbol{T}}(\boldsymbol{s},-1)=\pi^{j_A(x)}$, 且有

$$V_\infty(x,y,j)=\begin{cases}\pi^{J_1}, & \text{当 } j_A(x)=J_1,\ y<0,\\ \pi^{j}, & \text{当 } y=0.\end{cases}\tag{6.67}$$

在公式 (6.66) 中, $P_{\boldsymbol{s}\boldsymbol{s}'}(a)$ 表示当执行行动 a 时 (在状态转移之前), 从状态 $\boldsymbol{s}$ 到状态 $\boldsymbol{s}'$ 的转移概率 (可控马尔可夫链可以参见附录 B 中的 B.3 节).

6.6 补充材料和习题

1. **VWAP 和 TWAP** 容量加权的平均价格 (VWAP) 和时间加权的平均价格 (TWAP) 是两个常用的衡量交易表现的指标, 用于评估执行的算法. 本节与 6.1 节使用同样的符号, 将离散时间的两个指标定义为

$$\mathrm{VWAP}=\frac{\sum_{t=1}^T P_tN_t}{\sum_{t=1}^T N_t},\quad \mathrm{TWAP}=\frac{1}{T}\sum_{t=1}^T P_t,$$

连续时间的指标定义为

$$\mathrm{VWAP}=\frac{\int_0^T P_t\mathrm{d}Y_t}{\int_0^T \mathrm{d}Y_t},\quad \mathrm{TWAP}=\frac{1}{T}\int_0^T P_t\mathrm{d}t.$$

考虑一个包含线性和附加价格影响的随机游走模型 (6.1). 正如 6.1 节中讨论过, Bertsimas 和 Lo (1998) 发现执行问题 (6.2) 的最优解是一个在时间上均匀分布的执行策略, 即有 $N_t^*=\tilde{N}/T$, 此时两种指标的结果是一样的.

$$\mathrm{VWAP}_{N^*}=\frac{\sum_{t=1}^T P_t}{T}=\mathrm{TWAP}_{N^*}.$$

同时, 根据 (6.13),

$$E(\mathrm{VWAP}_{N^*})=\frac{V_1(P_0,Y_1)}{\tilde{N}}=P_0+\frac{\theta}{2}\left(1+\frac{1}{T}\right),$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

该值随着“影响敏感度”参数 θ 线性增长. 另外, 当 θ 很小时, $E(\mathrm{VWAP}_{N^*})$ 的值为初始价格 P_0.

习题 将 Bertsimas 和 Lo 中提到的最优执行规则 N^* 与 Obizhaeva 和 Wang (OW) 在模型 (6.1) 下的最优执行规则相比, 并以 VWAP 和 TWAP 为指标比较. 离散时间内的 OW 规则可以通过将 6.3.4 节中连续时间模型解中的 $(N-1)(1-e^{-\rho\tau})$ 替代为 $\rho(T-2)$, 证明对于 $t=2,\dots,T-1$,

$$\mathrm{OW}_1^* = \mathrm{OW}_T^* = \frac{\tilde{N}}{\rho(T-2)+2} \quad \text{且} \quad \mathrm{OW}_t^* = \frac{\rho\tilde{N}}{\rho(T-2)+2}.$$

请证明通常情况下 $\mathrm{VWAP}_{\mathrm{OW}} \neq \mathrm{TWAP}_{\mathrm{OW}}$. 同时, 请对于不同的 ρ 值, 分析 $E(\mathrm{VWAP}_{\mathrm{OW}})$ 和 $E(\mathrm{VWAP}_{N^*})$ 的关系.

2. 另一个常用的表现指标是容量比例 (PoV), 该指标使用在一定时间段内实际交易的容量作为基准指标, 并结合一段时间段内的参与率 β 来分析. 举例来说, 如果还有 Q 股将被购买, PoV 计算时间段 $(t-\Delta t, t)$ 内交易的容量 V, 并执行 $\min(Q, \beta V)$.

3. **Bertsimas 和 Lo 的最优执行问题** 通过直接改写求解优化问题 (6.2), 求解时通过将目标函数写成 $N_1, N_2, \dots, N_T$ 的二次函数的形式直接进行求解.

4. 基本的随机模型 (6.1) 的一个变化是加入了价格影响项, 并考虑了市场情形/私有信息项 X_t, 其服从 AR(1) 过程:

$$\begin{aligned} P_t &= P_{t-1} + \theta N_t + \gamma X_t + \epsilon_t, \\ X_t &= \rho X_{t-1} + \eta_t, \end{aligned}$$

其中 $-1<\rho<1$, $\{\epsilon_t\}_{t\geqslant 1}$ 和 $\{\eta_t\}_{t\geqslant 1}$ 是独立同分布的序列, 均值为 0, 方差为有限. 请证明最优交易策略是确定性的, 即与 P_t 无关, 并具体证明如下的最优解 $N_{T-k}^* = \delta_{Y,k} Y_{T-k} + \delta_{X,k} X_{T-k}$, 其中

$$\delta_{Y,k} = \frac{1}{1+k}, \quad \delta_{X,k} = \frac{\rho b_{k-1}}{2a_{k-1}}, \quad a_k = \frac{\theta}{2}\left(1+\frac{1}{k+1}\right),$$
$$b_k = \gamma + \frac{\theta\rho b_{k-1}}{2a_{k-1}} \ (>0, \text{ 当 } \theta,\rho,\gamma>0), \quad b_0 = \gamma.$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

5. 6.1 节中的业绩评价指标是 Bertsimas 和 Lo (1998) 中的期望成本与 Almgren 和 Chriss (2001) 中的均值方差准则. 在文献中提到的其他的准则包括 Schied 和 Schöneborn (2009) 中提到的期望效用与 Forsyth 等 (2012) 中提到的 GBM 模型的均值二次变差. Gatheral 和 Schied (2013) 则提出使用风险中性期望收益或者成本优化准则来作为评价指标, 尤其是在市场情形与投资者的偏好独立的情况下更适用. Schied 等 (2010) 利用 Euler-Lagrange 模型证明最小化成本与最小化交易策略中的价格影响对于一系列的模型是等价的.

6. **关于 Almgren 和 Chriss 模型的习题** 假设我们将 6.1.2 节中 Almgren 和 Chriss 中的方差项替换为其他的风险测度. 考虑所有可行的交易策略求解
$$\inf_{Y_t\in\mathcal{A}} E\left[\eta\int_0^T \dot{Y}^2(t)dt + \lambda\sigma^2\int_0^T Y(t)dt\right].$$
同时证明, 最优清算策略为如下形式:
$$Y^*(t) = (\tilde{N} - Ct)\left(1 - \frac{t}{T}\right),$$
其中 $C = \dfrac{\lambda\sigma^2 T}{4\eta}$.

7. **关于 Gatheral 和 Schied 模型的习题** 根据 Gatheral 和 Schied (2011), 分析包含一个时间平均的风险价值 (VaR) 最优执行问题的 GBM 形式. 具体来说,
$$C(T,\tilde{N},P_0) = \inf_{\dot{Y}\in\mathcal{A}} E\left[\int_0^T \dot{Y}^2(t)dt + \lambda\int_0^T P(t)Y(t)dt\right],$$
约束条件为 $Y(0) = \tilde{N}$, $Y(T) = 0$, 其中 $dP(t) = \sigma P(t)dW(t)$. 求解 $C(t,Y,P)$ 的 HJB 方程:
$$C_t = \frac{1}{2}\sigma^2 P^2 C_{PP} + \lambda PY + \inf_{\dot{Y}}(\dot{Y}^2 - \dot{Y}C_Y).$$
令 $\tilde{N} = 0$ 和 $T \to 0$, 求解 HJB 的边界条件, 并证明唯一的最优执行策略为
$$Y_t^* = \frac{T-t}{T}\left[\tilde{N} - \frac{\lambda T}{4}\int_0^t P_u du\right].$$

8. 对于 6.2.2 节中的乘数价格影响模型, Guo 和 Zervos (2015, 引理 2) 在不使用 HJB 方程的情况下, 发现最优解具有以下特性:

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

- 如果 $\mu \leqslant \delta$ 且 $C_{\mathrm{s}} = 0$, 则最优策略是时刻 0 时卖出所有的股票.
- 如果 $\mu \leqslant \delta$, 则最优策略中一定不包含买入股票.
- 如果 $\mu > \delta$, 则最优策略是在时刻 T 时卖出所有的股票.

9. 在分析均值回归模型的最优配置问题时,6.4 节使用马尔可夫随机游走模型 A_t 的特性和其最小值 $M_t = \min\limits_{0\leqslant u\leqslant t} A_u, 0 \leqslant t \leqslant T$ 来求解. 特别地, Guo 等 (2013) 证明了包含独立同分布的增量的随机游走模型和其最小值的拓展情况:

   (a) 如果 $(T-k)/2 \in \mathbb{Z}$ 且 $k > 0$, 则有 $P(M_T = -k) = P(A_T = -k)$. 该结果称为马尔可夫随机游走模型的部分反射准则.

   (b)

$$P(A_T = k) = \begin{cases} \displaystyle\sum_{i=1}^{T-|k|} p^{T-i-1}(1-p)^i L_{T,k}^i, & \text{当 } \dfrac{T+k}{2} \in \mathbb{Z} \text{ 且 } |k| \leqslant T, \\ 0, & \text{其他,} \end{cases}$$

   其中

$$\begin{aligned} L_{T,k}^i = (1-\bar{p}) &\binom{\frac{T+k}{2}-1}{\left\lfloor \frac{i+1}{2} \right\rfloor - 1} \binom{\frac{T-k}{2}-1}{\left\lfloor \frac{i+2}{2} \right\rfloor - 1} \\ &+ \bar{p} \binom{\frac{T+k}{2}-1}{\left\lfloor \frac{i+2}{2} \right\rfloor - 1} \binom{\frac{T-k}{2}-1}{\left\lfloor \frac{i+1}{2} \right\rfloor - 1}. \end{aligned}$$

   (c) 对于 $k = 1, \ldots, T$, $P(M_T = -k)$ 是关于 k 的一个递减函数.

10. **关于价格操纵的评论** 尽管不存在价格操纵是短期 (如几秒或几分钟) 模型的一个非常好的特点, 但是对于一个长期 (如几周或几个星期) 模型而言, 该条件有一定的局限性. 实际上, 在任何不包含价格影响的模型中, 对于在时刻 0 买入 (或者做空) 一股, 并随后卖出 (或者买回) 的策略而言, 当股票价格是一个下鞅过程 (或者上鞅过程) 时, 如包含严格为正 (或者为负) 的漂移项的几何布朗运动模型, 即存在价格操纵. 具体可参见 Guo

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

和 Zervos (2015) 与 Gatheral 和 Schied (2011) 以及后续的论文对于价格操纵的讨论.

习题 比较无套利情况下的衍生品定价模型和不存在价格操纵情况下的定价模型, 后者在数量金融领域中不太常见且取决于优化问题中的表现评价指标.

11. 除了最优策略以外, Hult 和 Kiessling (2010) 也讨论了一系列其他的启发性配置策略. 具体而言, 论文讨论了如下的简单策略, 该策略在价格水平为 J_0 时买入一单位的股票, 并在价格水平为 J_1 时取消限价订单, 论文指出: 由于该订单需要支付 π^{J_0} (如果订单被执行) 或者 π^{J_1}(如果订单被取消, 并在该价格水平下买入一个市场订单), "只需求得限价订单被执行的概率就可以计算得到期望买入价格". 论文指出, 利用势论可以将计算执行概率简化成求解一系列的线性方程组. 更多细节可以参见附录 B.

12. **将最优执行问题与最优配置问题相结合** Guéant 等 (2012) 将最优交易流程设计 (即最优执行问题) 与最优配置限价问题相结合, 从而在时刻 T 之前, 将 $\tilde{N}$ 股股票进行清算的问题. 存货为 $Y_t = \tilde{N} - \xi_t^a$, 其中 ξ^a 是计算卖出的股票数的一个跳跃过程. 假设跳跃的幅度为 1, 且强度函数为 $\lambda(\delta^a)$, 其取决于交易者报价 $X_t^a = X_t^0 + \delta_t^a$, 其中 X_t^0 为有效价格过程, 该过程中不存在交易摩擦, 服从一个带漂移项的布朗运动过程: $dX_t^0 = \mu dt + \sigma dB_t$. 与 6.2.1 节中的模型类似, 此时并不能直接得到 ξ^a 过程. Guéant 等 (2012) 假设 $\lambda(\delta^a) = A\exp(-\kappa\delta^a)$, 并分析了如下的最大化指数效用问题:

$$\sup_{\delta^a \in \mathcal{A}} E\left(-\exp\left\{-\gamma\left[X_T + Y_{T-}(S_T - b)\right]\right\}\right), \tag{6.68}$$

其中 γ 是 CARA 效用函数 $U(z) = 1 - e^{-\gamma z}$ 里的风险厌恶系数. $\mathcal{A}$ 是可行交易策略集 (一个在 $[0,T]$ 上的存在下界的可预测过程). b 是在时刻 T 时将所有剩余的股份进行清算的成本, $dX_t = X_t^a d\xi_t^a$ 是交易者在时刻 t 时卖出股票得到的资金 X_t 所服从的过程. 论文利用动态规划准则来求解价值方程 $u(s,x,y,t)$ 的 HJB 方程, 即将 (6.68) 式中的 E 替换为 $(X_t^0, X_t, Y_t) = (s,x,y)$ 时的条件期望公式. HJB 方程为

$$u_t + \mu u_s + \frac{1}{2}\sigma^2 u_{ss} + \max_{\delta^a} \lambda^a(\delta^a)\left[u(s, x+s+\delta^a, y-1, t) - u(s,x,y,t)\right] = 0,$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其边界条件为

$$u(s,x,0,t) = -\exp(-\gamma x), \quad u(s,x,y,T) = -\exp\{-\gamma[x+y(s-b)]\}.$$

论文通过变量转换的方法将 PDE 转换为一系列的 ODE 方程来求解 HJB 方程. 在求解时是根据边界条件来 "猜" 解的形式. 在得到了正确的解的形式后, 论文利用常见的验证方法来验证该结果是否为最优解. 论文的解的形式为 $u(s,x,y,t) = -\exp[-\gamma(x+ys)][v_y(t)]^{-\gamma/\kappa}$, 其中 v_y 服从如下的线性 ODE 过程:

$$\frac{dv_y(t)}{dt} = (\alpha y^2 - \beta y)v_y(t) - \eta v_{y-1}(t), \tag{6.69}$$

其中 $v_y(T) = \exp(-\kappa y b)$, $v_0(t) = 1$, $\alpha = \gamma\sigma^2/2$, $\eta = A[1+(\gamma/\kappa)]^{-(1+\gamma/\kappa)}$. 最优的 δ_*^a 由下式求得:

$$\delta_*^a(t,y) = \frac{1}{\kappa}\ln\left(\frac{v_y(t)}{v_{y-1}(t)}\right) + \frac{1}{\gamma}\ln\left(1+\frac{\gamma}{\kappa}\right). \tag{6.70}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

第七章　做市和智能订单路径

做市商和任何金融交易所一样, 都有着悠久的历史. 他们可以追溯到 18 世纪末, 当时在伦敦被称为股票经纪人. 除了这个几乎被遗忘的名字之外, 他们更常被称为纳斯达克的交易商和纽交所的专家. 他们的主要商业模式很简单: 从卖方那里购买证券, 目的是在相对较短的时间内以更高的价格转售给买方. 交易所对做市商规定了最低限度的义务, 例如, 最小价差、最低报价额、做市商必须报价的每日最低百分比, 以及做市商报价的最短有效时间; 详细信息请参见 Weaver (2012).

从经济学角度看, 一个做市商与一个靠近很多家庭且 24 小时营业的零售便利店相似. 没有这样的商店, 想要在半夜喝牛奶的消费者可能需要等到下一次牛奶从农场送来. 此外, 每周提供 100 加仑牛奶的农民能够直接从商店得到报酬, 并不需要等到牛奶被消费者买完. 当然, 如果附近有很多农场能够按照不同时间段提供牛奶, 而且容器大小齐全, 那么对售卖牛奶的便利店的需求就会减少. 否则, 便利店就能直接解决农民和消费者的问题. 如果流动性被定义为资产与货币相互转换的难易程度, 那么便利店无疑向农民和消费者提供了流动性. 同样地, 在股票市场中, 做市是为了向市场中的证券买卖双方提供流动性. 当做市商从前一个证券获得的利润过低时, 则可以转到其他证券. 此外, 当证券价格波动性过强时, 该证券的做市商可以选择放弃该证券, 而在价格波动恢复到某种可接受的水平时再重新开始交易. 由于做市商在预算限制和未来供求预测的不确定性下运作, 因此必须妥善管理好存货. 另外, 由于像伦敦交易所和纳斯达克等采用了竞争性的交易系统, 做市商还需要面对来自其他做市商的竞

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

争. 交易所给予做市商的补偿可以是不同的形式. 例如, 纽交所提供回扣和 (或) 更低的费用, 而德意志交易所和多伦多交易所则分别提供信息优势和时间优先权.

本章首先给出了与做市相关的随机优化问题的概述. 7.1 节介绍了 Ho 和 Stoll (1981) 的最优化模型与 Avellaneda 和 Stoikov (2008) 的最新进展. 7.2 节介绍了 Guéant 等 (2013) 的研究结论, 该结论中考虑了存货约束和后续拓展的影响. 在 7.3 节中, 我们介绍了 Guilbaud 和 Pham (2013) 的另一个推广, 将限制和市场订单纳入做市的考量. 他们的工作使用了脉冲控制, 这属于随机控制领域, 我们没有在前面的章节中考虑, 并在此进行了介绍. 7.4 节概述了零散的初级市场和暗池的智能订单路径. 7.5–7.7 节考虑了交易所和暗池中的最优订单拆分问题. 7.8 节给出补充和评论, 7.9 节是练习题.

7.1 Ho 和 Stoll 模型以及 Avellanedo-Stoikov 决策

假设做市商的参考价格过程 S_t 是布朗运动, 波动率为 σ. 假设 CARA 效用[①] 函数为 $1-e^{-\gamma z}$, 代理人持有 q 份股票直至到期 T 的期望效用为 $1-u(x,s,q,t)$, 并且

$$\begin{aligned} u(x,s,q,t) &= E_t\left\{[-\exp(-\gamma(x+qS_T)]\right\} \\ &= -\exp(-\gamma x)\exp(-\gamma q s)\exp\left[\gamma^2 q^2\sigma^2\frac{T-t}{2}\right], \end{aligned} \tag{7.1}$$

其中 x 是初始财富, s 是初始每股价格, γ 是连续的折现率. 注意到代理人的预定 (无差异) 买价 r_t^b 满足 $u(x-r_t^b(s,q),s,q+1,t)=u(x,s,q,t)$, 因此满足 $r_t^b(s,q)=s+(-1-2q)\gamma\sigma^2\dfrac{T-t}{2}$. 类似地, 求解 $u(x+r_t^a(s,q),s,q-1,t)=u(x,s,q,t)$ 即可得到预定卖价 $r_t^a(s,q)=s+(1-2q)\gamma\sigma^2\dfrac{T-t}{2}$.

代理人的做市商策略是一个随机控制问题, 涉及两个控制 δ_t^a 和 δ_t^b 以及三个状态变量: S_t、现金和存货. 代理人的买价 s_t^b, 卖价 s_t^a, 并且与代理人订单相对的市场买入 (或卖出) 订单以泊松率 $\lambda^a(\delta_t^a)$ (或 $\lambda^b(\delta_t^b)$) 出现, 其中 $\delta_t^b=S_t-s_t^b$ 和 $\delta_t^a=s_t^a-S_t$ 是利差, λ^a 和 λ^b 是递减函数. 令 N_t^a 和 N_t^b 分别为代理人在时刻 t 出售和购买的份额. 则两者都服从泊松过程, 各自强度分别为 $\lambda^a(\delta^a)$ 和 $\lambda^b(\delta^b)$. 代理人持有的现金 X_t 满足如下方程:

$$dX_t = s_t^a dN_t^a - s_t^b dN_t^b. \tag{7.2}$$

① CARA 代表 "固定绝对风险厌恶"; 3.3.1 节的脚注 1 介绍了绝对风险厌恶. 指数效用函数 $U(z)=1-e^{-\gamma z}$ 有绝对风险厌恶系数 γ, 其不随 z 而改变, 因此是固定的.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

代理人在时刻 t 的库存 (持有的份额) 为 $q_t = N_t^b - N_t^a$. 因此, 代理人的最终财富为 $X_T + q_T S_T$. 优化问题的目标是选择非预期的过程 $\delta_a = \{\delta_t^a : 1 \leqslant t \leqslant T\}$ 和 $\delta_b = \{\delta_t^b : 1 \leqslant t \leqslant T\}$ 以最大化指数效用, 因此价值函数 (减去 1 后) 等价于

$$V(s,x,q,t) = \max_{\delta^a,\delta^b} E\left\{-\exp\left[-\gamma(X_T + q_T S_T)\right] | (S_t, X_t, q_t) = (s,x,q)\right\}. \quad (7.3)$$

该问题最初由 Ho 和 Stoll (1981) 提出.

Avellaneda 和 Stoikov (2008) 选择 $\lambda_a(\delta) = \lambda_b(\delta) = Ae^{-\kappa\delta}$ 做进一步讨论, 其中 A 为正的常数项并且 κ 在统计意义上刻画了资产的流动性①. 在使用 CARA 效用函数时, 价值函数简化为

$$V(s,x,q,t) = -\exp\left\{-\gamma[x + v(t,s,q)]\right\} \quad (7.4)$$

且 v 的 HJB 方程为

$$\begin{aligned}
&\max_{\delta^a} \gamma^{-1} A e^{-\kappa\delta^a} \left(1 - \exp\left\{-\gamma[s + \delta^a + v(t,s,q-1) - v(t,s,q)]\right\}\right) \\
&+ \max_{\delta^b} \gamma^{-1} A e^{-\kappa\delta^b} \left(1 - \exp\left\{-\gamma\left[-s + \delta^b + v(t,s,q+1) - v(t,s,q)\right]\right\}\right) \\
&+ v_t + \sigma^2 \frac{(v_{ss} - \gamma v_s^2)}{2} = 0,
\end{aligned} \quad (7.5)$$

边界条件为 $v(T,s,q) = qs$. (7.5) 式中的极大值点为

$$\begin{aligned}
\delta_*^a &= \gamma^{-1}\log\left(1 + \frac{\gamma}{\kappa}\right) - s - v(t,s,q-1) + v(t,s,q), \\
\delta_*^b &= \gamma^{-1}\log\left(1 + \frac{\gamma}{\kappa}\right) + s - v(t,s,q+1) + v(t,s,q).
\end{aligned} \quad (7.6)$$

虽然上面已经对于 v 的 PDE 方程 (7.5)–(7.6) 在边界条件为 $v(T,s,q) = qs$ 时的情形进行了简化, 但是 PDE 方程仍然是高度非线性的并且涉及连续变量 t 和 s, 以及一个离散变量 q. Avellaneda 和 Stoikov (2008) 利用渐近扩展方法对 v 的最优策略进行了近似.

7.2 HJB 方程的求解和后续的拓展

Guéant 等 (2013) 将 (7.5) 和 (7.6) 转化为 ODE 的一个系统. 使用的技术类似于 6.6 节补充材料 12 中求解另一个问题的方法. 在求解时, 将有限库存限

① 注意 Guéant 等 (2012) 在 6.6 节的附注 12 中使用了相同的选择, 当其与指数效用耦合时具有易于处理的优点.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

制 $|q| \leqslant Q$ 这一额外的约束纳入了分析, 此时 ODE 系统是有限的并且求得的显式解如下:

$$\exp\{-\gamma v(t,s,q)\} = \exp\{-\gamma q s\}\left(v_q(t)\right)^{-\gamma/\kappa}, \tag{7.7}$$

其中 $(v_{-Q}(t),\ldots,v_{-1}(T),v_0(T),v_1(T),\ldots,v_Q(t))^\top$ 可以由 $e^{-(T-t)\boldsymbol{M}}(1,\ldots,1)^\top$ 显式给出. 令 $\beta = \kappa\gamma\dfrac{\sigma^2}{2}$ 和 $\eta = A\left(1+\dfrac{\gamma}{\kappa}\right)^{-\left(1+\frac{\kappa}{\gamma}\right)}$, 则有

$$\boldsymbol{M} = \begin{pmatrix} \beta Q^2 & -\eta & 0 & \cdots & \cdots & \cdots & 0 \\ -\eta & \beta(Q-1)^2 & -\eta & 0 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 & -\eta & \beta(Q-1)^2 & -\eta \\ 0 & \cdots & \cdots & \cdots & 0 & -\eta & \beta Q^2 \end{pmatrix},$$

其中 $e^{\boldsymbol{A}} = \sum_{n=0}^{\infty}\dfrac{\boldsymbol{A}^n}{n!}$ 是矩阵 $\boldsymbol{A}$ 的指数项.

其他参考价格模型的拓展 前面的 HJB 方程及其解可以很容易地扩展到包含漂移项布朗运动形式的参考价格上, 如 $dS_t = \mu dt + \sigma dB_t$. Fodra 和 Labadie (2012) 与 Ahuja 等 (2016) 将模型拓展到如下的 Ornstein-Uhlenbeck (OU) 过程中:

$$dS_t = \alpha(\mu - S_t)dt + \sigma dB_t. \tag{7.8}$$

由于 OU 过程是均值回复的, Fodra 和 Labadie 分析了参考价格与其稳态平均值相差很大的情况, 并只考虑一个 (或部分) 均值回复周期. 而 Ahuja 等则考虑了更长时间 T 内包含多个均值回复周期的情形. 在参考价格中考虑 $\alpha(\mu - S_t)$ 项时, 会对于 (7.5) 式进行明显修正, (7.5) 式的右侧目前存在一个明显的被加数 $\alpha(\mu - s)$. 修正后, 边界条件依旧为 $v(T,q,s) = qs$. 而包含 $\alpha(\mu - s)(\alpha > 0)$ 的高度非线性的偏微分方程 (7.5) 和 (7.6) 式都不能被简化成有限的常微分方程系统. Ahuja 等 (2016) 利用两种数值方法来解决该问题并计算得到最优策略. 一种是隐式有限差分方法, 另一种是将偏微分方程分解为两个偏微分方程的分步方法, 其中一个偏微分方程也可以进一步简化为有限库存约束下的有限常微分方程系统.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

7.3 包含限制和市场订单的脉冲控制

在如今的金融市场上, 限价订单交易者根据报价的离散 (分笔) 约束来设定其订单价格, 而市场则决定订单的执行速度. 前一节中所描述的关于做市的随机控制方法主要集中于最大化其终期财富的效用函数的限价做市商. 该模型假定市场订单的到达服从双重随机点过程, 其密度函数形式为 $A\exp(-\kappa\delta_t)$, 其中 δ_t 是限价订单价格与基础参考价格之间的价差. 随机控制问题在没有使用来自 LOB 的关于实际到达模式的信息并且甚至忽略 LOB 的离散价格水平的情况下, 对于在该市场订单到达假设下做市商设置限制订单的价格进行了建模.

Guilbaud 和 Pham (2013) 提出了与做市相关的一个更灵活的随机控制问题, 建模时同时使用限价和市场订单作为控制, 并使用 LOB 数据作为信息集. 因此, 代理商可以根据最优买入价格和卖出价格进行买入和卖出, 也可以在最优买入价上方和最优卖出价格下方一个价位进行买入和卖出. 此外, 对于中间价 P_t 和买卖价差 D_t, 他们假设如下的连续时间马尔可夫链模型.

中间价和买卖价差的动态模型

- 在一个给定无穷小生成元的状态空间中, 股票的中间价格 P_t 是一个外生的马尔可夫过程.
- LOB 买卖差价 D_t 的建模是基于一个有限状态空间 $\mathbb{S}$ 上的连续时间马尔可夫链, 该有限状态空间由最小报价单位 δ 的倍数组成, 并且服从刻度时钟的泊松过程, 因此 $D_t = \Delta_{N_t}$.
- 当市场中的买入/卖出指令影响当前最优买入与卖出价格之间的价差时, 刻度时钟对应于非齐次泊松过程 N_t, 其确定性的密度函数为 $\lambda(t)$, 时间是随机的.
- Δ_n, $n = 0, 1, 2, \ldots$ 是 $\mathbb{S}$ 上的状态为 $i\delta$, $i = 1, \cdots, M$ 的马尔可夫链, 转移概率矩阵为 $\rho_{ij} = P[\Delta_{n+1} = j\delta | \Delta_n = i\delta]$, 满足 $\rho_{ii} = 0$.
- P_t 和 D_t, $t \geqslant 0$ 相互独立.

7.3.1 做市商的脉冲控制

限价订单策略 做市商可以随时以最优买入/卖出价格提交买入/卖出指令, 也可以以比当前最优买入/卖出价更高/更低的价格下单. 这四个价格分别

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

被定义为 $P_t^b = P_t - \dfrac{D_t}{2}$, $P_t^a = P_t + \dfrac{D_t}{2}$, $P_t^b - l_t$, $P_t^a + l_t$, 其中 $l_t = 0$ 或 δ 表示代理人的选择. 此外, 买入订单的股票数量 q_t^b 和卖出订单的股票数量 q_t^a 同样是被选择的. 因此, 做市商在时间 t 点的控制是可预测的过程

$$\boldsymbol{L}_t = (q_t^b, q_t^a, l_t), \quad t \geqslant 0. \tag{7.9}$$

对限价订单策略 $\boldsymbol{L}_t$, 持有的现金 X_t 和股票数量 Y_t 是由下式给定的:

$$\begin{aligned} dY_t &= q_t^b dN_t^b - q_t^a dN_t^a, \\ dX_t &= -(P_t^b - l_t) q_t^b dN_t^b + (P_t^a + l_t) q_t^a dN_t^a. \end{aligned} \tag{7.10}$$

市价订单策略 市价订单策略是通过在停止时间 τ_i 发生的脉冲控制来建模的, 并且包括股票数量 $\zeta_i > 0$ (对于买入订单) 或者 $\zeta_i < 0$ (对于卖出订单单), ζ_i 是 $\mathcal{F}_{\tau_i}$-可测的. 具体而言, 市价订单策略可以被一系列脉冲控制表示为

$$\mathcal{C} = \{(\tau_i, \zeta_i) : i = 1, 2, \dots\}.$$

提交的市价订单提供流动性, 且通常需要支付交易费, Guilbaud 和 Pham (2013) 假定交易费固定为 $\phi > 0$. 因此, 持有现金和存货在 τ_n 点发生的跳跃如下所示:

$$\begin{aligned} Y_{\tau_n} &= Y_{\tau_n -} + \zeta_n, \\ X_{\tau_n} &= X_{\tau_n -} - c(\zeta_n, P_{\tau_n}, \Delta_{\tau_n}), \end{aligned} \tag{7.11}$$

其中 $c(\zeta, p, \Delta) = \zeta p + |\zeta| \dfrac{\Delta}{2} + \phi$ 代表固定的交易费 ϕ 与市场订单的线性价格冲击之和.

7.3.2 最优限价和市价订单策略的控制公式

做市商必须在终期 T 出清库存. 这就产生了出清函数

$$L(x, y, p, \Delta) = x - c(-y, p, \Delta) = x + yp - |y| \frac{\Delta}{2} - \phi,$$

该函数对应于满足 T 时刻 $(X_T, Y_T, P_T, \Delta_T) = (x, y, p, \Delta)$ 情况下, 投资者出清所有库存所获得的价值. 假设 U 是一个递增的奖励函数, g 是一个非负的凸函数. Guilbaud 和 Pham (2013) 对于最优做市问题进行建模, 研究在常规控制 $\boldsymbol{L}_t$ 和脉冲控制集 $\mathcal{C}$, $0 \leqslant t \leqslant T$ 下的最优选择, 即对应于如下的价值函数的马尔可夫控制问题:

$$v(t, x, y, p, \Delta) = \sup_{\boldsymbol{L}, \mathcal{C}} E_{t,x,y,p,\Delta} \left\{ U\left[L(X_T, Y_T, P_T, \Delta_T)\right] - \gamma \int_t^T g(Y_u)\, du \right\}, \tag{7.12}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中 $E_{t,x,y,p,\Delta}$ 代表基于 $(X_t, Y_t, P_t, \Delta_t) = (t, x, y, p, \Delta)$ 的条件期望.

随机控制问题 (7.12) 中的拟变分不等式 (QVI) Guilbaud 和 Pham 利用动态规划准则发现价值函数 (7.12) 是 HJB 方程 (7.5) 的如下扩展方程的唯一黏性解:

$$\min\left[-\frac{\partial v}{\partial t} - \sup_{q^b, q^a, l} \mathcal{L}^{\boldsymbol{L}} v + \gamma g, v - \mathcal{M} v\right] = 0, \tag{7.13}$$

其边界条件为 $v(T, x, y, p, s) = U[L(x, y, p, \Delta)]$. 在 (7.13) 中, $\mathcal{L}^{\boldsymbol{L}}$ 是与马尔可夫控制过程相关的无穷小生成元[①], 该过程包含有限阶常规控制 $\boldsymbol{L} = (q^a, q^b, l)$, 且 $\sup\limits_{q^b, q^a, l}$ 指 $l \in \{0, 1\}$ 上的最大化. 类似于 7.2 节, 买入订单的报价 q^b 和卖出订单的报价 q^a 的上界为 Q. 除了 $\mathcal{L}^{\boldsymbol{L}}$ 以外, (7.13) 也指出了与市价订单脉冲控制有关的脉冲算子 $\mathcal{M}$, 其被定义为

$$\mathcal{M}\psi(t, x, y, p, \Delta) = \sup_{|\zeta| \leqslant Q} \psi\left[t, x - c(\zeta, p, \Delta), y + \Delta, p, \Delta\right]. \tag{7.14}$$

Pham (2005) 对于 HJB 方程/更一般的变分不等式/拟变分不等式的黏性解进行了综述. 他首先指出, HJB 方程是随机控制问题 $\sup\limits_{a_s \in \mathcal{A}} E\Big[\int_t^T f(s, X_s^{t,x}, a_s)\, ds + g(X_T^{t,x})\Big]$ 的动态规划原理 (DPP) 的无穷小版本, 具体如下:

$$-\frac{\partial v}{\partial t}(t, x) - \sup_{a \in \mathcal{A}} [\mathcal{L}^a v(t, x) + f(t, x, a)] = 0 \quad \text{在 } [0, T] \times \mathbb{R}^n \text{ 上},$$

其中 $\mathcal{L}^a$ 是与具有恒定控制 a 的扩散项 X 相关的二阶无穷小生成元. 函数 $H(t, x, v_x, v_{xx}) = \sup\limits_{a \in A}\{\mathcal{L}^a v(t, x) + f(t, x, a)\}$ 称为相关控制问题的 Hamilton 函数, 其中有 $v_x = \partial v/\partial x$ 和 $v_{xx} = \partial^2 v/\partial x \partial x^\top$ 是 v 在 (t, x) 点的梯度向量和 Hessian 矩阵. HJB 方程 "自然需要 Hamilton 函数 H 的有限性, 当一系列控制 $\mathcal{A}$ 有界时, 该条件通常满足". 当 $\mathcal{A}$ 无界时, H 在一些领域可能取值为 ∞. 在这种情况下, 控制问题称作奇异问题. 对于奇异随机控制问题, 用变分不等式代替 HJB 方程

$$\min\left\{-\frac{\partial v}{\partial t}(t, x) - H(t, x, v_x, v_{xx}), G(t, x, v_x, v_{xx})\right\} = 0,$$

其中 G 是连续函数, $G \geqslant 0$ 当且仅当 $H < \infty$ 时. 这个变分不等式实际上由两个不等式 $-\partial v/\partial t - H \geqslant 0$ 和 $G \geqslant 0$ 组成, 如果其中一个不等式是严格

① 在 3.4 节附注 5 中引入了受控马尔可夫过程的最小生成元 $\mathcal{L}^{\boldsymbol{L}}$. 我们在下一段中引入变分和拟变分不等式, 它们将 HJB 方程推广到没有通常正则性条件的情形, 这也解释了在这些情形下黏性解的概念.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

的, 那么另一个成为等式. 另一个扩展, 称为拟变分不等式, 利用其他函数例如 $v - \mathcal{M}v$ 来取代上述的 G, 其中 $\mathcal{M}$ 是与 (7.14) 中的脉冲控制或与制度转换模型中的跳跃相关联的算子; 详细细节参见 Bensoussan 和 Lions (1982) 与 Davis 等 (2010).

作为非线性 PDE 的 HJB 方程, 其解的存在性需要满足在许多随机控制问题中所违反的强正则条件. 通过考虑值函数 v 的变分 (或准变分) 不等式 (而不是 HJB 方程) 的黏性解, 可以避免这些技术上的困难. 若不要求 v 连续, 基本的思想是去考虑光滑的测试函数 φ 被 v 的下 (上) 半连续包络所界定的下 (上) 界, 使得这些光滑测试函数满足不等式 $-\dfrac{\partial \psi}{\partial t} - H(t, x, \varphi_x, \varphi_{xx}) \leqslant 0\ (\geqslant 0)$, 其中 v 称作变分 (拟变分) 不等式的黏性子解 (超解). 如果 v 同时是黏性子解和超解, 那么 v 是黏性解; Phom (2005) 的第 3 章和第 5 章也同时描述了求解黏性解的数值方法.

7.4 智能订单路径和暗池

智能订单路径 订单路径是指订单从最终用户到交易所的过程. 如图 7.1 所示, 智能订单路径 (SOR) 对价格、成本、价差等因素进行考量, 并利用交易平台的多选择性来实现 “最优” 的订单路径. 在 2010 年, 印度证券交易委员会 (SEBI) 批准了每一类投资者的 SOR, 孟买股票交易所 (BSE) 则于 2010 年 9 月引入 SOR. SOR 的存在将有助于交易执行和价格发现的顺利实现, 同时增

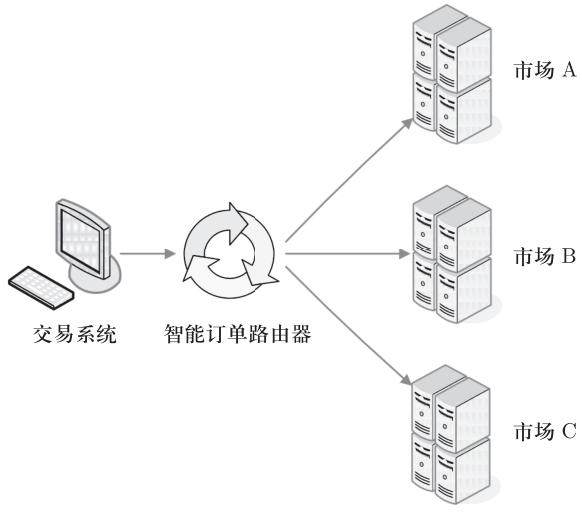


图 7.1 智能订单路径

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

加电子交易量.

正如我们在 6.5, 7.2 和 7.3 节中指出的, 在优化布局和做市方面, 算法交易的主要风险之一是执行风险. 执行风险是交易者在特定的时间段内与规模上无法执行其预设订单的风险, 这与运筹学中的存货风险十分类似. 在实践中, 为了降低执行风险, 大多数交易员利用有形 (或主要) 市场和暗池市场的组合, 后面我们将对此进一步解释.

暗池 在有形市场执行大额交易的主要困难之一是难以隐藏交易以及应对大额交易产生的价格影响. 暗池提供了一个替代交易系统, 用来为大额交易者解决这些问题. 在一个典型的暗池中, 订单和交易身份都是不公开的. 相反地, 买方和卖方通过简单地指定他们希望购买或卖出的股票总量来提交订单, 交易价格由特定暗池的匹配规则外生确定. 在提交买入 (卖出) 订单时, 交易员将被列入买方 (卖方). 买方和卖方之间根据订单到达的顺序进行匹配. 然而, 与提供 LOB 信息的主要市场不同, "暗池" 不提供任意给定时刻参与者或股票数量的任何信息. 在任意给定时期, 交易者只会被告知被执行的股票数量 (以暗池的最大容量为上限).

暗池已经占据了一定的市场份额, 并越来越受欢迎. 在美国, 据估计大约有 10%–20% 的交易量是通过暗池完成的, 约超过 40 个暗池为美国股市独有. 暗池具有一系列优势, 除了交易的保密性和减少市场冲击外, 由于交易价格通常是在最优买价和卖价的中点及附近, 暗池还可以帮助交易者节省价差成本. 许多暗池通常也提供较低的执行成本. 不同的暗池在回扣和收费结构以及流动性水平上都有一定区别. 暗池可以根据价格的匹配规则进行粗略分类. 一类是不可显示的限价订单交易, 这可能是代理人/经纪人或交易所拥有的. 这些暗池主要由大型经纪商经营, 其本质上是连续不显示的限价订单. 暗池的执行价格来自交易所的交易, 如全国最优买卖报价 (NBBO) 或交易量加权平均价格 (VWAP) 的中间价. 在这种情况下是没有价格发现能力的. 代理人/经纪人暗池包括 Instinet, Liquidnet 和 ITG Posit 等. 交易所拥有的暗池包括由 BATS Trading 和 NYSE Euronext 提供的暗池等. 另一类暗池由大型经纪商为客户设立 (并拥有), 也可能包括自营交易商. 这些暗池从订单流中获得价格, 所以能够存在价格发现能力. 这方面的例子包括 Credit Suisse Cross Finder, Goldman Sachs Sigma X, Morgan Stanley MS Pool 和 Citigroup's Citi Cross 等. 第三种类型的暗池则由 KCG (Knight Capital 和 Getco Holdings 合并后成立) 等独立电子做市商提供, 这些电子做市商也从订单流中获取价格.

在接下来的三节中, 我们考虑了交易所 (7.5 节) 和暗池 (7.6 节和 7.7 节)

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

的最优订单拆分问题. 对于暗池来说, 一个重要的特征是设计最优订单拆分方案所需的某些重要信息是不可得的 (因此被称作 “暗池”). 7.6 节和 7.7 节阐述了如何利用随机自适应控制和强化学习模型来解决这个难题. 其基本思路是首先考虑缺失信息可得的情况, 并找到最优规则, 然后利用次优规则来依次修正之, 这一规则会生成一系列数据以对缺失信息进行估计. 自适应控制规则的序贯性使得估计所得信息可作为新数据加入原数据集中. 7.5 节使用了类似的递归方法, 且在不同交易所的 SOR 中也进行了使用, 从而基本避免了由于最优订单拆分所导致的计算困难.

7.5 基于交易所 SOR 的最优订单分拆

每个交易所都有处理限制和市场订单的奖励和收费结构, 即使在相同的市场条件下, 交易所也可以显示不同的市场深度, 并且在填写订单时具有不同的速度. 因此, 一个自然的问题是如何将大额订单拆分成一系列合适规模的买入订单, 从而在有限的时间范围内以最小的成本最优地降低执行风险? Cont 和 Kukanov (2016) 在简化设定中研究了最优订单拆分问题. 在这种简化设定中, (a) 交易者需要在时间 T 内购买 N 股股票, 并且 (b) 在任何交易所中规模达到 N 的订单都可以立即被提交. 在时刻 0, 交易者可以将限价买入订单按照最优买入价格提交到 K 个交易场所中. 定义 L_i 为交易者在时刻 0 时在第 i 家交易所的限价买单规模, $i = 1, 2, \ldots, K$. 使用限价订单的 (每股) 返利和在第 i 家交易所使用市价订单的 (每股) 费用分别由 r_i 和 f_i 表示. 此外, η_i 指订单的总流出数, 包含交易所 i 在时间 $[0, T]$ 内的取消订单数和被执行的报价数. 假设 $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_K)$ 是一个联合分布函数 F 已知的随机向量. 在时刻 0, 除了限价订单, 交易者还可以提交一个可销售大小为 M 的买入订单到能以最优价格成交的交易所, 以保证 f_i 是所有交易所中最小的. 如果被执行的份额正好在时刻 T 之前少于 N, 交易者可以提交最终的 (有费用) 市场订单来购买剩余的数额.

7.5.1 成本函数与优化问题

交易者配置的订单可以用 $\boldsymbol{\Theta} = (M, L_1, \ldots, L_K) \in \mathbb{R}_+^{K+1}$ 表示. 在时刻 0, $L_1, \ldots, L_K$ 数量的限价订单加入 K 个交易所的订单队列 $Q_1, \ldots, Q_K$, 其中 $L_i \geqslant 0$ 和 $Q_i \geqslant 0$ 对所有 i 成立. Cont 和 Kukanov 假定所有这些限价订单都具有相同的价格, 其为全国最优出价, 即全交易所最高出价. $Q_i = 0$ 对应于第 i 家交易所的限价单的价格在买卖价差之内. 假设限价单在时刻 T 前未被修改,

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

则交易所会具体计算这些限价单在 T 时刻的 (全部或部分) 提交数量作为其初始队列位置和未来订单流出的函数 $\boldsymbol{\eta}$:

$$\min\left[\max(\eta_i - Q_i, 0), L_i\right] = (\eta_i - Q_i)^+ - (\eta_i - Q_i - L_i)^+,$$

$i = 1, \ldots, K$. 图 7.2 描述了时间 0 到 T 先进先出队列中限价单提交的演化过程. 因此, 时刻 T 前执行订单总额由下式给出:

$$A(\boldsymbol{\eta}, \boldsymbol{\Theta}) = M + \sum_{i=1}^{K} \left[(\eta_i - Q_i)^+ - (\eta_i - Q_i - L_i)^+\right]. \tag{7.15}$$

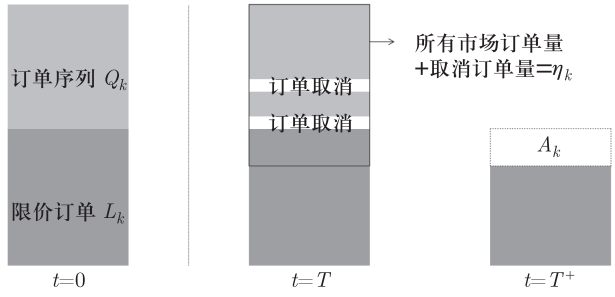


图 7.2 第 k 家交易所限价订单的演化过程. 执行规模 L_k 由 $A_k = (\eta_k - Q_k)^+ - (\eta_k - Q_k - L_k)^+$ 给出. 此处 L_k 被部分填充, 因此 $A_k = \eta_k - Q_k$

利用中间报价作为基准, 与 (7.15) 相关的执行成本为

$$(h+f)M - \sum_{i=1}^{K}(h + r_i)\left[(\eta_i - Q_i)^+ - (\eta_i - Q_i - L_i)^+\right], \tag{7.16}$$

其中 h 是时刻 0 的买卖价差的一半, f 则是最低的市场订单费用 (即最低流动性费用). 当 (7.15) 中的 $A(\boldsymbol{\eta}, \boldsymbol{\Theta})$ 与 N 不同时, 存在每股 λ_u 和 λ_o 的超出和低于正常调整的罚款. 此外, Cont 和 Kukanov 对于时刻 0 产生的市场冲击进行建模, 冲击为 $\theta\left(M + \sum_{i=1}^{K} L_i + [N - A(\boldsymbol{\eta}, \boldsymbol{\Theta})]^+\right)$, 并定义 $\theta > 0$ 为市场影响因子. 总成本 $c(\boldsymbol{\eta}, \boldsymbol{\Theta})$ 是执行成本 (7.16) 和这些额外成本的总和, 即

$$\begin{aligned} c(\boldsymbol{\eta}, \boldsymbol{\Theta}) = {} & (h+f)M - \sum_{i=1}^{K}(h + r_i)\left[(\eta_i - Q_i)^+ - (\eta_i - Q_i - L_i)^+\right] \\ & + \lambda_u[N - A(\boldsymbol{\eta}, \boldsymbol{\Theta})]^+ + \lambda_o\left[A(\boldsymbol{\eta}, \boldsymbol{\Theta}) - N\right]^+ \\ & + \theta\left(M + \sum_{i=1}^{K} L_i + [N - A(\boldsymbol{\eta}, \boldsymbol{\Theta})]^+\right). \end{aligned} \tag{7.17}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

因此, 最优化问题就是在时刻 0 选择 $\boldsymbol{\Theta} = (M, L_1, \ldots, L_K)$, 以最小化 $E[c(\boldsymbol{\eta}, \boldsymbol{\Theta})] = \int c(y, \boldsymbol{\Theta})\, dF(y)$. 最优选择可以表示为 $\boldsymbol{\Theta}^* = (M^*, L_1^*, \ldots, L_K^*)$.

7.5.2 K 个交易所的最优下单策略

Cont 和 Kukanov (2016, 3.2 节) 假设 F 是连续的, 且有

$$\max_i \{F_i(Q_i + N)\} < 1,\ \lambda_u < \max_i \left\{(2h + f + r_i)[F_i(Q_i)]^{-1} - (h + r_i + \theta)\right\},$$

其中 $F_i(x) = P(\eta_i \leqslant x)$, 以及

$$\begin{aligned} &\min_i r_i + h > 0,\ \lambda_u > h + f, \\ &\lambda_o > \max\left\{-(h + f),\ h + \max_i r_i\right\}. \end{aligned} \tag{7.18}$$

论文指出, 在 (7.18) 中的三个不等式有以下的经济解释: "即使一些有效的折扣 r_i 是负的, 限价订单执行仍然让交易者赚取一部分的买卖价差" 以及 "交易者没有动机超过目标数量 N 美元". 论文的主要结果如下:

- 如果 $\lambda_u \geqslant \dfrac{2h + f + \max\limits_i r_i}{F(Q_1, \ldots, Q_K)} - (h + \max\limits_i r_i)$, 则 $M^* > 0$.
- 如果 $(h + r_i - \lambda_o)P(\eta_i > Q_i) > \theta$, 则 $L_i^* > 0$. 此外, 当 $\boldsymbol{\Theta}^*$ 是 $E[c(\boldsymbol{\eta}, \boldsymbol{\Theta})]$ 的最小值时, 对于 $j = 1, \ldots, K$ 的充要条件是:

$$\begin{aligned} &P\left\{M^* + \sum_{i=1}^K \left[(\eta_i - Q_i)^+ - (\eta_i - Q_i - L_i^*)^+\right] < N\right\} = \frac{h + f + \lambda_o + \theta}{\lambda_u + \lambda_o + \theta}, \\ &P\left\{M^* + \sum_{i=1}^K \left[(\eta_i - Q_i)^+ - (\eta_i - Q_i - L_i^*)^+\right] < N \middle| \eta_j > Q_j + L_j^*\right\} \\ &= \frac{[\theta / P(\eta_j > Q_j + L_j^*)] + \lambda_o - (h + r_j)}{\lambda_u + \lambda_o + \theta}. \end{aligned}$$

7.5.3 一个随机逼近方法

Cont 和 Kukanov (2016, 第 4 节) 指出, 由于 $E[c(\boldsymbol{\eta}, \boldsymbol{\Theta})]$ 在 F 绝对连续时是 K 维积分, 随着交易所数量的增加, 计算难度越来越大, 找到最小值 $\boldsymbol{\Theta}^*$ 变得更加困难. 然而, "实际应用需要一种快速灵活的方法来优化多个交易场所的下单策略". 为克服这一困难, 论文使用 Nemirovski 等 (2009) 提出的随机逼近方法来进行求解. 定义 $\mathbf{g}(\boldsymbol{\eta}, \boldsymbol{\Theta}) = \dfrac{\partial c(\boldsymbol{\eta}, \boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}}$. 随机逼近方法是一个递归方法,

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

该方法首先用 $\mathbb{R}_+^{K+1}$ 中的初步估计量 $\boldsymbol{\Theta}^{(0)}$ 进行初始化. 接下来使用取决于 m 的步长 $\gamma_m > 0$, 并从分布 F 中生成独立同分布的样本 $\boldsymbol{\eta}_j^*$. 递归的形式如下:

$$\boldsymbol{\Theta}_j = \boldsymbol{\Theta}_{j-1} - \gamma \mathbf{g}(\boldsymbol{\eta}_j^*, \boldsymbol{\Theta}_{j-1}), \quad j = 1, \ldots, m. \tag{7.19}$$

最优规则 $\boldsymbol{\Theta}^*$ 的估计值为 $\hat{\boldsymbol{\Theta}}_m^* = \sum_{j=1}^m \dfrac{\boldsymbol{\Theta}_j}{m}$. 7.8 节的补充材料 5 和补充材料 6 给出了 (7.19) 的非参变体和 $\hat{\boldsymbol{\Theta}}_m^*$ 到 $\boldsymbol{\Theta}^*$ 的收敛性质, 以及 (7.19) 的统计背景及其变体.

7.6 暗池的删失探索和开发

Ganchev 等 (2010) 引入序贯学习和适应控制方法来在暗池中下单, 并讨论与前述交易所中 SOR 的区别, 论文称之为 "明" 交易所 (与 "暗" 相对). 论文指出:

> 在一个典型的暗池中, 买方和卖方提交的订单只是简单地说明他们希望买卖的股份总量, 而交易的价格是由 "市场" 外生确定的 (即 "明" 交易所中买价和卖价的中间值). 在提交买入 (或卖出) ν 份股票的指令后, 交易者被置于等待交易的买方 (或卖方) 队列中. 类似于明交易所, 买方和卖方之间的匹配发生在订单的连续到达序列中. 然而, 与明交易所不同的是, 交易者没有被提供关于在任何给定时刻可能有多少参与方或者可得股份的信息. 因此, 在给定的时间段内, 提交 ν 份股票订单的交易者只会知道其中有多少被执行. ····· 如何在多个独立的暗池中最优分配大量交易? ····· 我们分析了一个更一般的多场所探测问题的框架和算法 (基于设限观察值).

7.6.1 贪心配置算法

假设在每个时刻 $t = 1, \ldots, T$, 交易者将 ν_t 股股票在 K 个交易所卖出. 第 i 家交易所只能在 t 时间卖出最大数额 $U_t^i \in \mathbb{N}$ 的股票. 然而, 由于这些交易所是暗池, 交易者并不知道这个信息. 假设 $U_1^i, \ldots, U_T^i$ 是独立同分布的, 且其生存函数 T_i 未知, 即 $T_i(u) = P(U_t^i \geqslant u)$. 定义 $\boldsymbol{n}_t = (n_t^1, \ldots, n_t^K) \in \mathbb{N}^K$ 为交易者在时刻 t 的配置, 且有 $\sum_{i=1}^K n_t^i = \nu_t$. 因此, 交易者目标的优化问题为

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$$\max_{\boldsymbol{n}_t\in\mathbb{N}^K}\sum_{i=1}^{K}E\left[\min(n_t^i,U_t^i)\right]\quad\text{满足}\ \sum_{i=1}^{K}n_t^i=\nu_t,\tag{7.20}$$

对于每个 $t=1,\ldots,T$. 由于 $E[\min(n,U_t^i)]=\sum_{s=1}^{n}T_i(s)$, Ganchev 等 (2010) 将 (7.20) 表述为

$$\max_{\boldsymbol{n}_t\in\mathbb{N}^K}\sum_{i=1}^{K}\sum_{s=1}^{n_t^i}T_i(s)\quad\text{满足}\ \sum_{i=1}^{K}n_t^i=\nu_t,$$

该问题可以用如下方法来求解:

- 如果 $\nu_t=1$, 则交易的最优场所为 $j_1=\arg\max_{i=1,\ldots,K}T_i(1)$. 因此, 最优化 $\boldsymbol{n}_t=(n_t^1,\ldots,n_t^K)$ 时, 对于 $i\neq j_1$, $n_t^{j_1}=1$ 和 $n_t^i=0$.
- 如果 $\nu_t=2$, 则 $j_1=\arg\max_{i=1,\ldots,K}T_i(1)$ 为首先进行交易的场所. 接下来的最优场所为 $j_2=\arg\max_{i=1,\ldots,K}T_i(v_i+1)$, 当 $i=j_1$ 时 $v_i=1$, 否则 $v_i=0$.

在以这种方式进行处理时, 最大化 (7.20) 的最优分配规则 $\boldsymbol{n}_{t,\text{opt}}$ 的贪心方法是每次分配一个份额, 并按顺序使 K 个交易所上的尾部概率最大的贪心算法, 其伪代码如下:

贪心配置算法

输入: 份额 ν_t, $T_i, i=1,\ldots,K$ 的尾部概率分布函数

输出: 最优配置 $\boldsymbol{n}$

$\boldsymbol{n}=(n_1,\ldots,n_K)\leftarrow\mathbf{0}$

for $\ell\in\{1,\ldots,\nu_t\}$ **do**

  $j\leftarrow\arg\max_i T_i(n_i+1)$

  $n_j\leftarrow n_j+1$

end for

return $\boldsymbol{n}$

7.6.2 T_i 的改进 Kaplan-Meier 估计量 $\hat{T}_i^t$

在 T_i 未知的情况下, Ganchev 等 (2010) 基于删失数据 $R_t^i=\max(n_t^i,U_t^i)$, 利用 Kaplan-Meier 估计量对 T_i 进行序贯估计, 该估计量也是基于删失样本

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$R_s^i, s \leqslant t$ (而不是独立同分布的 U_s^i) 的非参数极大似然估计量 (NPMLE). 定义

$$M_{t,j}^i = \sum_{s=1}^{t} I_{\{R_s^i=j, n_s^i>j\}},$$
$$N_{t,j}^i = \sum_{s=1}^{t} I_{\{R_s^i\geqslant j, n_s^i>j\}},$$

$h_j^i = P(U_t^i = j | U_t^i \geqslant j)$ 的 NPMLE 为 $\hat{h}_{t,j}^i = \dfrac{M_{t,j}^i}{N_{t,j}^i}$. 约定 $\dfrac{0}{0} = 0$, 则 $T_i(\ell)$ 的 NPMLE 为

$$\hat{T}_i^t(\ell) = \prod_{j=0}^{\ell-1} \left(1 - \hat{h}_{t,j}^i\right); \tag{7.21}$$

具体可参见 Lai 和 Xing (2016) 的 2.4 节. Ganchev 等 (2010) 提出, $\hat{T}_i^t(\ell)$ 可以通过如下递归法进行更新:

$$\begin{aligned} M_{t+1,j}^i &= M_{t,j}^i + I_{\{R_{t+1}^i=j, n_{t+1}^i>j\}}, \\ N_{t+1,j}^i &= N_{t,j}^i + I_{\{R_{t+1}^i\geqslant j, n_{t+1}^i>j\}}. \end{aligned} \tag{7.22}$$

Lai 和 Ying (1991a) 指出, $\hat{T}_i^t(\ell)$ 的统一一致性只有在 "风险集大小" $N_{t,\ell}^i$ 充分大的情况下才成立, 并且提出了具有更好收敛性质的 Kaplan-Meier 估计的改进. Ganchev 等 (2010) 提出了一个不同的改进方法, 即利用 $\hat{T}_i^t(\ell)$ 的 (随时间 t 变化) 连续性来估值. 该方法使用一个界限值 c_i^t, 使得在 $\ell \leqslant c_i^t$ 时, $\hat{T}_i^t(\ell)$ 能够对于 $T_i(\ell)$ 进行很好的估计. 论文在选择 c_i^t 时, 使用了 2 个调节参数 $\varepsilon > 0$ 和 $\delta > 0$:

$$c_i^t = \max\left\{\ell \leqslant \nu : \ell = 0 \text{ 或 } N_{t,\ell-1}^i \geqslant 128\left(\frac{\ell\nu}{\varepsilon}\right)^2 \ln\left(\frac{2\nu}{\delta}\right)\right\}, \tag{7.23}$$

其中, $\nu = \sum_{t=1}^{T} \nu_t$ 是时刻 T 时交易者卖出股票的总份额数. 论文定义改进的 Kaplan-Meier 估计值为

$$\tilde{T}_i^t(\ell) = \begin{cases} \hat{T}_i^t(\ell), & \text{当 } \ell < c_i^t + 1, \\ \hat{T}_i^t(c_i^t), & \text{当 } \ell \geqslant c_i^t + 1. \end{cases}$$

论文的引理 1 和 2 指出, 以下的不等式至少以概率 $1-\delta$ 成立:

$$|\tilde{T}_i^t(\ell) - T_i(\ell)| \leqslant \frac{\varepsilon}{8\nu} \quad \text{对于所有 } \ell \leqslant c_i^t; \tag{7.24}$$

$$|\tilde{T}_i^t(\ell) - T_i(\ell)| \leqslant \frac{\varepsilon}{2\nu} \quad \text{对于所有 } c_i^t < \ell \leqslant \nu, \text{ 如果 } \hat{T}_i^t(c_i^t) \leqslant \frac{\varepsilon}{4\nu}. \tag{7.25}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

7.6.3 探索、开发和 ε 最优配置

Ganchev 等 (2010, 第 102 页) 指出:“通过对每个交易所的尾部概率 $\hat{T}_i^t(c_i^t+1)$ 进行乐观的改进, 由于该算法运行次数很少, 因此我们确信没有交易所不被探索过.” 基于该想法, 在贪心配置算法 $\boldsymbol{n}_{t,\text{opt}}$ 中, 在每个时刻 t, 使用 $\hat{T}_i^t(\cdot)$ 来替换 $T_i(\cdot)$, 其适应配置准则 $\hat{\boldsymbol{n}}_t$ 是 ε 最优的. 这个论点涉及 “开发” 引理和 “探索” 引理, 二者分别为 Ganchev 等 (2010) 论文中的引理 3 和 4.

开发引理 *在任何时刻 t, 如果 (7.24) 成立并且对每个交易所 i (a) $\hat{n}_t^i \leqslant c_i^t$ 或 (b) $\hat{T}_i^t(c_i^t) \leqslant \dfrac{\varepsilon}{4\nu}$ 成立, 则 $\max\limits_{i=1,\dots,K}|\hat{n}_t^i - n_{t,\text{opt}}^i| < \varepsilon$.* 这个引理提出, 如果 (7.24) 成立, 并且卖出的股票数量不超过界限值, 或者如果 (b) 对于每个交易所都成立, 那么可以证明 $\hat{\boldsymbol{n}}_t$ 是 ε 最优.

探索引理 Ganchev 等 (2010) 的引理 4 表明, 在任意时刻 t, 如果 $\hat{\boldsymbol{n}}_t$ 不是 ε 最优, 则 $\hat{\boldsymbol{n}}_t$ 生成的 “有用观测量” 的概率值存在正的下限, 因而会使得相关风险集大小增加 1. 论文中引理的证明具体阐述了使用探索引理来生成有用观测量的基本想法:

> 假设在时刻 t, 配置不是 ε 最优. 根据开发引理可知, 必然存在交易所 i, 满足 $\hat{n}_t^i > c_i^t$ 且 $\hat{T}_i^t(c_i^t) > \dfrac{\varepsilon}{4\nu}$. 即在该交易所中, 算法已经配置了超过界限的单位, 但是界限处的尾部概率并不接近 0. 定义 ℓ 为一个真实的场所. 由于 $\hat{n}_t^i > c_i^t$, 贪心算法会得到一个有用的观测值来探索这个交易所 (即能够使风险集大小 $N_{t,c_\ell^t}^\ell$ 增加的观测值)$\cdots\cdots$ 既然 $\hat{T}_\ell^t(c_\ell^t) > \dfrac{\varepsilon}{4\nu}$, (7.24) 意味着 $T_\ell^t(c_\ell^t) > \dfrac{\varepsilon}{8\nu}$.

将探索和开发结合为一个 ε 最优规则 根据开发和探索引理, 可以将 $\hat{n}_t$ 定义为 “删失开发–探索准则”. Ganchev 等 (2010) 的定理 3 将上述引理相结合, 证明存在一个随机的烧机时刻 τ, 其概率至少为 $1-\delta$, τ 是 K, ν, $\dfrac{1}{\varepsilon}$ 和 $\ln\dfrac{1}{\varepsilon}$ 的多项式, 且满足

$$P\left\{\max_{i=1,\dots,K}|\hat{n}_t^i - n_{t,\text{opt}}^i| < \varepsilon|\tau\right\} \geqslant 1-\varepsilon \quad \text{对于所有 } t > \tau. \tag{7.26}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

7.7 暗池中的随机拉格朗日优化

Laruelle 等 (2011) 在前一节的最优订单拆分问题中介绍了一些附加特征. 最重要的特征为取值于 $(0,\infty]$ 的 ν_t 和 U_t^i, 而不是 Ganchev 等 (2010) 中使用的 $\mathbb{N}$. 因此, 暗池的订单数不需要是整数, 我们可以使用 r_t^i 来表示分配给第 i 个暗池的 ν_t 的比例, 且有 $\sum_{i=1}^K r_t^i = 1$. 另一个主要区别是, 第 i 个暗池的返利结构是基于 $\theta_i \in (0,1]$ 来建模的, 此时第 i 个暗池中的买价是 $\theta_i b_t$, 其中 b_t 是时刻 t 时主要市场资产的最优买入价. 因此, 在第 K 个交易所中拆分购买 ν_t 份股票的成本 $c(t)$ 是

$$\begin{aligned} c(t) &= b_t \sum_{i=1}^K \theta_i \min(r_t^i \nu_t, U_t^i) + b_t \left[\nu_t - \sum_{i=1}^K \min(r_t^i \nu_t, U_t^i) \right] \\ &= b_t \left[\nu_t - \sum_{i=1}^K \rho_i \min(r_t^i \nu_t, U_t^i) \right], \end{aligned}$$

其中 $\rho_i = 1 - \theta_i > 0$ 是第 i 个暗池的吸引力系数. 在该框架下, 最优的订单拆分问题变为在给定 b_t 情况下 $c(t)$ 期望的最小化问题. 假设 $(\nu_t, U_t^1, \ldots, U_t^K)$ 独立于 b_t, 此时订单拆分问题等价于最优化问题

$$\max_{\boldsymbol{r}_t \in \mathbb{S}} \sum_{i=1}^K \rho_i E\left[\min(r_t^i \nu_t, U_t^i)\right], \tag{7.27}$$

其中 $\boldsymbol{r}_t = (r_t^1, \ldots, r_t^K)$ 且 $\mathbb{S}$ 是 K 维的单纯形, 即

$$\mathbb{S} = \left\{ (r_1, \ldots, r_K) \in \mathbb{R}_+^K : \sum_{i=1}^K r_i = 1 \text{ 且对于任意 } j, 0 \leqslant r_j \leqslant 1 \right\}.$$

Ganchev 等 (2010) 使用了 7.6.1 节中贪心算法的表达式 $E[\min(n, U_t^i)] = \sum_{s=1}^n T_i(s)$, Laruelle 等 (2011) 基于下式的凹性和可微性推导出它们的随机拉格朗日算法

$$\varphi_i(r) = E\left[\min(r\nu_t, U_t^i)\right], \tag{7.28}$$

论文中称之为第 i 个交易所的平均执行函数. 尽管 $\min(r\nu, U)$ 在 $r \in [0,1]$ 中不是一个平滑函数, 论文表明在满足以下假设的情况下 φ_i 在 $[0,1]$ 上是可微的:

$$P\{\nu_t > 0\} = 1, \quad P\{U_t^i > 0\} = 1, \quad \text{且 } \frac{U_t^i}{\nu_t} \text{ 连续.}^{①} \tag{7.29}$$

① 如果随机变量具有连续分布函数, 则它称为连续变量.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

在 (7.29) 的情况下, φ 在 $[0,1]$ 上可微, 且有

$$\varphi_i'(r) = E[\nu_t I_{\{r\nu_t \leqslant U_t^i\}}], \quad \varphi_i'(0) = E[\nu_t I_{\{U_t^i \geqslant 0\}}] > 0. \tag{7.30}$$

7.7.1 基于随机近似的拉格朗日方法

本节在 7.5.2 节中 $L_1, \ldots, L_K$ 的基础上, 放宽整数约束, 从而可以利用导数来求解优化问题 (7.27). 事实上, $\sum_{i=1}^K \rho_i E[\min(r_t^i \nu_t, U_t^i)] = \sum_{i=1}^K \rho_i \varphi_i(r_i)$. 约束优化问题 (7.27) 可以用拉格朗日方法来求解. 令 $\dfrac{\partial G}{\partial \boldsymbol{r}_t} = 0$, 其中

$$G(\boldsymbol{r}_t) = \sum_{i=1}^K \rho_i \varphi_i(r_t^i) - \lambda \sum_{i=1}^K r_t^i,$$

而 λ 是拉格朗日乘子. 最优配置规则 $\boldsymbol{r}_{t,*} = (r_{t,*}^1, \ldots, r_{t,*}^K)$, 其中 $*$ 表示 "最优". 因此, 方程系统的解为:

$$\rho_j \varphi_j'(r_{t,*}^j) = \frac{1}{K} \sum_{i=1}^K \rho_i \varphi_i'(r_{t,*}^i), \quad j = 1, \ldots, K. \tag{7.31}$$

尽管 (7.31) 通常需要 $\boldsymbol{r}_{t,*} \in \mathrm{Int}(\mathbb{S})$, Laruelle 等 (2011, 命题 1) 证明, 当满足如下条件时, (7.31) 即使是在 $\boldsymbol{r}_{t,*} \in \partial\mathbb{S}$ 的情况下也成立:

$$\min_{1 \leqslant i \leqslant K} \varphi_i'(0) \geqslant \max_{1 \leqslant i \leqslant K} \varphi_i' \left[(K-1)^{-1} \right]. \tag{7.32}$$

论文同时讨论了这些条件的解释, 且证明当所有的 ρ_i 相等, 即所有 K 个交易所提供相同的返利时, 上述条件在 (7.29) 下成立.

基于随机近似 (SA) 的执行 将 (7.30) 和 (7.31) 相结合, 可得

$$E\left[\nu_t \left(\rho_j I_{\{r_{t,*}^j \nu_t \leqslant U_t^j\}} - \frac{1}{K} \sum_{i=1}^K \rho_i I_{\{r_{t,*}^i \nu_t \leqslant U_t^i\}} \right) \right] = 0, \quad j = 1, \ldots, K. \tag{7.33}$$

在对 (7.33) 进行直接执行时, 需要首先对 $(\nu_t, U_t', \ldots, U_t^K)$ 的联合分布进行描述, 从而利用蒙特卡罗方法来计算上式中的期望. Laruelle 等 (2011) 的 2.1 节讨论了一些可行模型, 但是由于基于 $(\nu_t, U_t', \ldots, U_t^K)$ 的 $\boldsymbol{r}_{t,*}$ 形式仍需确定, 这些模型并不实用. 该论文的 2.2 节放弃了这些模型, 并使用随机近似方法[①] 求解, 该方法中递归模型随时间变化. 由于该问题的解 $\boldsymbol{r}_{t,*}$ 约束在 $\mathbb{S}$ 范围内, 因

① 7.8 节的补充材料 6 对 SA 方法的背景和细节进行了进一步介绍.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

此论文的 3.3 节使用了另外一种技术, 从而将该问题转化为一个非约束问题以使得 SA 方法在此处可以使用. 具体而言, 此处的 SA 递归为

$$\boldsymbol{r}_{t+1} = \boldsymbol{r}_t + \gamma_t \boldsymbol{H}(\boldsymbol{r}_t, \nu_t, U_t^1, \dots, U_t^K), \tag{7.34}$$

其中 $\gamma_t > 0$ 是学习率序列, $\boldsymbol{r}_0$ 是初始估计值, 而 $\boldsymbol{H}(\boldsymbol{r}_t, \nu_t, U_t^1, \dots, U_t^K)$ 是一个 K 维向量, 其第 j 项定义为如下方程:

$$\begin{aligned} &H_j(\boldsymbol{r}, \nu, U_1, \dots, U_K) \\ &= \nu \left[\rho_j I_{\{r_j \nu \leqslant U_j\}} - \frac{1}{K} \sum_{i=1}^{K} \rho_i I_{\{r_i \nu \leqslant U_i\}} + \xi_j(\boldsymbol{r}, \nu, U_1, \dots, U_K) \right], \end{aligned} \tag{7.35}$$

其中 ξ_j 有一种 "均值回复" 效果, 从而将 $\boldsymbol{r}$ 拉进单纯形 $\mathbb{S}$:

$$\begin{aligned} &\xi_j(\boldsymbol{r}, \nu, U_1, \dots, U_K) \\ &= \rho_j \left[(1 - r_j) I_{\{0 \leqslant U_j, r_j < 0\}} + \frac{1}{r_j} I_{\{\nu \leqslant U_j, r_j > 1\}} \right] \\ &\quad - \frac{1}{K} \sum_{i=1}^{K} \rho_i \left[(1 - r_i) I_{\{0 \leqslant U_i, r_i < 0\}} + \frac{1}{r_i} I_{\{\nu \leqslant U_i, r_i > 1\}} \right], \end{aligned} \tag{7.36}$$

上式可以通过将单位区间内的 φ_j 拓展到一个实轴上的可微凹函数来求得. 7.7.2 节对于上述递归在 $t \to \infty$ 上的收敛进行了讨论.

7.7.2 拉格朗日递归到优化函数的收敛性

Laruelle 等 (2011, 第 1054 页) 对随机近似递归过程 (7.34)–(7.36) 给出了如下的启发式解释:

> 只要 r 是真正的配置向量, 即属于单纯形 …… 首先假设所有因素 ρ_i 相等 (都等于 1). 然后对发送的订单指令 ($r_i \nu \leqslant U_i$) 全部执行的暗池会得到一定比例的奖励. 另一方面, 不能执行所有订单指令的暗池则会受到一定比例的惩罚.

因此, 只要前面的 L 个暗池在时刻 t 已经完全执行了订单指令, 才满足 $r_{t+1}^i = r_t^i + \gamma_t \left(1 - \dfrac{L}{K}\right) \nu_t$, $i = 1, \dots, L$, 即这些交易所由于在时刻 $t+1$ 增加执行订单的比例而获得回报. 另一方面, $r_{t+1}^j = r_t^j - \gamma_t \left(\dfrac{L}{K}\right) \nu_t$, $j = L+1, \dots, K$, 即这些交易所会因为在时刻 $t+1$ 减少执行订单的比例而受到处罚. 通过权衡惩罚与奖励的系数 ρ_i, 上述解释也可以扩展到 ρ_i 不完全相等的情况.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

Laruelle 等 (2011, 定理 4.1) 研究了随机近似递归模型 (7.34) 在 $(\nu_t, U_t^1, \ldots, U_t^K)$ 独立同分布假设下, 具有相同分布 $(\nu, U_1, \ldots, U_K)$ 的收敛性. 除了独立同分布假设外, 论文也假设 (7.32) 与下面条件成立:

(i) $\nu > 0$ 几乎处处成立, 且 $E(\nu^2) < \infty$;

(ii) $F_i(x) = P\left\{\dfrac{U_i}{\nu} \leqslant x\right\}$ 是 $\mathbb{R}_+$ 上的连续函数, 对于所有的 $i = 1, \ldots, K$, 满足 $P(U_i > 0) > 0$;

(iii) $\sum_{t=1}^{\infty} \gamma_t = \infty$ 且 $\sum_{t=1}^{\infty} \gamma_t^2 < \infty$.

在这些假设下, 论文证明 (7.34) 在 $t \to \infty$ 时几乎处处收敛于

$$\boldsymbol{r}_* = \arg\max_{\boldsymbol{r} \in \mathbb{S}} \sum_{i=1}^{K} \rho_i E\left[\min(r_i, \nu, U_i)\right]. \tag{7.37}$$

而且, 如果严格不等式 (7.32) 成立, 则有 $\boldsymbol{r}_* \in \mathrm{Int}(\mathbb{S})$.

Laruelle 等 (2011) 论文的第 5 节将上述严格的独立同分布假设 (定义该假设是为了便于使用传统的 SA 理论) 进行放宽, 假设 $(\nu_t, U_t^1, \ldots, U_t^K)$ 是一个遍历序列, 其弱收敛于 $(\nu, U_1, \ldots, U_K)$. 此时假设 (i) $\nu > 0$, 且满足 $\sup_t E\nu_t^2 < \infty$. 另外, 论文中将假设 (iii) 替代为

(a) $T^{-1}\sum_{t=1}^{T} \nu_t I_{\{y\nu_t \leqslant U_t^j\}} \to E[\nu I_{\{y\nu \leqslant U_j\}}] = O(T^{-\alpha_j})$ 对于所有的 $y \in \mathbb{R}_+$ 和 $j = 1, \ldots, K$ 以及一些 $\alpha_j \in (0, 1]$ 几乎处处成立;

(b) $\sum_{t=1}^{\infty} \gamma_t = \infty$, $\gamma_t = o(t^{a-1})$, 且 $\sum_{t=1}^{\infty} t^{1-a} \max(\gamma_t^2, |\gamma_t - \gamma_{t+1}|) < \infty$,

其中 $a = \min\limits_{1 \leqslant i \leqslant K} \alpha_j$. 在这个更一般的设定下, 论文证明 $\boldsymbol{r}_t \to \boldsymbol{r}_*$ 几乎处处成立; 具体细节参见 7.8 节的补充材料 6.

7.8 补充材料和注释

1. **做市商仓位控制** 7.1–7.3 节展示了学术上在最优化市场订单上具体的模型和结果, 这节补充介绍这些模型的变形以及库存控制, 尤其是那些运用在实证研究中或者数据驱动的内容. 在 7.3.2 节中, 做市商在到期日 T 之前会对于仓位进行清算, 且必要的话会在 T 时交易市场订单. 然而, 在 7.1 节和 7.2 节中只考虑了限价订单, 因此没有考虑他们清算的要求. 在 Hasbrouck 和 Sofianos (1993), Madhavan 和 Smidt (1993) 与 Hendershott 和 Seasholes (2007) 的实证研究中, 做市商 (或 NYSE 的 "特种经纪人") 通常会持有一些仓位. 特别地, Hendershott 和 Seasholes (2007) 利用 NYSE 在 1994 年至 2004 年间的总结文件 (SPETS, 文件中提供了每天每只证券的特种经纪人的仓位数据) 中的内部数据, 对于交易

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

日收盘时的仓位数据与日度和月度的收益之间的关系进行研究. 研究发现, 多头和空头仓位存在非对称性, 且 “边际仓位的增加会带来收益的增加”, 因此对特种经纪人投资组合而言, “大部分的大额买入 (卖出) 订单会出现在价格下降 (增加) 的交易日”. 另外, “价格呈现出与预先形成的收益率相关的均值回复特征, 因此这些大额交易在特种经纪人看来是无利可图的”.

Huang 等 (2012) 对汇率市场上的做市商进行了研究. 论文指出 “之前大量的工作对于仓位如何影响交易者的在提供买方和卖方时候的行为进行研究”, 该论文同时引用了本书 7.1 节中介绍的 Ho 和 Stoll 模型. 论文指出, “实证研究表明, 与其他因素的影响相比, 仓位水平对价格的影响相当有限”, 而且 “一项对美国外汇交易员的研究显示, 市场规范是买卖价差的重要决定因素, 只有一小部分买卖价差与传统价差有所不同.” 因此, 该论文致力于寻找 “适用于不对于价格进行控制、且很少利用积极型交易来调整仓位的做市商的有效策略”. 论文使用了与经典的仓位控制模型类似的模型来求解 “仓位过高的风险与其他做市商交易的成本 (以弥补仓位的不足) 的最优权衡问题”. 论文同时发现, 经典的仓位控制理论与论文中的做市商仓位控制理论有着显著的不同. 在经典的理论中 “订单数量必须是非负的, 单位仓位的持有成本是确定的, 订单成本是线性的”. 论文在对优化问题进行建模时, 包含了与市场交易者的客户交易中获得的利润以及与其他交易者进行交易所产生的成本项. 控制项 q_t 表示买方 (当 $q_t > 0$) 或卖方 (当 $q_t < 0$) 在时刻 $t = 1, \ldots, T$ 时交易的数量. 论文假设总利润 (减去成本) 服从指数效用, 求解得到价值函数的 Bellman 方程, 研究发现最优策略中遵从阈值规则, 具体而言, (a) 当仓位位于两个阈值之间时, 不进行交易; (b) 如果仓位超过上限, 卖出股票, 以及 (c) 如果库存低于下限, 买入股票.

Cartea 等 (2014) 指出, 如今在证券市场中, “仓位与交易的执行都由计算机完成”, 而报价驱动的市场 (至少在证券市场上) 则让位于限价订单市场. 在报价驱动的市场中, 做市商对于买入和卖出订单进行交易, 而投资者则基于报价进行投资 (如 Huang 等 (2012) 考虑的情形). 而在限价订单市场中, 所有参与者都可以表现得像旧的订单驱动的做市商一样. 论文指出, “在交易日期间, 所有的订单都在 LOB 中累积, 直到找到交易对手执行, 或者被发布订单的代理人取消. 而交易对手是指市场订单 ……” 因此, 在这个框架中, 市场交易是 LOB 驱动的, 与我们在 7.3 节所描述的

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

思想相似. Guilbaud 和 Pham (2013) 也使用了这种方法. 该论文进一步假设市场订单 (MOs) 主要有两种类型:

> 第一类订单是有影响力的订单, 激活市场并诱导其他交易者增加他们提交的市场订单的数量. 举例来说, 一个有影响力的市场卖出指令的到来增加了下一期卖出订单出现的可能性, 且 (小幅) 增加了下一期买入订单出现的可能性. 另一方面, 当没有影响的订单到达时, 市场订单的到达强度不会改变. 由此可知, 市场中存在着一类订单, 这些订单在其他交易者看来并没有传递信息, 因此并不会使得这些交易者改变他们购买市场订单的意愿. 从这个角度来说, 本论文中关于市场订单的模型能够捕捉到的交易集聚可以是片面的, 也可以是双面的, 并且允许交易的活动显示出算法交易给市场环境带来的积极反馈. 近年来, 多元的 Hawkes 过程被用于金融计量经济学文献中 …… 然而, 本文第一次将这种影响引入到与算法交易相关的最优控制问题中 ……. 市场订单的到达通常被认为包含一定的信息, 因为它可以传递有关随后价格变动和逆向选择风险的信息. 这里我们假设资产的中间价格的动态受到有影响力的市场卖出和买入订单的短期不平衡的影响; …… 有影响力的订单的到来对 LOB 的买入卖出两侧产生了短暂的影响. 更具体地说, 由于一些市场交易者预期卖出订单和买入订单的强度会发生变化, 做市商考虑到可能增加的市场交易和逆向选择风险, 会重新配置订单簿, 因此订单簿的买卖两侧会产生短期的变化.

2. **衍生品市场的做市** 本节补充材料 1 指出, 市场交易者的约束和目标在不同市场上可能有所不同. 7.1 – 7.3 节已经对于股票市场中的模型和方法进行介绍. 在衍生品市场中, 做市商通常假设市场上存在的交易摩擦不符合经典的 Black-Sholes 类型的假设 (参见 Lai 和 Xing (2008, 第 183 页)), 其中最重要的一个区别就是衍生品市场存在交易成本. 在交易成本存在时, 期权定价涉及在 3.4 节的补充材料 5 中所讨论的奇异型随机控制; 参见 Davis 等 (1993) 与 Lai 和 Lim (2009). Stoikov 和 Saglam (2009) 指出, 欧式期权 "状态空间可以被分解成交易和非交易区域", 而做市商不能 "持续地进行交易并且可能也会在股票市场中做市". 做市商控制为股票中间价 S_t 附近的溢价 δ_t^a, δ_t^b 和期权的中间价 $C_t = C(S_t, t)$ 附近的溢价 $\tilde{\delta}_t^a, \tilde{\delta}_t^b$,

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

其中

$$dC(S_t,t)=\Theta_t dt+\Delta_t dS_t+\Gamma_t(dS_t)^2/2,$$

其中 Δ_t,Γ_t 和 Θ_t 分别是期权的 Delta, Gamma 和 Theta (参见 Lai 和 Xing (2008, 第 313 页)). 交易的总收益 Z_t 服从以下过程:

$$dZ_t=\delta_t^a dN_t^a+\delta_t^b dN_t^b+\tilde{\delta}_t^a d\tilde{N}_t^a+\tilde{\delta}_t^b d\tilde{N}_t^b,$$

仓位的价值 I_t 服从 $dI_t=q_t dS_t+\tilde{q}_t dC_t$, 此处 N_t^a, N_t^b 和 q_t 的定义与 7.1 节相同, 而且 $\widetilde{N}_t^a$, $\widetilde{N}_t^b$ 和 $\widetilde{q}_t$ 均是与期权相关的值. 论文使用均值方差法求得最优解 $(\delta^a,\delta^b,\tilde{\delta}^a,\tilde{\delta}^b)$, 其价值函数为

$$v(z,s,q,\tilde{q})=\sup_{(\delta^a,\delta^b,\tilde{\delta}^a,\tilde{\delta}^b):t\leqslant s\leqslant T}\left\{E_t(Z_T)-\gamma\mathrm{Var}_t(I_T)\right\},$$

其中 E_t 为给定 $(Z_t,S_t,q_t,\tilde{q}_t)=(z,s,q,\tilde{q})$ 时的条件期望.

3. **生存分析和 Kaplan-Meier 估计量** 正如 Ganchev 等 (2010, 第 103 页) 指出, 由于 "ℓ 大的尾部概率必须小于 c_i^t 处的真实概率", 因此 (7.25) 成立, 同时 (7.24) 也成立. 论文在证明 (7.24) 时, 使用了论文中的定理 2, 即对于任意 $\delta>0$, 任意 $i=1,\ldots,K$, 满足

$$P\left\{|\hat{T}_i^t(\ell)-T_i(\ell)|\leqslant \ell\sqrt{\frac{2\ln\frac{2v_{\max}}{\delta}}{N_{t,\ell-1}^i}}\quad \text{对于所有 } 1\leqslant\ell\leqslant\nu\right\}\geqslant 1-\delta. \tag{7.38}$$

在证明时, 论文首先考虑 $n_1^i,\ldots,n_t^i$ 独立同分布的情形, 并利用 Földes 和 Rejtö (1981) 实证的结果来求得边界值. Gu 和 Lai (1990) 对这些边界值进行了锐化, 并将其与鞅理论相结合, 从而得到了一个更好的 Lai 和 Ying (1991a) 类型的 Kaplan-Meier 统计量. Ganchev 等 (2010) 利用鞅理论以及 Azuma 不等式 (参见附录 A 中定理 A.5) 从而去掉了独立同分布的假设. 生存分析与 4.6.2 节和 5.3 节的密度模型也有比较紧密的联系, 且在信用评分和评级方面也有诸多应用, 具体细节请参见 Lai 和 Xing (2016).

4. **Ganchev 等 (2010, 5.1 节) 的参数生存模型及实证研究** 论文使用了美国一家券商在四个活跃的暗池中的数据. 每一个暗池集中在 12 个相对活跃的股票上, 这样就产生了 48 个不同的仓位数据集. 这些股票在所有交易所 (明面和暗池) 的日均交易量从 1 到 6000 万股不等. 为了研究

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

7.6.3 节中探索和开发方案的表现, 论文估计了对于每一个 i 的分布 U_t^i. Ganchev 等研究发现一个非常重要的特点: 绝大多数提交到暗池的交易 (84%) 的交易不会执行. 因此, $P\{U_i^t = 0\}$ 与 $P\{U_t^i > 0\}$ 的情形需要分别建模. 建模时使用了非参数 Kaplan-Meier 估计量, 并对于零膨胀均匀分布、泊松分布、指数分布以及幂定律模型等模型进行考虑. 其中零膨胀分布是混合分布, 且其中一部分是 0 点处的质点, 而幂定律模型即为 2.6.1 节中的幂定律模型. 研究发现, 零膨胀幂定律模型对数据的拟合效果最好. 论文同时比较了两种准则下开发–探索策略在订单执行概率和半衰期方面的差异.

5. **随机梯度法和 (7.19) 的变形** 在利用 (7.19) 进行递归时, 需要首先知道 F, 从而对 η_j^* 进行采样. Cont 和 Kukanov (2013, 4.1 节) 在没有假设 F 已知的情况下, 提出 F 在一定的情况下可以进行序贯估计, 且在 (7.19) 的非参数形式下用这些估计值来替代已知的 F. 具体而言, 论文对 $j = 1, \ldots, K$, 定义 $\dfrac{\partial c(\boldsymbol{\eta}, \boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} = \mathbf{g}(\boldsymbol{\eta}, \boldsymbol{\Theta}) = (g_1, \ldots, g_{K+1})$, 其中

$$\begin{aligned} g_1 &= h + f + \theta - (\lambda_u + \theta) I_{\{A(\boldsymbol{\eta}, \boldsymbol{\Theta}) < N\}} + \lambda_o I_{\{A(\boldsymbol{\eta}, \boldsymbol{\Theta}) > N\}}, \\ g_{j+1} &= \theta + I_{\{\eta_j > Q_j + L_j\}} \left[-(h + r_j) - (\lambda_u + \theta) I_{\{A(\boldsymbol{\eta}, \boldsymbol{\Theta}) < N\}} \right. \\ &\qquad \left. + \lambda_o I_{\{A(\boldsymbol{\eta}, \boldsymbol{\Theta}) > N\}} \right], \end{aligned}$$

因此, $\mathbf{g}(\boldsymbol{\eta}, \boldsymbol{\Theta})$ 只通过指示变量 $I_{\{A(\boldsymbol{\eta}, \boldsymbol{\Theta}) < N\}}$, $I_{\{A(\boldsymbol{\eta}, \boldsymbol{\Theta}) > N\}}$ 和 $I_{\{\eta_j > Q_j + L_j\}}$ 与 $\boldsymbol{\eta}$ 相关. 具体而言, $\{A(\boldsymbol{\eta}, \boldsymbol{\Theta}) < N\}$ 和 $\{A(\boldsymbol{\eta}, \boldsymbol{\Theta}) > N\}$ 分别对应于未达目标和超过目标的情形. $\{\eta_j > Q_j + L_j\}$ 指规模为 L_j 的限价订单在第 j 个交易场所 (其最高买入规模为 Q_j) 被完全执行的事件. 因此, 与前人的研究中使用指定的 F 来对于 η 建模不同, 论文提出使用现有累积数据来进行建模, 数据包括指示变量中的数据, 以及从数据集中有放回抽样的数据, 后者与 2.4.3 节中的 Bootstrap 抽样类似.

6. **随机逼近 (SA) 和补充材料 5、7.5.3 节、7.7.1 节的递归方法** SA 方法源起于 Robbins 和 Monro (1951) 的递归方法, 该论文称之为随机逼近, 用于求解回归方程 M 的唯一解 x_*:

$$y_t = M(x_t) + \varepsilon_t, \quad t = 1, 2, \ldots, \tag{7.39}$$

其中 ε_t 表示无法观测的随机误差. 假设 M 是一个平滑函数, 且 $M(x_*) \neq$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

0. 在确定性情况 $\varepsilon_t \equiv 0$ 下, 利用牛顿法 $x_{t+1} = x_t - \dfrac{y_t}{M'(x_t)}$, 可得

$$x_{t+1} = x_t - \frac{M(x_t)}{M'(x_t)} - \frac{\varepsilon_t}{M'(x_t)}. \tag{7.40}$$

因此, 如果 x_t 收敛于 x_*, 则有 $M(x_t) \to 0$ 和 $M'(x_t) \to M'(x_*)$, 此时 (7.40) 需满足 $\varepsilon_t \to 0$, 而该条件在大多数的随机误差中并不成立 (如 ε_t 独立同分布, 且方差为正时). 为平均误差项 ε_t, Robbins 和 Monro (1951) 提出了如下的递归方法:

$$x_{t+1} = x_t - \gamma_t y_t, \tag{7.41}$$

其中 γ_t 是正的常数, 满足 $\sum_{t=1}^{\infty} \gamma_t^2 < \infty$ 和 $\sum_{t=1}^{\infty} \gamma_t = \infty$. 假设当 $x > x_*$ 时, $M(x) > 0$, 以及当 $x < x_*$ 时 $M(x) < 0$. 此时在 $\sum_{t=1}^{\infty} \gamma_t^2 < \infty$ 情况下, 在很多的噪声项的随机模型中 (如 ε_t 独立同分布), $\sum_{t=1}^{\infty} \gamma_t \varepsilon_t$ 在 L_2 中几乎处处收敛. 在正则性条件下, 该结果反过来意味着 $x_t - x_*$ 在 L_2 中几乎处处收敛. 另外, 在 $\sum_{t=1}^{\infty} \gamma_t = \infty$ 情况下, $x_t - x_*$ 的极限为 0. 除了求解 $M(x)$ 的根, SA 还可以用递归估计回归方程的极大值. 该方法最初由 Kiefer 和 Wolfowitz (1952) 提出, 选取满足 $x_t^{(l)} = x_t - \delta_t$ 和 $x_t^{(r)} = x_t + \delta_t$ 的 $x_t^{(l)}$ 和 $x_t^{(r)}$, 并基于下式进行递归:

$$x_{t+1} = x_t + \gamma_t \frac{y_t^{(r)} - y_t^{(l)}}{2\delta_t}, \tag{7.42}$$

满足 $\lim\limits_{t\to\infty} \delta_t = 0$. 为了消除 $\varepsilon_t^{(r)}$ 和 $\varepsilon_t^{(l)}$ 在

$$\frac{y_t^{(r)} - y_t^{(l)}}{2\delta_t} = \frac{M(x_t^{(r)}) - M(x_t^{(l)})}{2\delta_t} + \frac{\varepsilon_t^{(r)} - \varepsilon_t^{(l)}}{2\delta_t}$$

中的影响, Kiefer 和 Wolfowitz 在求解时, 选取了满足 $\sum_{t=1}^{\infty} \left(\dfrac{\gamma_t}{\delta_t}\right)^2 < \infty$, $\sum_{t=1}^{\infty} \gamma_t = \infty$ 和 $\sum_{t=1}^{\infty} \gamma_t \delta_t < \infty$ 的 $\gamma_t > 0$. Blum (1954) 已提供了 Robbins-Monro 和 Kiefer-Wolfowitz 框架的多元形式. Lai (2003) 对 SA 进行了综述, 并对 SA 方法如何演化为 "优化、控制和信号处理领域成为重要且充满活力的话题" 进行了梳理.

SA 递归 (7.31) 在方程组 (7.30) 中, $(\nu_t, U_t^1, \ldots, U_t^K)$ 为独立同分布, $(\nu, U_1, \ldots, U_K)$ 的分布也相同. 方程组也可以改写成 $\boldsymbol{M}(\boldsymbol{r}) = 0$, 其中

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$\boldsymbol{r}=(r_1,\ldots,r_K)\in\mathbb{S}$, $\boldsymbol{M}(\boldsymbol{r})=(M_1(\boldsymbol{r}),\ldots,M_K(\boldsymbol{r}))$, 且

$$M_j(\boldsymbol{r})=E\left[\nu\left(\rho_j I_{\{r_j\nu\leqslant U_j\}}-\frac{1}{K}\sum_{i=1}^{K}\rho_i I_{\{r_i\nu\leqslant U_i\}}\right)\right],\quad j=1,\ldots,K. \tag{7.43}$$

定义 $\boldsymbol{r}_t=(r_t^1,\ldots,r_t^K)$, $\boldsymbol{y}_t=(y_t^1,\ldots,y_t^K)$, 其中

$$y_t^j=\nu_t\left(\rho_j I_{\{r_t^j\nu_t\leqslant U_t^j)\}}-\frac{1}{K}\sum_{i=1}^{K}\rho_i I_{\{r_t^i\nu_t\leqslant U_t^i\}}\right),\quad j=1,\ldots,K, \tag{7.44}$$

则可得如下回归模型

$$\boldsymbol{y}_t=\boldsymbol{M}(\boldsymbol{r}_t)+\boldsymbol{\varepsilon}_t, \tag{7.45}$$

其中 $\boldsymbol{\varepsilon}_t=(\varepsilon_t^1,\ldots,\varepsilon_t^K)$ 代表零均值随机误差. 在此情况下, SA 递归过程 (7.41) 的形式为 $\boldsymbol{r}_{t+1}=\boldsymbol{r}_t+\gamma_t\boldsymbol{y}_t$, 但此时可能并不满足约束条件 $\boldsymbol{r}_t\in\mathbb{S}$. Laruelle 等 (2011) 通过重新定义回归函数

$$M_j(\boldsymbol{r})=E\left[\nu\left(\rho_j I_{\{r_j\nu\leqslant U_j\}}-\frac{1}{K}\sum_{i=1}^{K}\rho_i I_{\{r_i\nu\leqslant U_i\}}+\xi_j\right)\right] \tag{7.46}$$

来消除这个约束, 其中 $j=1,\ldots,K$, ξ_j 由 (7.36) 定义. $(\nu_t,U_t^1,\ldots,U_t^K)$ 的独立同分布假设是为了将 (7.30) 与回归模型 (7.45)–(7.46) 相联系. 事实上, 收敛理论也适用于满足如下条件的随机回归模型: ε_t 是鞅差分序列, 且 $\sup_n(\varepsilon_n^2|\mathcal{F}_{n-1})<\infty$ 几乎处处成立. Blum (1954) 对递归方程 (7.40) 的鞅结构进行了研究. Lai (2003) 的第 3 节也对 Robbins-Siegmund 引理进行了介绍, 该引理最初在 Laruelle 等 (2011, 第 1059 页) 中用于将独立同分布假设拓展到遍历的 $(\nu_t,U_t^1,\ldots,U_t^K)$ 过程中, 后者已经在 7.7.3 节的最后一段进行了介绍.

SA 递归 (7.19) 或者其非参数变异　Cont 和 Kukanov (2013, 4.1 节) 针对 (7.19), 指出 "最优递归步长" 为 $\gamma_m=C/m$, 其中

$$C=\max_{\boldsymbol{\Theta},\boldsymbol{\Theta}'\in\mathbb{S}}\|\boldsymbol{\Theta}-\boldsymbol{\Theta}'\|\Bigg/\left(\max_{\boldsymbol{\Theta}\in\mathbb{S}}E\left[\|\mathbf{g}(\boldsymbol{\eta},\boldsymbol{\Theta})\|\right]\right)^{1/2},$$

$$\mathbb{S}=\left\{\boldsymbol{\Theta}\in\mathbb{R}_+^{K+1}:M\leqslant N,M+\sum_{i=1}^{K}L_i\geqslant N,\right.$$
$$\left.L_i\leqslant N-M\quad\text{对于 } i=1,\ldots,K\right\}.$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

然而, 论文并没有解释最优选择的直观意义. 注意到 $\sum_{m=1}^{\infty}\gamma_m=\infty$ 和 $\sum_{m=1}^{\infty}\gamma_m^2=\infty$, 满足 SA 递归几乎处处收敛的准则. 论文接下来提出使用另外的步长 $\gamma_m=C'm^{-1/2}$, 其中 C' 为一个常数, 且该常数 "可以根据参数来进行比例调整". 然而, 在此情况下 $\sum_{m=1}^{\infty}\gamma_m^2=\infty$, 且论文并没有对这一 "可尺度化" 选择的 (如果存在) 收敛结果进行阐述. 在论文中, 递归步长定义为 γ, 且对 "最优" 和 "可尺度化" 模型中的 K, m 和其他参数进行了定义, 但是并没有对 m 在收敛中的作用进行讨论. Lai (2003) 的第 4 节对 SA 中 γ_m 的渐近最优选择进行了梳理. Powell (2007) 的第 6 章对 SA 中步长的选择和近似动态规划问题中更一般的随机梯度算法进行了进一步讨论.

7.9 习题

1. 考虑如下的做市问题, 其中间价格 $S(t)$ 对于特定的 $b(\cdot,\cdot)$ 和 $\sigma(\cdot,\cdot)$, 满足如下过程:

$$dS(t)=b(t,S(t))dt+\sigma(t,S(t))dW(t),$$

使得该问题有唯一的解. 对于特定的 ϕ, 令

$$u(t,s,q,x)=\max_{\delta^a,\delta^b}E_{t,s,q,x}\left\{\phi\left[S(T),Q(T),X(T)\right]\right\}.$$

利用动态规划准则, 求解关于 $u(t,s,q,x)$ 的 HJB 方程

$$\begin{aligned}(\partial_t+\mathbb{L})u+\max_{\delta^+}\lambda^+(\delta^+)\left\{u\left[t,s,q-1,x+(s+\delta^+)\right]-u(t,s,q,x)\right\}\\+\max_{\delta^-}\lambda^-(\delta^-)\left\{u\left[t,s,q+1,x-(s-\delta^-)\right]-u(t,s,q,x)\right\}=0,\end{aligned}$$

其边界条件 $u(T,s,q,x)=\phi(s,q,x)$. 此处 $\mathbb{L}=b(t,x)\partial_s+(2\sigma^2(t,x))^{-1}\partial_{ss}$, 其中 $\partial_s=\partial/\partial s$, $\partial_{ss}=\partial^2/\partial s^2$.

2. (a) 令 X 为 $\mathbb{Z}_+$ 上的随机变量, 满足 $E(X)<\infty$. 证明

$$E(X)=\sum_{s=1}^{\infty}P(X\geqslant s).$$

(b) 使用 (a) 或其他结果, 证明 7.6.1 节中的

$$E[\min(n,U_t^i)]=\sum_{s=1}^{n}T_i(s).$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

(c) 利用归纳法证明 (7.20) 的最优解可以由 7.6.1 节中的贪心算法求得.

3. (a) 类似于 7.7 节, 定义 $\varphi(r) = E[\min(r\nu, U)]$. 证明当 $E[\nu I_{\{U>0\}}] > 0$ 且分布函数 $\dfrac{U}{\nu}$ 在 $\mathbb{R}_+$ 上连续时, $\varphi(r)$ 在 $[0,1]$ 上处处可微.

   (b) 在 (a) 的条件下, 对于所有 $r \in [0,1]$, 证明 $\varphi'(r) = E[\nu I_{\{r\nu \leqslant D\}}]$ 成立.

   (c) 证明 $\varphi(r)$ 可以扩展到整个实数域, 且当其定义如下时, 始终为凹函数:

$$\varphi(r) = \begin{cases} \left(r - \dfrac{r^2}{2}\right)\varphi'(0), & \text{当 } r < 0, \\ \varphi(1) + \varphi'(1)\log r & \text{当 } r > 1. \end{cases}$$

   上述拓展与 (7.36) 中 ξ_j 的关系是什么?

4. 定义 X 是一个均值为 μ 的随机变量. 定义 $x_1, x_2, \ldots$ 为 X 上的独立同分布序列. 考虑满足 $\mu_1 = x_1$, $i \geqslant 2$ 的如下序列:

$$\mu_i = (1-\alpha_i)\mu_{i-1} + \alpha_i x_i = \mu_{i-1} + \alpha_i(x_i - \mu_{i-1}).$$

   其中 α_i 是步长参数. 当满足 $\sum_{i=2}^{\infty} \alpha_i = \infty$ 和 $\sum_{i=2}^{\infty} \alpha_i^2 < \infty$ 时, 证明 $\mu_i \to \mu$ 几乎处处成立.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

第八章　信息学、监管和风险管理

信息学和基础设施是典型的自动交易系统的组成部分, 特别在软件和硬件方面. 从广义上讲, 它们可以分为两大支柱: 策略和交易. 策略支柱是指在交易主体的控制下, 由主体开发和维护的部分. 交易策略对市场事件作出反应, 如价格变化或者消息到达, 并根据这些事件的信息内容做出交易决策. 交易支柱是指交易领域中的部分, 包括处理传入订单请求、验证订单属性、匹配买卖双方的订单, 最后向市场数据订阅者发送匹配结果. 从基础设施的角度来看, 这两个支柱通过一些标准的传输协议来连接, 从而促进了它们之间的联系. 图 8.1 阐述了两大支柱的关系.

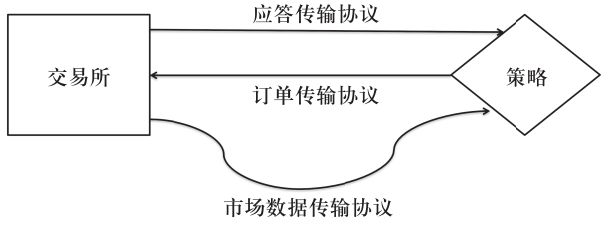


图 8.1　交易支柱和策略支柱的图解

在电子交易和自动交易出现之前, 证券交易主要是通过三种机制进行的: (1) 专属做市商, 如纽约证券交易所的经纪人; (2) 交易池, 如芝加哥商品交易所期货交易池; (3) 场外交易市场, 如利率互换交易的经济人市场. 在交易池演化为电子交易平台之后, 专家、池内交易员以及场外经纪商和交易员的作用都是通过自动交易算法进行的. 传统的经纪人角色被限价订单和匹配引擎所取代, 许多经销商和专家则被能够在分秒内发布各种类型的订单的算法所代替. 尽管

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

如此, 在场外交易中, 经纪人帮助初始交易商匿名寻找交易对手的角色仍然存在,[①]特别是针对非标准利率互换等专业或结构性产品.

一个典型的电子交易, 所有被批准的参与者可以通过网关提交市价交易、限价交易或取消交易的订单. 基于订单类型, 交易平台可以处理取消订单, 或者直接匹配市价和限价订单. 在这样一个过程之后, 平台将通过网关发送确认消息, 如果成功完成相应的订单交易, 则通知订单发送方. 例如, 如果计划取消的限价订单在取消订单到达平台之前被执行, 那么取消订单的发送者将收到拒绝的回复. 当订单成功执行时, 更新后的限价订单将通过交易平台的广播系统向所有市场数据用户发送. 则清算由第三方清算所完成. 在交易中, 交易双方提供担保品以满足初始和维持保证金的要求, 提供担保的保证金金额和特殊的跨市场保证金的特许是由交易所规定的. 如今, 在大多数交易所, 大部分的日常交易都是由某种自动交易算法所完成的, 它可以作为一种人为发出指令的执行引擎, 或可以作为一种完全自动化的交易算法, 在触发事件后生成买卖指令.

在对基础设施讨论时, 8.1 节回顾了一些在第 1–3 章和高频交易策略中的定量策略. 其中, 我们特别解释了为什么对交易和策略支柱的理解是实现这些自动化算法策略的关键. 8.2 节以纳斯达克和芝加哥商品交易所为例, 介绍了交易所基础设施的概况, 并讨论了匹配引擎的作用、一些常用的订单类型的匹配机制和一些市场数据协议. 交易策略方面的基础设施和信息学在 8.3 节介绍. 8.4 节介绍了美国交易所的一些现行规定和交易规则. 8.5 节讨论了管理操作、建模和交易风险的一些问题. 有关主题和当代发展的补充材料在 8.6 节中列出. 8.7 节提供了一些练习.

8.1 量化策略

广义上讲, 算法和系统的量化交易可以分为相对价值交易、统计套利和延迟套利. 尽管统计套利之类的术语有时被用来描述基于统计方法的通用建模范式, 但在从业者中, 下列术语已被合理地标准化, 以表示特定的交易方式.

相对价值交易 如 1.2 节所讨论的, 相对价值交易通常是指固定收益中的一种交易方式, 其交易策略是基于分析收益率曲线不同部分之间的历史关系, 参见 Lai 和 Xing (2008) 的第 10 章——收益率曲线分析和利率市场的介绍. 这些策略会找出曲线上的错位, 并根据均值回归的速度和概率来进行定位. 通

① 场外交易市场的参与者通常会以一种成对的方式进行谈判. 没有 LOB 信息, 除主要经销商以外的其他参与者没有关于目前市场上其他地方可用的价格信息. 因此, 由于缺乏价格信息, OTC 市场有时被称为 “黑市”; 参见 Duffie (2012) 对这一市场的匹配过程、资产定价和信息动态进行的定量分析.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

常, 一种交易由多个子部分组成, 即有着不同期限的 “支节”. 通常, 低频相对价值策略依赖于处理历史债券收益率或利率互换数据来确定长期稳定关系的模型, 并在这些关系偏离阈值水平时建立头寸. 高频相对价值策略也有类似的情况, 受暂时供给和需求的可比较因素驱动的影响. 这些策略的实质是为模型拟合找到合适的数据, 并确定在交易中进出的最佳时机. 固定收益相对价值策略中使用的典型工具是主权债券、利率互换和通胀挂钩债券. 期权, 如互换期权、互换期权的上限和下限, 也经常被用来表达相对价值交易的条件视图. 除固定收益之外, 相对价值交易描述了一种一般的交易方式, 即由一只股票或一篮子股票对其他股票的相对价格分析驱动. 这种交易方式既可以是低频, 也可以是高频.

统计套利 统计套利有时指的是 Markowitz 的投资组合优化和其变形形式的统计基础. 在 2.3–2.5 和 2.7 节中, 我们已经研究了这种风格的交易方式从经典方法到最近的一些发展. 2.8 节还描述了其他交易方式. 此外, 第三章分析了将交易成本考虑在内的多周期投资组合优化问题. 第三章同时讨论了采用随机控制理论、近似动态规划、滤波器、大数据分析和机器学习的现代方法. 还有一些非参数的方法, 如对数–最优策略和预期效用理论, 它们代替了经典公式中的均方投资组合优化理论. 这种交易方式既可以是低频, 也可以是高频; 关于如何使用高频数据交易一篮子股票, 请参见 Fan 等 (2012). 统计套利也被用作一种通用术语, 用来描述使用一些统计分析的量化交易.

延迟套利 这指的是高度相关的、有时甚至是等价的交易资产. 例如, ETF SPY 和 E-mini S&P 500 期货合约 (由芝加哥商品交易所的 ES 代码所表示) 都基于标准普尔指数, 忽略了股息和利率部分. 然而, ES 在芝加哥的芝加哥商品交易所交易平台上交易, 而 SPY 在纽约市内和周边的几家交易所的交易平台上进行交易. 当 ES 上升时, SPY 也应该上升, 尽管会有拖延. 这种延迟是套利者可以操作的速度的函数. 这种延迟可能在几十毫秒的范围内, 主要是由于两个交易所的地理差距: 芝加哥和纽约之间的距离大约是 714 英里, 光速为每小时 670616629 英里, 如果用一束光将两者连接起来, 需要花费约 3.8 毫秒[①]. 因此, SPY 和 ES 之间存在相当大的延迟. 延迟套利策略将考虑这类机会, 它们作为两个交易中心之间信息传递的促进者. 另一种类型的延迟套利与可分解为其他交易部分的可交易资产有关. 在一些期货交易所, A, B, C 之间的直接

① 在这一章节毫秒单位缩写为 ms.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

合同、利差[①]{A-B}, {B-C} 和捆绑包[②]{A, B, C} 都是可以交易的. 欧洲美元期货就是一个例子, 在这种情况下, 利差、捆绑包与其组合的交易都是活跃的. 高频套利者监测这些可交易资产与普通成分之间的价格一致性, 并在短期不一致的情况下套利.

所有这些方法在业界都很受欢迎. 要将它们作为自动交易算法实现, 对交易基础设施的理解是绝对必要的. 例如, 尽管相对价值交易可以应用于固定收益的证券和股票, 但它们的匹配规则却非常不同. 对于用于固定收益匹配的按比例规则的详细信息, 以及股票交易平台上常用的价格–时间优先规则, 参见 8.2.2 节. 高频交易的货币交易方法需要监控许多独立平台的 LOB. 此外, 所有的量化交易实现都需要对市场数据进行大量研究. 对市场数据进行存储是重要的, 这样, 在模拟市场节奏的模拟环境中, 该策略可以进行后验测试.

8.2 交易基础设施

表 8.1 列出了一些主要的美国交易所以及它们最受欢迎的股票和衍生产品的每日成交量和持仓量[③]. 表 8.2 列出了芝加哥商品交易所的十大算法交

表 8.1 2015 年 11 月 29 日, 美国主要交易所、其主要交易产品及持仓量或日交易量. 例如, 那天苹果公司的交易量是 72056345 股, 而 ES 的持仓量是 2920713 份合约

交易所	主要交易产品	简称	持仓量/日交易量
芝加哥商品交易所	Eurodollar Futures	ED	11089332
	E-mini S&P Futures	ES	2920713
	10-Year T-Note Futures	ZN	2764946
纳斯达克	Apple	AAPL	72056345
	PowerShares QQQ Trust	QQQ	46277773
	Facebook	FB	41294632
纽约证券交易所	Bank of America	BAC	79806275
	Sprint Corp	S	47012913
	Ford Motor Company	F	44414913

来源: NASDAQ OMX,CME.

① 买入一份合同利差 {A-B}, 意味着买入一份合同 A, 同时卖出一份合同 B, 价格根据 A 与 B 之间的差价而定. 因此价差可以为负值, 有关完全市场和分散市场的更多讨论, 请参见 8.2.1 节.

② 购买一个捆绑包 {A, B, C} 通常意味着同时购买 A,B,C 的合同各一份, 其价格是根据这些资产之前收盘价的平均价格变化来计算的.

③ 未平仓量指的是在某一交易日结束时未被关闭或交付的期权和/或期货合约的总数.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

表 8.2 在 CME Globex 上的十大算法交易期货类别占总量的百分比. 数字是基于交易指示器 (FIX Tag 1028). “自动” 指令由计算机算法提交, 无需人工干预; “人工” 指令是由人工干预所提交的交易; “其他” 指令包括在场内或场外进行的交易. 请参考 Haynes 和 Roberts (2015) 在 2012 年 11 月 12 日至 2014 年 10 月 31 日的芝加哥商品交易所交易数据集的完整分类和描述.

产品群/子群名称	自动	人工	其他
FX/G10	80.7	16.4	2.8
FX/Emerging Markets	70.3	17.1	12.5
Equities/International Index	69.3	30.3	0.4
Equities/Equity Index	66.5	32.6	0.9
FX/E Micros	66.3	33.7	0.0
Interest Rate/US Treasury	64.0	30.2	5.9
Interest Rate/STIRS	60.3	37.5	2.2
Equities/Select Sector Index	59.2	22.0	18.8
Energy/Crude Oil	54.3	40.2	5.5
Metal/Base	49.2	47.7	3.1

易期货. 要处理如此庞大的交易量, 交易所必须配备强大的电子平台, 处理入站订单, 匹配买方和卖方的订单, 并通过对外渠道向所有市场数据的订户发送最新的 LOB. 概言之, 交易基础设施含有三个主要部分: 订单网关、匹配引擎和市场数据广播引擎, 如图 8.2 所示.

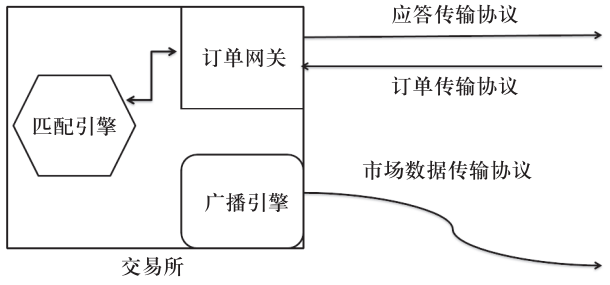


图 8.2 典型交易平台的组件

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

8.2.1　订单网关

一个真正的交易平台显然要比图 8.2 中所示的简化模型复杂得多, 而订单网关通常设在一个名为订单路由接口的系统中. 对芝加哥商品交易所的交易平台 Globex 来说, 这一接口被称为 *iLink*; iLink 原理图为图 8.3. 图顶部的方框代表交易公司和策略过程, 如询价引擎、Alpha 引擎和订单管理引擎, 这些都是交易公司的组成部分, 这将在 8.3 节进一步讨论. 市场参与者通过称为会话的节点连接到交易所, 该节点唯一标识参与者. 订单通过特定的会话发送到订单网关, 该网关被表示为图的第二行. 客户可以有多个会话, 而交易通常有多个订单网关来处理入站订单消息. 匹配引擎, 即买卖订单的实际匹配, 由图底部的方框表示. 每个网关处理和验证来自市场参与者的入站订单信息, 然后根据特定的规则路由订单. 例如, 在芝加哥商品交易所 iLink 订单消息框架下, 有多个订单网关, 每个订单都独立工作, 根据订单的产品组对客户订单进行解码和路由. 这些产品组是根据期货或期权合约的标的资产来定义的. 例如, 西德克萨

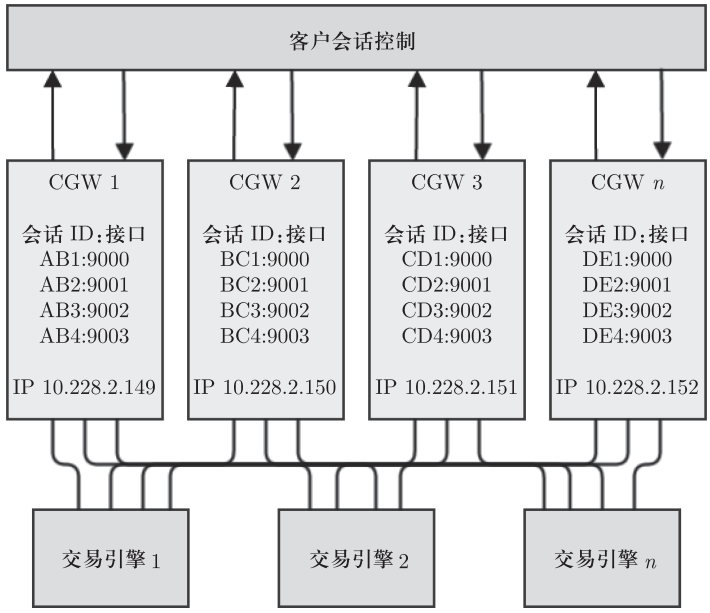


图 8.3　iLink 订单路由接口的原理图和 CME Globex 的匹配引擎. CGW1, CGW2, CGW3, CGW4 是 4 个网关节点; 交易引擎 1, 交易引擎 2, 交易引擎 3 是 3 个匹配的引擎. CGW1 正在处理 4 个会话: AB1:9000, AB2:9001, . . ., AB4:9003, 其中 9000 到 9003 是会话连接的端口. 会话 AB1 只能在一个网关上存在; AB1 有一个单一的序列流, 由来往所有市场的消息组成 (在 CME 的许可下使用)

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

斯中质原油期货 (简称 WTI) 和布伦特原油期货将被路由到相同的匹配引擎, 而小麦和玉米期货将被路由到另一个.

一旦网关解析并验证了传入订单的基本属性, 它将通过出站通道发送确认消息, 如图 8.2 所示, 声明该订单是被交易所接受或拒绝. 如果订单被接受, 它将被提交给匹配引擎. 这样的确认是由策略方的订单管理引擎接收, 然后将订单回复事件传递给 Alpha 引擎, 用于策略级处理, 如 8.3 节所述.

8.2.2 匹配引擎

匹配引擎接受订单网关路由的订单, 并根据某些预先确定的算法进行买卖订单的实际匹配. 当一个市场订单或有市场限制的订单 (这被统称为*侵略*或*攻击*订单) 到来时, 匹配引擎会根据匹配规则确定买卖的数量. 在美国交易所常用的匹配算法如下.

- 先进先出 (First-In-First-Out): 最常用的算法是用价格和时间来填写订单, 所以在同一价格按时间顺序下订单; 也被称为价格时间优先规则, 这在 5.5.2 节中已经讨论过了. 大多数美国股票交易所根据这种算法匹配订单.

- 按比例分配: 该算法根据价格、订单大小和时间优先级来填写订单, 在队列中的每个订单都被填写为相对于整个队列大小的订单百分比函数, 并进行适当的四舍五入处理. 由于初始分配的交易数量的减少, 多余的数量会通过先进先出法分配.

- 分配: 这是一种按比例分配算法的修正, 是在收到订单的时候, 就会有一个优于市价的价格的订单. 在匹配的过程中, 在订单中建立更激进的价格等级的订单被给予了时间优先级. 这一算法的微小变化目前被用于欧洲美元期货, 这是 CME Globex 上最具流动性的期货之一.

对于芝加哥商品交易所和类似的交易所来说, 一个或多个资产之间的价差交易本身可作为一种资产, ES 即为解释完全市场和价差市场的概念. 特别是在 2016 年 6 月到期的 ESM6, 是一个典型的完全市场例子. 价差市场指的是由各种设备组成的多支工具. 例如, ESM6-ESU6 由两条*关节* ESM6 和 ESU6 组成,ESU6 是指在 2016 年 9 月到期的 ES. 价差市场的报价惯例是 ESM6 的价格减去 ESU6 的价格, 因此价差市场的价格可能是负的. 价差交易代表两种或更完全的期货工具同时买卖. 例如, 购买 ESM6-ESU6 价差意味着购买一份

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

ESM6 合同, 并销售一份 ESU6 合同. 在这一特例中, 在执行时, 交换器将发送三个执行报告, 一个用于价差, 另一个用于两个关节. 由于交易是在价差中进行的, 因此, 两个单独关节的价格需要通过一些交易所定义的规则来确定, 这样, 关节的价格就会被分配到接近标的资产或指数的价格.

8.2.3 市场数据传播

每当交易平台上有任何一个订单事件发生, 无论是取消订单、执行订单或增加订单, 交易市场数据传播系统通过网络或一些指定的结构向市场数据用户播放这个 LOB 相应的改变. 由于 LOB 事件发生频率非常高, 网络资源的最优使用是在最短的时间内将数据交付给每个订阅者, 同时避免网络拥塞是一个非常重要的任务. 为了实现高速广播和产品频谱与用户数量的可扩展性, 许多交易在广播其市场数据时使用多播技术而不是传统的单播方式作为网络数据包, 参见 8.6 节中的补充材料 2.[①]

除了考虑网络应用外, 交易所还需要选择市场数据传输协议——编码和解码订单信息的方案——来最大限度地适应它们的市场、匹配引擎基础设施和数据的目标用户. 两种流行的协议是 FIX/FAST 和 ITCH. FIX/FAST 协议是 FIX 公司开发的一种技术标准, 专门针对大容量和低延迟的市场数据流. 目前, 该协议已被纽约证券交易所和欧洲期货交易所等交易所采用. ITCH 是一开始由纳斯达克开发和采用的对外直接存取市场数据协议; ITCH 也被其他交易所和暗池使用, 如伦敦证券交易所和 Chi-X 公司.

芝加哥商品交易所的广播系统称为芝加哥商品交易所集团市场数据平台 (MDP), 它使用自己的协议 MDP 3.0.[②]这个协议包括 (i) 简单的二进制编码 (SBE) 和 (ii) 事件驱动的消息传递. (i) 和 (ii) 两者都有能力在单个包中交付多个事件, 以实现在带宽利用方面的高效率. 图 8.4 阐述了一个 MDP 3.0 包, 其中包含两个提供事件数据的 FIX 消息. MDP 3.0 还允许一个单一事件通过多个包来交付. 当这种情况发生时, MDP 3.0 确保策略侧解析过程[③]可以在收到第一个包时启动. 关于 MDP 3.0 包的更多细节, 请查看 8.6 节中的补充材料 3 和 4.

---

① 如果在传输过程中出现数据包丢失, 单播协议通常可以检测到这种丢失并自动执行重传, 而多播协议可能需要订阅者付出更多的努力来管理这种情况.

② CME MDP 3.0 是最新的协议. 参阅协议和消息规范的官方文档, 网址: `http://www.cmegroup.com/confluence/display/EPICSANDBOX/CME+MDP+3.0+Market+Data`. 请注意, 除了 MDP 3.0 之外, MDP 还对不可操作的价格数据提供 "流线型固定/快递" 端口支持, 详见官方文件.

③ 解析过程是将网络数据包转换为数据的计算机子程序.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

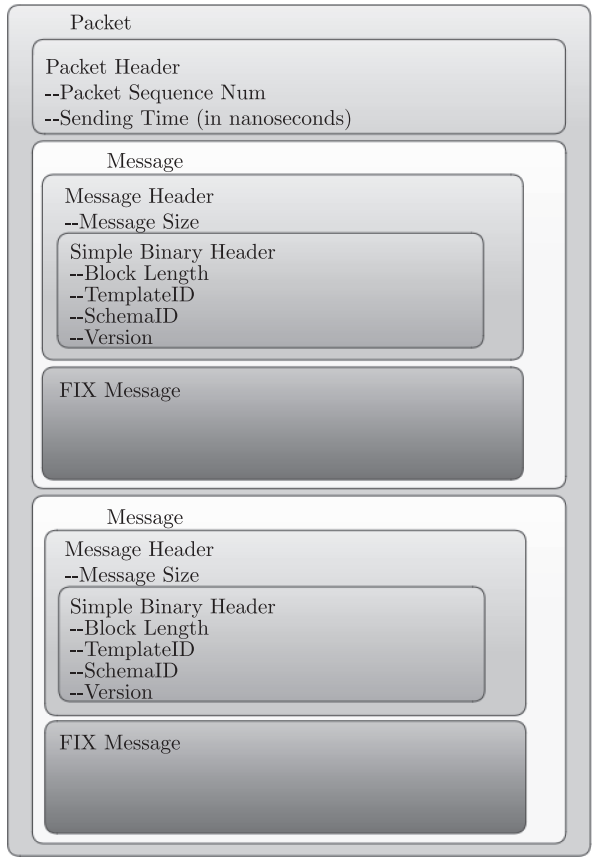


图 8.4 包含两个 FIX 消息的 CME MDP3.0 包 (经 CME 许可使用)

8.2.4 订单的收费结构

交易所的收费结构因地点、资产类别、订单类型和汇总成交量 (TCV) 而异.[①]表 8.3 总结了美国一些代表性交易所收取的费用. 大多数交易所对那些会消耗流动性的订单收取一定费用, 并向那些增加流动性的公司提供回扣. 这种费用回扣结构的基本原理是促进订单簿的流动性, 并补偿做市商的仓储风险.

然而, 在一些交易所, 如波士顿证券交易所 (缩写为 BX), 这种结构发生了逆转: 流动性接受者得到了回扣, 而流动性提供者则收取了寄送订单的费用. 从经济学的直觉来说, 那些不希望交叉买卖价差但仍想获得订单队列优势的市场参与者可能愿意支付一定的费用, 因为他们能够靠近订单队列的前端.

① TCV 是用所有交易所的交易量和交易报告计算出来的, 含有费用的交易报告为当月的统一交易报告计划.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

表 8.3 2015 年 10 月 5 日的收费结构. 括号里的费用意味着向流动性提供者提供回扣: 纽约证券交易所收取和提供显示流动性的费用; 纳斯达克通过单一的 MPID 获取流动性, 并提供显示指定的零售流动性; 从公司拿走至少 0.20% 的 TCV, 并提供显示和非显示的流动性, 从公司获得流动性的回扣; CME 作为非交易所成员, 为 E-mini 标准普尔承担和增加流动性

交易所	收取	提供显示	提供非显示
纽约证券交易所	$0.00270	$(0.00140)	$(0.00000)
纳斯达克	$0.00295	$(0.00340)	$(0.00250)
波士顿股票交易所	$(0.0017)	$0.00140	$0.00240
芝加哥商品交易所	$0.40000	–	–

8.2.5 主机托管服务

对于高频交易来说, 几微秒就能产生重大影响. 因此, 策略部分驻留的服务器盒到交易数据中心 (匹配过程发生和市场数据的来源之处) 的位置是影响策略整体延迟的一个重要因素. “主机托管” 指的是将交易服务器放置在或靠近交易数据中心的做法. 例如, 芝加哥商品交易所在伊利诺伊州的奥罗拉市芝加哥商品交易所集团中心和在 350 E. Cermak 的芝加哥市中心均提供了主机托管服务. 纳斯达克也在新泽西州中心卡特雷市提供类似的托管服务, 同时也提供了卡特雷和奥罗拉之间的微波连接.[①]提供主机托管服务的交易被要求提供与所有参与者同等的访问权限. 具体来说, 将所有服务器连接到网关的物理电缆都要求相同的长度, 因此所有的主机托管客户机都受到相同的入站和出站延迟. 对于托管服务, 交易所收取的费用取决于客户端服务器所需要的空间和功耗.

8.2.6 清算和结算

结算和清算指的是在买方和卖方订单匹配后所引发的后交易过程. 清算所作为结算机构将集中对交易对手进行清算, 在原始的买方和卖方之间进行交易, 并承担着原合同双方今后结算的责任. 这一约务更新有效地用两份独立合同替代了一份合同, 集中的交易对手方面对的是原来的买卖双方. 对于芝加哥商品交易所来说, 清算所的定义是一个提供结算和清算服务的机构, 它由那些将资金存入清算所以满足交易和保证金要求的成员组成. 其他交易所也使用类似的定义. 清算所, 作为集中交易对手, 有助于减少违约风险, 并要求成员保持足够的抵押存款来弥补潜在损失, 通过提供资金来吸收多余的抵押品在发生违

① 根据纳斯达克的数据, 微波连接的单向延迟大约是 4.13 毫秒, 而在最快的光纤线路上则是 6.65 毫秒.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

约时减少经济影响. 清算所还通过在多个交易对手之间的抵销交易, 帮助提高资本效率. 清算成员代表他们自己的账户和客户的交易, 并执行与交易相关的服务, 如基于交易量保证金计算和库存跟踪收费.

8.3 策略信息和基础设施

策略信息学和基础设施处理交易公司的流程和程序. 它有三个主要的处理单元: 市场数据处理、Alpha 引擎处理和订单管理. 每个单元负责整个过程中的一个特定任务. 图 8.5 说明了典型的策略基础结构.

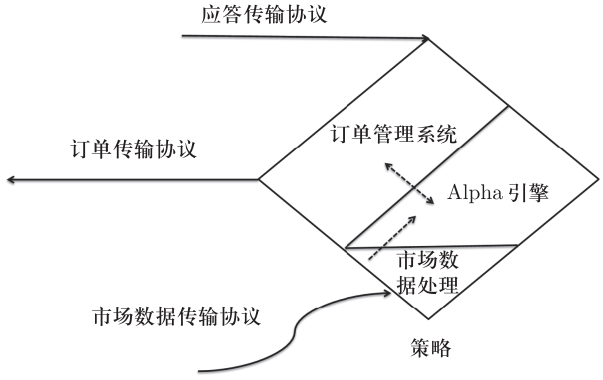


图 8.5 典型策略基础设施的主要处理单元. 虚线箭头表示三个处理单元之间信息流的方向

8.3.1 市场数据处理

市场数据处理单元监测市场数据广播通道, 解码原始市场数据, 更新订单状态和处理交易信息. 市场数据有两大主要来源[①]: 报价和频率. 报价数据包含与订单状态有关的信息, 例如订单的每一层的价格、交易量和订单数量. 频率数据包含与执行有关的信息. 市场数据处理单元将市场数据从线路格式解码到与订单和执行状态相关的特定信息. 解码信息集由一般的订单事件属性组成, 如时间标记、事件类型、符号、价格、大小. 根据消息协议, 它还可以包含诸如价格的类型[②]、执行的匹配序列号等信息. 使用解码后的消息, 该单元将更新订单.

① 存在与匹配过程相关的其他数据流, 比如拍卖状态与开放和结束的不平衡, 但市场数据的大部分都与报价和频率有关.

② 对于在芝加哥商品交易所平台上交易的一些期货合约, 可交易价格可能是直接价格或隐含价格.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

目前, 两种主要的引用数据类型分别是基于顺序的和基于级别的. 一些交易所, 比如纳斯达克, 提供了两种类型的报价流, 而另一些交易所, 比如芝加哥商品交易所, 只提供了基于级别的报价并称为级别驱动型交易, 请参见 4.1 节. 基于顺序的流允许重新构建订单, 该订单包含每个价格级别的特定订单的订单细节, 因此该策略可以跟踪订单的生命周期, 并在新订单添加和现有订单执行或取消时, 推断准确的队列转换. 另一方面, 基于级别的市场数据只提供受影响的订单级别的快照更新. 例如, 在订单取消的情况下, 交易所将发送关于总的交易量和订单数量的信息, 但是没有关于订单被取消的信息. 这增加了推断自己的订单队列位置策略的复杂性, 并且需要一些非平凡的和主观的建模来获得一个合理的估计.

由于市场数据处理和订单构建流程是事件触发器的关键路径的顶部节点[①], 因此其实现的效率对于策略的整体性能非常重要. 需要软件工程专业知识, 以确保订单构建过程以最有效的方式进行. 数据结构类型、内存管理方法和簿记算法都需要使用最有效的计算机语言进行精心设计. C++ 是一种用于延迟敏感任务的流行语言. Java 也被广泛使用, 因为它被认为更容易使用, 而且不易出错. 然而, Java的内存管理方面, 例如垃圾收集算法, 必须进行微调, 以实现总体延迟的高稳定性.

除了提供实时市场数据的高频算法, 市场数据处理的单元还负责收集离线 Alpha 研究的数据.[②]开发高频率算法, 将市场数据存储在原始数据包级别, 而不是在重建的 LOB 级别上, 是一种常见的做法. 这些原始数据可以通过标准的读文件过程或通过多播广播进行重新广播的适当应用程序进行后验检测的回放. 使用读文件的优点是, 回放的速度很容易被研究者控制; 这种方法的缺点是, 回放没有模拟实际交易发生的环境. 原始数据包重新广播能够在回放速度控制方面以牺牲灵活性为代价来模拟环境.

市场数据收集的另一个问题是时间标记同步. 这对于依赖跨交易信息集聚的策略尤其重要. 例如, 依赖 10 个交易所的集聚 LOB 数据的延迟套利策略对不同市场数据源的同步非常敏感. 为了准确地收集和回放这些市场数据, 收集必须由被精确时间协议同步的托管服务器执行, 它通常缩写为 PTP. 然而, 即使是使用了 PTP, 最基本的错配概率也仍然是不可忽视的, 尤其是在交易高峰时段. 求解错配问题的通用同步方法请参阅 4.4.2 节和 4.7 节的补充材料 9.

① 事件触发器是一种高频算法中的子程序, 它可以检测来自事件数据包生成的数据的交易信号; 参见 8.6 节补充材料 4 的进一步讨论.

② 参见 8.6 节补充材料 5 中关于 Alpha 研究的系统介绍.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

8.3.2 Alpha 引擎

一旦市场数据被解码, 订单随交易信息更新, 市场数据处理部门就传送结果, 通常通过回调机制, 传递给策略 Alpha 处理单元. Alpha 引擎实现了市场数据处理单元的回调功能, 并根据最近更新的市场数据和其他策略属性如位置库存、未完成订单和风险参数等制定交易决策. 交易决策是由模型或一系列模型驱动的, 这些模型将当前的策略状态作为输入, 并输出可执行的指令, 例如提交新订单、取消或替换现有订单, 或者什么也不做, 等待下一个事件.

为了说明这一问题, 考虑一个典型的市场制定策略, 该策略由 Alpha 单位执行短期价格预测和最优价差估计. 在一个新的市场事件被解码和订单更新之后, Alpha 引擎首先更新价格模型的参数估计, 并且在这个过程中, 同时计算一个新的价格预测和错误边界. 一旦它计算了预期的公允价值价格, 它就会检查策略的最新库存, 更新价差模型, 然后计算最优报价的价格和大小. 可操作的指令是通过新的最优报价调整、取消或替换现有报价. 注意, Alpha 引擎的触发器不需要从市场数据供给中获得任何市场事件. 订单事件, 例如一个已存在的订单, 也可以触发 Alpha 引擎的某些部分重新计算和重新提交新订单. 在这种情况下, Alpha 引擎的操作是由来自交换的订单网关的确认信息生成的.

考虑到一个高频率策略的交易需求, 所有必要的策略单元的计算都需要在几微秒内完成. 这需要定量研究人员和软件工程师的共同努力, 以找出模型复杂性、准确性和延迟影响之间的最佳权衡. 一旦战略部门做出了一个交易决策, 它就会将必要的订单属性, 如方向、大小和价格, 传递给订单管理部门. 8.6 节的补充材料 5–7 将会给出 Alpha 研究的进一步讨论, 补充材料 8–10 将提供延迟减少技术和硬件设置算法交易的其他细节和背景.

8.3.3 订单管理

订单管理单元从交换的订单消息通道接收实时信息, 并更新策略库存和订单状态. 它还与 Alpha 引擎单元进行接口, 在将这些订单发送到交易所之前将策略订单编码成有线格式. 通过对外部交换订单消息通道和内部 Alpha 引擎的接口, 订单管理单元非常适合用来跟踪订单状态和策略库存, 并在订单提交给交易所之前执行风险检查.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

8.3.4 订单类型和订单限定条件

策略方面的每个传出订单都有两个组成部分, 即订单类型和订单限定条件. 如第五章引言所述, 订单类型通常可以划分为限价订单、市价订单和取消订单. 订单限定条件是交易所提供的额外的灵活性, 以满足市场参与者的各种需求. 例如, 一个常用的订单限定条件是 GTC, 它意味着 “有效直到撤单”, GTC 允许附带的限制令保持在 LOB 上, 直到该命令完全执行、取消或 (衍生) 工具到期. 在没有使用 GTC 的情况下, 所有订单都被认为是在交易日结束时即当市场收盘时从 LOB 中删除的日订单.

每个交易所都提供自己的订单类型和订单限定条件. 例如, 纳斯达克平台提供的订单超过 10 类, 而芝加哥商品交易所提供的订单超过 5 类. 下文介绍三种限定条件:

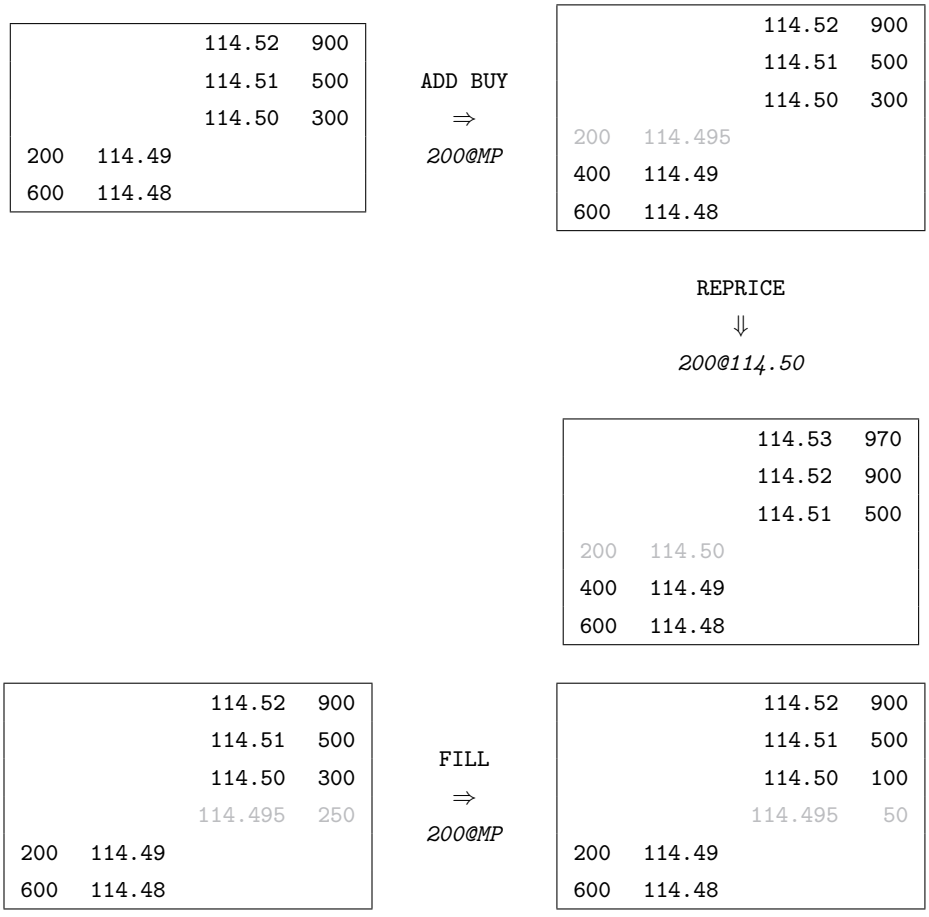
半价单顺序 半价单 (Midpoint-peg, 缩写 MP) 指令允许市场参与者提交与全国最佳报价和要约 (NBBO) 市场的中点挂钩的订单; 有关 NBBO 市场的背景, 请参见 8.4.1 节. 一旦交易所的匹配机器接收了一个挂钩订单, 它就会被隐含在 LOB 中. 只有执行这些命令才能显示它们的存在. 对于奇差而言, 挂钩的订单可以执行半便士的增量, 有效地将成本降低到小于最小报价单位. 这个订单限定条件的关键好处是价格的改善和匿名性. 这一限定条件可以在许多美国股票交易所上市, 如纳斯达克, 巴兹全球市场 (BATS global markets, 缩写 BATS)、群岛交易所 (Archipelago Exchange, 缩写 Arca)、波士顿股票交易所 (BX) 和费城证券交易所 (Philadelphia Stock Exchange, 缩写 PSX).

为了进一步解释, 表 8.4 给出了两种情况. 在第一种情况中, 该策略提交了一份订单: “在纳斯达克以半价单购买 200 股苹果股票.” 当 NBBO 是 (114.49 美元,114.50 美元) 时, 这个半价单就被接受了. 因此, 相对于 NBBO 的价格, 这个订单作为隐含订单放在 LOB 中, 其价格为 114.495 美元, 等于 (114.49+114.50)/2 美元. 当全国最好的订单在 114.50 美元被完全取消之后, 新的全国最好的卖价将会变成 114.51 美元. 这样的取消使得盯住价格从 114.495 美元变到 114.50 美元, 因为 114.50 美元是第二大投标价格和新的最好的报价的平均值. 也就是说, 取消导致隐含的买卖价差从半个点增加到 1 个点, 而被显示的买卖价差从 1 个最小交易单位到 2 个最小交易单位. 表 8.4 中的第二种情况对应于隐含订单位于卖价为 114.495 美元而全国最好的买价为 114.49 美元的情况. 因此, 新接受的 “盯住买入” 指令将立即执行, 交易量为 200 股, 交易价格为 114.495 美元. 注意, 在相同的 NBBO 配置下, 给出的价格 114.495

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

表 8.4 半价单指令购买 200 股股票. 纳斯达克在全国最佳报价和要约市场中. 上面部分: 半价单指令购买订单的结果是, 作为隐含的订单, 在中间市场定价为 114.495 美元. 然后, 在第 1 阶段的订单被取消, 一个新的一级请求级别被确定, 它将新中间价重新定价为 114.50 美元. 下面部分: 半价单指令购买订单结果, 立即以 114.495 美元的价格执行



美元要比在不提供 MP 限定条件的交换条件下使用市场订单的价格更好.

冰山订单 像半价单指令一样, 冰山订单也有一个无形的组成部分. 在 CME Globex 上, 限定条件采用了 *FIX Tag 210 MaxShow* 的形式, 它指定了订单的显示数量. 在执行显示的数量之后, 订单会自动地补充, 直到全部金额被填满. 例如, 考虑下列订单: "在芝加哥商品交易所, 购买 50 份 CLH6 合同[①]并把 *MaxShow* 设定为 10, 限定价格为 48.65 美元." 市场上只有 10 份合同. 一旦执行了 10 个合同, 就会向市场展示另外 10 个合同. 这个过程将重复, 直到整

① CLH6 是芝加哥商品交易所在 2016 年 3 月到期的原油 (WTI) 期货的代码.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

个订单的大小为 50. 注意, 在队列优先级中, 补充订单被视为一个新订单.

被动委托 纳斯达克的被动委托订单让做市商能够控制报价与现有流动性的相互作用. 只有在价格超过市场的情况下, 才会执行一个被动委托订单, 如果价格在经济上有利于执行, 那么订单就会从最好的买卖价格中重新定价. 对于那些利用回扣来控制交易成本的做市商来说, 这一规定的一个重要优势在于, 即使当做市商的订单快照被延迟时, 获取流动性和支付费用的风险也会降低. 被动委托订单必须标明为显示和限制, 否则将被交易所拒绝. 请注意, 这一限定条件也可以在许多美国股票交易平台上使用, 例如, 纳斯达克、BATS、Arca、BX 和 PSX. 为了说明, 考虑表 8.5 中的两种情况. 在第一种情况下, 该策略提交了一份订单: “在纳斯达克购买 200 股苹果股票, 被动委托订单价格为 114.50 美元.” 该订单将被接受, 并以 114.49 美元的价格添加. 在第二种情况下, 订单价格为 114.51 美元, 订单将以 114.50 美元的价格填充, 因为放弃的折扣 (净收益率) 是 0.0041 美元, 低于 0.01 美元, 因此对执行的订单具有经济上的优势.

表 8.5 购买 200 股被动委托订单. 上面部分: 限制价格为 114.50 美元, 导致订单被添加到订单簿中, 其与卖价相差 1 个单位, 并被定价为 114.49 美元. 下面部分: 限定价格为 114.51 美元, 订单将以 114.50 美元的价格填充

		114.52	900			114.52	900
		114.51	500	ADD BUY		114.51	500
		114.50	300	⇒		114.50	300
200	114.49			200@$114.50	400	114.49	
600	114.48				600	114.48	

		114.52	900			114.52	900
		114.51	500	ADD BUY		114.51	500
		114.50	300	⇒		114.50	100
200	114.49			200@$114.51	200	114.49	
600	114.48				600	114.48	

8.4 交易所规章制度

8.4.1 证券信息处理程序和国家市场体系管理规则

在美国, 1975 年的证券法修正案提出了证券信息处理程序 (SIP) 和股权国家市场体系 (NMS). 证券信息处理程序通过处理和合并所有参与中心的受

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

保护的报价和交易, 将所有的交换连接到一个数据流中.

证券工业自动化公司 (SIAC) 代表统一证券行情协会 (CTA)[①]运作并维护两个独立的计算机环境来处理交易和报价信息: 交易信息由统一证券行情系统 (CTS) 处理; 报价信息由统一报价系统 (CQS) 处理. CTA SIP 收集来自所有市场中心的报价和交易, 交易行情 A 和行情 B 的证券, 并传播实时交易和报价信息. 行情 A 的证券是在纽约证券交易所上市的, 行情 B 的证券是在 BATS、NYSE Arca、NYSE MKT[②]和其他地区交易所上市的. 对于在纳斯达克上市的证券 (行情 C), UTP[③] SIP 可以提供数据馈送的 UTP 一级信息——由合并行情系统处理的交易信息, 以及由合并报价系统处理的报价信息. 每天都有相当数量的 SIPs 交易进出. 例如, 根据 2015 年第一季度 CTA SIP 报告, 而处理所有传入数据的中值在该季度延迟接近 400 毫秒.

2007 年, 美国证券交易委员会 (SEC) 通过了证券交易所的国家市场体系管理规则 (Reg NMS).[④]为了确保市场数据传播对 SIP 数据和直接市场供给消费者而言是公平的, 国家市场体系管理规则确认了同步数据分发的标准, 以便报价和交易信息在同一时间为不同的供给离开交易所. 一旦市场数据离开交易所, 它就可以通过多个网络跳跃和不同的路径传输, 因此市场数据的到达会因特定的设置和接近源的距离而有较大变化. 由于这个原因, 国家市场体系管理规则不需要同步数据来同时到达目的地. 它打算将市场位置连接起来, 并包含一项阻止自动市场上的订单从而避免被其他地方提交的劣质订单所取代的条款. 国家市场体系管理规则[⑤]的关键计划是订单保护或交易通过 (规则 611) 和一系列的例外情况, 如建立跨市场的扫描订单 (ISO) 以提供有效和可行的中间价格优先级的规则 611(b).

规则 611 要求交易中心执行书面政策和程序, 以防止以低于受保护报价的价格执行交易. 这是为了防止所谓的 “通行交易”. 对于国家市场体系的股票价格报价被认为是受保护的, 它必须是即时和自动的,[⑥]并且是最好的出价或最好的报价. 该交易所既可以拒绝有价的订单, 也可以将其发送到显示最佳价格的交易所. 美国的注册交易所必须在主要交易所的 SIPs 中提供最好的买价和报价, 这在原则上是建立了最低的卖价和最高的买价的 NBBO. 像暗池这样的替

① 统一证券行情协会 (CTA) 是统一报价系统 (CQS) 和统一证券行情系统 (CTS) 的正式运营机构.

② 纽约证券交易所 MKT 的前身为美国证券交易所.

③ UTP 是未上市交易特权的缩写, 本质上是指行情 C 的证券.

④ 此规则仅适用于股票交易所. 期货交易, 例如在芝加哥商品交易所平台上交易的期货, 不受国家市场体系管理规则的约束.

⑤ 另外三个值得注意的项目是访问规则 (规则 610)、子便士规则 (规则 612) 和市场访问规则.

⑥ 请注意, 即使隐含的价格比显示的价格好, 也不能立即获得反向或隐含的订单, 因此不受保护.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

代交易系统不需要为任何 SIP 提供最佳报价, 但必须在 NBBO 中检查和匹配交易, 并立即向金融行业监管机构 FINRA的交易报告机构报告交易.

ISO 订单提供了一种机制来绕过 NBBO 和贸易保护检查, 而这些检查在规则 611 下是必需的. 在所有受保护的场所进行交易的经纪自营商必须在所有受保护的场所进行交易, 并假定交易所遵循规则 611. 使用 ISO 指令的一个优点是减少了延迟, 减少了由于过期的 NBBO 价格而导致的流动性减少的可能性. 为了说明这是如何工作的, 考虑 XYZ, 一个在交易所 A、交易所 B 和交易所 C 交易的 NMS 证券. 表 8.6 给出了这三种交易所的订单簿状态. NBBO 的买价为 114.90 美元, 报价为 114.50 美元. 交易商 D 想支付 114.51 美元购买 500 股 XYZ 股票. 如果交易商 D 在交易所 A 以 114.51 美元的价格购买 500 股股票, 那么该订单将被拒绝或被传送到另一个交易所. 如果将订单提交给交易所 B 或 C, 则订单将部分被执行, 其余部分将作为改进的报价档. 另一种情况是, 交易商 D 可以同时发送两个指令, 一个在交易所 B 支付 114.50 美元购

表 8.6 在交易所 A、交易所 B 和交易所 C 的 XYZ 订单的快照; NBBO 的买价为 114.90 美元, 报价为 114.50 美元

交易所 A			114.53	980
			114.52	900
			114.51	400
	400	114.49		
	600	114.48		
	750	114.47		
交易所 B			114.52	350
			114.51	200
			114.50	100
	150	114.48		
	170	114.47		
	270	114.46		
交易所 C			114.52	300
			114.51	100
			114.50	50
	110	114.48		
	210	114.47		
	300	114.46		

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

买 100 股股票, 另一个在交易所 C 支付 114.50 美元购买 50 股股票. 在执行和随后的 SIP 更新之后, 新的 NBBO 将分别以 114.49 美元买价和 114.51 美元的卖价提供. 然后, 交易商 D 可以提交第三个购买订单, 支付 114.51 美元购买剩余的 350 股. 另一方面, 通过使用 ISO, 交易商 D 可以同时向三个交易所一次性发送三个订单, 标记为 ISO.

8.4.2 证券卖空规则

卖空指的是在不首先拥有证券的情况下出售证券, 通过向另一方提供担保, 将证券交付给买方, 然后保持向相同或不同的交易对手进行贷款, 直至空头最终被买回. 在这样做的时候, 卖空者的目的是为了从证券价值的下降中获利, 由于利润足够的大, 以至于可以抵消那些实际上是抵押贷款给证券公司的任何损失. 卖空为促进价格发现提供了一种有价值的机制. 自 1938 年首次引入卖空规则以来, 市场发展快速, 证券卖空规则也提供了必要的更新, 以保障市场不受这种机制的潜在副作用的影响. 它在 2004 年首次获得批准, 经过反复修正, 美国证券交易委员会在 2011 年通过了当前的修正案. 规则 200 规定, 要求适当地 "标记" 为：多头 —— 卖方在销售时拥有证券, 空头 —— 卖方在销售时不持有证券, 卖空豁免 —— 卖方是超额配售, 或者承销商进行超额配售. 规则 203(b) 指定了一个阈值, 在交付失败时触发强制关闭.

规则 201, 作为提价交易规则的替代, 被广泛讨论, 并最终被修改. 该规则适用于允许卖空的条件, 并且如果有担保证券比收盘价格下降了 10% 或更多, 则禁止以低于当时的买入价卖出或执行卖空. 这可以作为一个熔断器, 从而有助于缓解由于卖空力量的增加带来的下跌态势. 需要注意的是, 多头 (如持有和卖出证券) 以最佳的价格执行出售仍然是被允许的. 举例来说, 在纳斯达克上市的 XYZ 公司所发行的 NMS 证券前一天的收盘价为 100 美元. 假设由于一些新闻事件, XYZ 的全国最优报价 (NBBO) 已经迅速降至 89.99 美元的买价和 90.00 美元的卖价, 导致纳斯达克根据规则 201 触发熔断器. 经销商 A, 没有 XYZ 股票, 且正在寻求卖空, 在纳斯达克提交限价单, 卖价为 89.99 美元. 订单将被重新定价并且显示为 90.00 美元, 这是因为 XYZ 是规则 201 所涵盖的证券. 经销商 B 拥有 100 股 XYZ 股票, 在感受到持续的下跌压力时, 决定出售 50 股并以 89.99 美元的价格提交 50 股限价卖单, 此单则标记为 "多头". 并且, 这个命令输出后将立即执行.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

8.4.3 其他交易规则

纽交所规则 48 特定纽约证券交易所指定的做市商 (DMM) 有义务传播开盘指示性价格, 这些价格需要在交易实际开始之前得到股票市场经理们的批准. 然而, 在极端情况下, 为了加速开放市场, 交易所可以通过使用规则 48 有权力推翻这一义务. 最近的一个案例是在 2015 年 8 月的最后一周, 当时道琼斯工业指数由于中国市场大幅下滑, 下跌了 1 个百分点. 纽约证券交易所在预期市场会出现大幅波动时援引规则 48 以加速市场开放.

纳斯达克规则 5550(a)(2) 根据这条规则, 公司要满足持续上市要求, 必须确保其股票的最佳买入价格保持在 1 美元以上. 如果最佳买入价格连续 30 天保持在 1 美元以下, 那么纳斯达克将向公司发出 "补正通知". 公司将获得 180 天的时间来重新达标, 即在 180 天内至少连续 10 天的最佳买入价必须高于 1 美元. 如果在 180 天内没有达成, 公司将收到退市信.

8.4.4 熔断机制

替代 Reg SHO 上调规则的是熔断机制, 其通过指定允许卖空的一系列条件来限制价格大幅下跌后的下行价格波动. 目前许多交易所还列出了其他更明确的熔断机制. 这些熔断机制也被称为交易限制, 实际上可以在一段特定时间内完全停止交易, 以使市场以更有序的方式达到均衡价格. 它们能适用于整个市场的或特定股票. 对于每个 NMS 股票, 熔断机制根据 SIP 计算出的参考价格在指定的时间窗口内连续监测价格波段. 价格波段的宽度取决于证券的流动性等级. 8.6 节的补充材料 11 提供了熔断机制的更多细节.

8.4.5 市场操纵

以下是交易所和监管机构所监控到的市场滥用行为的例子, 这些交易所和监管机构将这种行为视为违反真实的交易活动或操纵市场行为.

分层 分层会造成订单的人为不平衡, 从而扭曲了供需关系, 导致有利的买卖或影响市场的方向.

欺骗 欺骗行为通过提交过多的订单而产生虚假的流动性; 还意味着为交易基础架构有意加重工作量, 增加竞争对手的处理延迟.

做尾盘 此违规的意图是影响资产的收盘价, 以获得更有利的利润, 或者影响依靠收盘价格进行未来交易决策的其他算法.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

8.5 风险管理

由于全球交易所交易量中有相当比例是由于某种形式的自动交易所致, 因此这种交易方式可能对全球金融市场产生深远的影响. 在交易公司内适当的风险管理和合规程序的实施以及一定程度的监管对长期的市场参与者都是有利的. 在交易公司层面, 风控官的角色是确保公司所做的所有交易活动都遵循一系列预设的良好实践和限制, 并根据战略的执行情况、交易量和评估策略的稳健性来更新这些实践和限制. 一些风险限制也由交易所监测和执行. 例如, CME 对指令速率有严格限制, 所以任何超过此阈值的订单都将被取消. 许多交易所还监控了交易成功量与指令总数之比的一些界定. 合规专员在公司层面的角色是监控所有交易活动, 确保遵守适当的交易规则和监管准则, 并调查剩余没有解决的任何可能导致合规和法律问题的潜在问题. 对于定量交易, 公司层面的两个主要风险来源是操作风险和策略风险.

8.5.1 操作风险

由于交易决策是在瞬间进行的, 确保算法和基础架构无差错依赖于软件工程和定量分析. 交易算法需要经过多个测试和认证阶段, 并进行重复检查. 看似不明显的错误 (如整数溢出或下溢) 可能会对策略产生重大影响. 算法交易过程中的操作风险主要是指由于基础架构中断和软件错误而导致的风险. 与任何大规模的复杂和关键任务应用程序的软件开发一样, 软件测试的方法也是不可或缺的.

软件错误风险 软件测试技术 (如单元测试、回归测试和非回归测试) 被广泛用作软件开发过程的一个组成部分. 单元测试旨在隔离特定的类或类成员函数, 以便测试独立于代码的其他部分. 有时, 开发人员还需要一种方法来检查新的错误修复程序或新版本是否不会破坏现有的代码. 因此, 除了针对特定修改的单元测试之外, 开发人员还执行回归测试, 它比较实际应用程序的输出而不是代码的某些抽象部分. 比较过程通过以下几个步骤进行: (a) 将应用程序的输出保存到文件中, (b) 执行修改, 然后 (c) 执行两个输出文件, 简单比较以确保修改不影响现有的正确表现. 鉴于软件的回归测试旨在确保在修复缺陷的过程中没有现有的功能被破坏, 对软件的新版本执行非回归测试, 可以确保任何新增功能都能产生预期的效果.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

订单传输协议风险 根据 CFTC(商品期货交易委员会) 指令, 为了使交易公司能够直接访问交易所配对和市场数据基础架构, 它们需要通过一系列一致性测试. 为了实现一致性测试, 交易公司连接到模拟真实产品环境的交换管理设施进行测试, 这是为了在产品环境中测试它们完整的端到端交易流程. 例如,CME 为当前产品环境 (认证环境) 和未来环境 (新版本环境) 提供专用测试环境. 希望测试其交易基础架构的客户可以从位于同一产品设施中的专用数据中心连接到环境, 并交叉连接到测试网关和匹配引擎.

其他基础架构的相关风险 如果传入消息的速率高于处理消息的速率, 操作系统内核可能会丢弃网络数据包; 一旦捕获缓冲区已满, 内核就会丢弃所有传入的数据包. 发生这种情况时, 某些信息会丢失, 并且文件状态可能会破坏. 当最好的买卖价格被锁定或交叉时, 这往往表现为滞后价格形式. 需要适当的机制来处理由于滞后信息导致文件状态不正确或怀疑的情景. 前者是相当简单的处理. 例如, 算法可以从交易所请求指令快照或者从另一个独立的数据捕获设备回放预先记录的数据.[①]然而, 后者存在的问题会多一些, 特别是当稳健状态正常而价格是滞后的还是实时存在不确定情况的时候.

意外关机或网络故障是基础设施风险的另一个来源. 这对于在订单簿队列中有许多未完成订单的做市策略尤为重要, 这些订单都需要不断的监控和调整. 策略或订单管理引擎的预期关闭或发生网络故障可导致意外执行未完成订单, 这将需要痛苦的人工干预来关闭或对冲. 如果网关检测到客户端引擎可能关闭或出现网络故障, 在交换中处理这种情况的一种方法是通过自动取消. 策略和订单管理引擎之间可以采取类似的机制.

合规风险 合规专员的职责是帮助检测和防止潜在的违规行为. 检测违规需要以微秒级处理进行大规模的数据分析. 监控交易活动是一项挑战, 交易活动要跨越多个交易所和多个资产类别的多个时区. 鉴于一个典型的高频策略所传输的大量订单需要一定程度的自动化来处理传入和传出的消息, 以此发现潜在的违规行为. 这可以通过回测覆盖在历史市场数据之上的战略指令, 通过合规性政策检查清单并寻找潜在的违规行为来完成. 虽然大多数交易所都明确规定了超额信息率等政策, 而且易于检测, 但惯性激发或分层等其他方法则需要复杂的模型来跟踪许多内部状态, 并提供必要的能力来区分市场活动流动性来采取行动.

① 像纳斯达克和 CME 这类的交易所还提供了丢弃拷贝服务, 它将客户端订单回复消息发送到指定的位置, 作为通过正常的生产连接将订单回复消息的备份副本发送到客户端.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

8.5.2 策略风险

对于策略实施, 除了采取合理的程序之外, 考虑可能对算法的整体完整性产生不利影响的风险也是很重要的.

模型风险 对于需要实时参数更新的模型, 应该有适当的机制来监视和纠正拟合误差, 例如, 由于需要求逆矩阵条件数目恶化. 这些错误可能会破坏未来的预测和估计.

交易风险 从风险管理的角度来看, 高频交易也有一些传统交易已经存在的常见风险参数. 因此应执行以下数量和方面的检查来控制交易风险:

- 头寸限制: 这限制了特定投资组合的净头寸. 它可以是简单股票的名义术语, 期权的 delta 术语, 或者是债券投资组合的 (美元) 期限术语. 对于一个普通的股票投资组合, 可能会限制具体行业部门, 以限制系统暴露于某一特定部门. 对于债券投资组合, 收益率曲线的不同部分可能有特定的限制, 以限制交易公司在收益率曲线形状上的变化.
- 盈亏限制: 这限制了特定策略的亏损;
- 期权希腊字母敞口: 这限制了期权 delta, vega, gama 和其他风险分解的敞口.
- 情景限制: 这是以一些经济情景为基础, 或者是基于历史实现, 或者是基于一些理论上的最糟糕的经济环境, 来限制盈亏. 这与常用的风险价值限制密切相关, 其中有一个特定的被放在尾部的阈值或相关的尾部事件.
- 库存调节: 在一天开始 (BOD) 时, 交易员有责任确定 BOD 的位置是否符合预期. 这涉及从 BOD 头寸的上一个结束日期 (EOD) 结束时检查库存, 并检查相对于结算公司的开放头寸. 这个检查可以确保系统运行的自动化过程已经找到了正确的位置, 而且没有任何未完成的交易中断.

对于自动化交易, 这里有几个其他相关的变量需要考虑:

- 订单大小: 这限制了每个提交订单的大小.
- 订单速率: 这限制了策略在一秒内可以提交订单的数量.
- 消息速率: 这限制了指令总数与完成订单数之比.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

- 潜在仓位: 这限制了如果所有未完成的订单都被实现, 投资组合可以持有的潜在仓位. 它与做市策略更为相关, 因为在限价订单的不同层次上可能存在多个订单, 从而执行的可能性减小.

在发送每个订单之前, 需要进行交易前风险检查, 在此期间, 策略循环执行所有必要的检查, 以确保发出的指令不违反上述任何风险检查.

另外, 策略必须监视所有接收到的订单回复消息, 以检查可能影响交易的拒绝信息或其他消息. 拒绝可以来自交易所或订单管理引擎. 前者通常被称为*交易所拒绝*, 后者被称为*内部拒绝*. 交易所可能会因为订单属性不正确、消息传递率过高或保证金不足等原因而拒绝订单. 一个策略需要有一个机制来处理交易所和内部拒绝, 这样它就不会重复发送, 这会对整体延迟产生负面影响, 更重要的是会阻塞交换带宽.

8.6 补充材料和注释

1. **场外交易市场的交易** 第二段介绍的场外交易市场主要涉及更为复杂和定制化的产品. 场外交易市场的交易通常涉及最少三个交易对手: 发起交易商, 对手交易商和作为调解人的经纪人. 经纪人的角色是协助原始交易商匿名寻找必要的流动性. 对于这项服务, 经纪人获得佣金. 对于利率掉期, 这个佣金相当于一个*连续基点*的一小部分.[①]为了能够解释清楚场外交易, 考虑一个简单的场外交易利率互换, 涉及两个虚构的交易对手: 宙斯银行 (经销商 A) 和雅努斯银行 (经销商 B). 下面的对话概述了场外交易机制如何在现实生活中起作用:

```
经销商A对经纪人: 2年期半年付息100MM的债券价格是多少?
经纪人: 让我查查.
经纪人对经销商B: 贵方2年期半年付息100MM的债券价格是多少?
经销商B: 100MM我出价1.25%和1.75%.
经纪人: 谢谢.
经纪人对经销商A: 100MM是1.25%和1.75%.
经销商A: 卖出100MM.
经纪人对经销商B: 宙斯银行以1.25%卖出100MM.
经销商B对经纪人: 成交.
```

① 基点是 $\frac{1}{100}$ 百分点, 一个连续基点是指一段时间内连续的基点的流量, 例如, 在两年内以每年 25 个连续基点为例, 其代表两个以 0.25% 乘以名义额的现金流量, 分别是第一年年底的 0.25% 和第二年年末的 0.25%.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

经纪人对经销商A: 成交,对象为雅努斯银行.
经销商A对经纪人: 成交.
经纪人对经销商A: 你以1.25%卖出2年期半年付息100MM的债券给雅努斯银行.
经纪人对经销商B: 你以1.25%从宙斯银行买入2年期半年付息100MM的债券.

2. **网络层传输协议** 网络传输协议有两种主要类型: 传输控制协议 (TCP) 和用户数据报协议 (UDP).TCP 是面向连接的, 因为它建立了两个主机之间的一对一连接或单一传播, 因此一旦连接建立, 数据就可以双向发送. TCP 连接具有多级检查和确认以确保传输可靠性. 另一方面,UDP 是无连接协议. UDP 不像 TCP 那样需要一对一连接, 而是可以在多播设置中使用 UDP, 一个主机将数据传播到多个主机接受特定多播信道. UDP 多播的优势在于带宽优化和输出最大化. 但是, 一旦 UDP 数据包离开主机, 就不能保证全部或部分被指定的接收者接收. 接收者有责任检测丢失的数据包并请求主机重播. 除非传播者和听众之间有多次跳跃, 否则如果接收者能够跟上处理到达的数据包, 则丢弃分组的发生通常是罕见的.

3. **CME MDP 3.0 信息** 此处对 8.2.3 节中关于 CME MDP 网络数据包提供了更多细节. 数据包遵循标准分层格式:

   - 二进制数据包标题给出分组的时间戳的唯一序列号, 例如
     - 10897345 2015101512345234563248 8

     这意味着网关在 2015-10-15 12:34:52.345632488 发送序列号为 10897345 的数据包.
   - Message Header 给出了解码 MDP 消息的方法; 见图 8.4, 其中 Message Size 指包含二进制标题的消息的长度, Block Length 是指消息其余部分的长度, TemplateID 是用于编码消息的 ID, SchemaID 是发布消息的系统的 ID, 以及 Version 表示 Schema 版本.
   - FIX Message Header 给出了 FIX 消息的类型, 例如图 8.6 中的 35=X.
   - FIX Message Body 给出以实际的 FIX 消息类型为条件匹配事件的类型.
   - FIX Message Block 给出匹配时间的细节.

   对于交易重要的两个主要 MDP 消息类型是交易概要和工作簿管理. 交易概要信息是针对交易事件的, 即当匹配过程发生交易或一系列交易时,

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

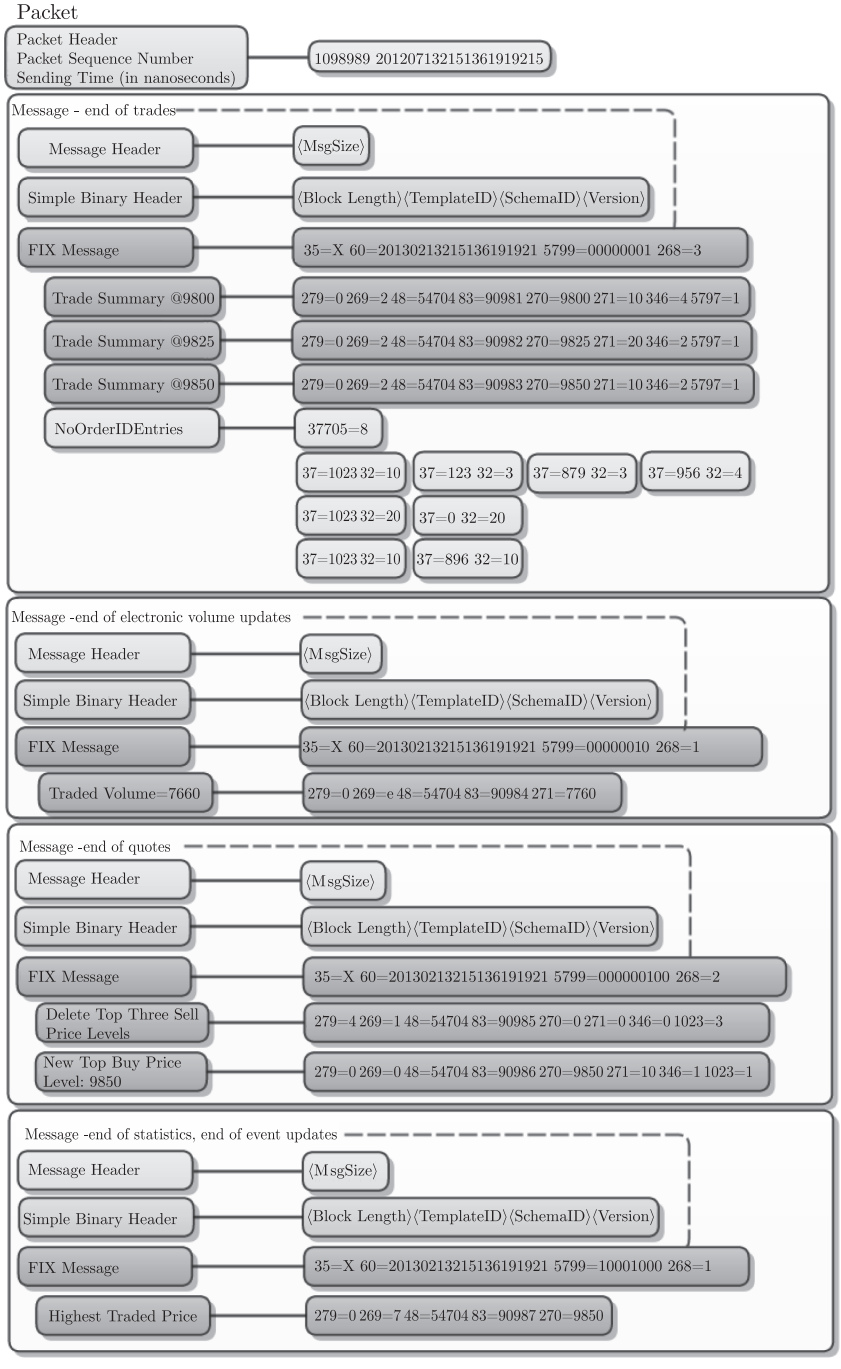


图 8.6　买方激进交易的 MDP3.0 信息, 该交易出现在 LOB 卖出的三个级别中 (由 CME 授权使用)

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

这些信息中的一个或多个将由交易所传播. 工作簿管理消息有两个子类型: 多深度簿和隐含簿. 簿通常按照贸易事件更新事件. 例如, 考虑一个买单在买方取出工作簿三个等级的情况, 由此产生的 MDP3.0 消息由图 8.6 表示. 对于完整的 MDP3.0 消息协议格式, 请参考 CME 在线资源.

4. **市场事件, 策略触发器和 CME MDP 3.0** 事件触发器处理触发策略计算和订单提交的事件序列. 市场数据来自网络数据包, 它可以包含原子簿转换, 多个原子簿转换或者部分或不完全转换的信息, 如 8.2.3 节中所讨论的. 由于触发事件, 原子簿转换由所有簿和订单事件组成, 因此作为策略的输入信息的订单簿的快照与在触发事件结束时的匹配引擎的快照相匹配. 例如, 图 8.6 定义了 40 个期货合约买入订单到达时触发的原子簿转换. 如果数据包仅包含交易摘要而不包含后续的簿管理更新, 则数据包仅包含部分簿转换, 因为在数据包末尾, 簿状态并不能完全反映传入采购订单触发的事件. 从战略的角度来看, 这个区别是重要的. 尽管策略需要在解析原始网络数据包后立即对事件作出反应, 但不完整的簿信息可能会导致不正确或不一致的簿状态更新. 为了说明这一点, 考虑一种在检测交易事件时重新计算内部模型状态的做市策略. 如果事件跨越多个数据包, 模型计算和报价单应等待, 直到所有数据包都被解析. CME MDP 3.0 的一个优点是解析过程可以在第一个数据包到达后立即开始, 而不需要等待所有的数据包.

5. **Alpha 研究** Alpha 策略模型研究通常包括以下几个关键步骤:

   - 问题范围形成: 这就定义了研究项目的性质和范围; 例如, 问题范围可能是 “精炼当前模型以提高基于水平的订单簿列位置推断的准确性”;
   - 文献综述: 根据研究课题的范围和定义, 对该领域及相关领域已有的工作进行研究;
   - 理论基础: 在文献综述的基础上, 研究人员可以选择最合适的建模框架, 并为需要修改的理论基础奠定基础;
   - 模式: 研究人员使用诸如 R 或 Python 之类的脚本语言对理论模型进行编码, 并执行一个简单的模拟作为概念验证;
   - 产品代码的实现和测试: 一旦满足典范仿真的结果, 研究人员可以用 C++ 或 Java 等编译语言实现原型, 然后用更精确的匹配引擎级仿真器进行测试; 这个往往是保守的风险限额的产品交易;

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

- 产品启动: 模型最终投入正常风险限额的生产, 并逐步将风险参数放宽至正常水平;
- 改进: 使用交易调试日志, 研究人员研究模型行为, 确定需要解决的技术问题并注意未来研究的其他功能.

6. **Alpha 研究的编程平台** 在定量研究中, 经常需要快速建立想法的原型, 即对模型进行编码, 使其适合于一些历史数据并评估结果. 脚本语言 (如 R, Python 或 Matlab) 非常适合这种快速原型. 脚本语言优于编译语言 (如 C++ 和 Java) 的优势在于, 研究人员可以轻松修改模型, 即时查看结果和查询变量. R 已被统计界广泛采用, 其中还包括机器学习和一般数据分析. 它是开源的, 代码库经常更新. 虽然主要的代码库提供了大量基本的数据分析功能, 但是 R 的强大功能来自许多包含广泛研究需求的软件包, 从尖端的机器学习库, 例如, `kernlab`, `nnet` 和 `glmnet`, 到复杂的优化例程, 例如 `nloptr` 和 `optim`. 对于处理大型数据集, R 有一些软件包, 例如 `biglm`, `ff`, 可以执行迭代模型更新, 因此不需要一次加载整个数据集. 对于具有多个内核或网格计算的现代计算机, 它还具有许多包, 例如 `snow`, `parallel`. 其他 R 包, 如 `gputools` 和 `gmatrix`, 也可用于图形处理单元 (GPU[①]) 上的并行计算, 其中 GPU 现在是现代计算机中的标准组件.[②] 对于喜欢 Python 的面向对象编程语言语法的研究人员来说, Python 是一个流行的选择. 与数据分析库 pandas 一起, Python 在模型复杂性和运用能力方面可以与 R 匹敌.

   原型设计完成后, 模型需要以编译器语言实现, 以便充分利用内核和计算机硬件的低延迟功能. 在理想情况下, 研究模型和生产模型共享相同的代码或调用相同的核心库集, 以尽量减少翻译阶段的错误. 做到这一点的一种方法是用最终选择的语言编写原型, 并用该语言进行研究. 这可能是低效率的, 因为编译和调试的过程阻碍了研究过程. 另一种方法是构建一套 C++ 库, 通过足够数量且设计良好的语法糖来更便于使用.

   提供了最佳平衡的方法是以包和库的形式扩展已经得到良好支持和广泛采用的语言, 如 R 和 Python. 交易公司的研究团队可以使用 C++ 或 Java 编写的高度优化的代码实现库, 支持特定的模型或模型的一部分, 并且可以添加专有的功能来支持固件实现, 如计算网格, GPU 或协处理器

① GPU 是图形处理单元的缩写. 由于 GPU 具有高度并行的结构, 现代 GPU 不仅用于计算机图形处理和图像处理, 而且用于并行计算.

② 关于 R 和 C++ 在现代数据分析中使用的并行计算的更多信息, 请参见 Matloff (2015).

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

集群, 以尽可能简单地访问历史市场数据和交易日志. R 可以加载这些库进行原型设计, 然后在生产代码中使用相同的库.

7. **数值准确性** 虽然交易算法是通过对实数集的运算推导出来的, 但它们是在基于小数进行数值运算的计算系统上实现的, 这些小数可以表示为如下形式:

$$\pm \left[d_0 + d_1\beta^{-1} + d_2\beta^{-2} + \cdots + d_{p-1}\beta^{-(p-1)}\right]\beta^e, \tag{8.1}$$

其中 $p, \beta \in \mathbb{N}$, $\beta > 1$, $e \in \mathbb{Z}$, $d_i \in \mathbb{Z}$, $0 \leqslant d_i < \beta$, 其中 $i = 0, \ldots, p-1$. 这被称为浮点数表示法, 包括精度 p, 基数 β 和指数 e. (8.1) 中的数 $d_0 + d_1\beta^{-1} + d_2\beta^{-2} + \cdots + d_{p-1}\beta^{-(p-1)}$ (记为 m) 称为有效位数或尾数表示法. 通常来说, (8.1) 假设 $d_0 \neq 0$, 从而 $1 \leqslant m < \beta$. d_0 和 d_{p-1} 分别称为最高位和最低位. IEEE 754 二进制格式是浮点位 (如 $\beta = 2$) 表示的标准化规则. 这一格式也能处理 0 值、负值和缺失值 (NaN)①. 尽管十进位制的数如 $0.25(= 1.0 \times 2^{-2})$ 和 $0.5(= 1.0 \times 2^{-1})$ 在二进制里有准确的表达方式, 但许多其他的数例如 $0.1(= 1.1001100110011\cdots \times 2^{-4})$ 并不能准确表述, 并且会导致一个近似错误, 而这一错误常常会出人意料地造成重大影响; 典型的例子是在非计划取消和通过比较浮点数产生的决策错误中丢失精度.

8. **硬件设置** 用于算法交易的适当的计算机硬件设置很大程度上取决于策略的类型. 对于高频交易, 对计算硬件的需求可能比中低频策略的要求高一些. 一般来说, 更高的时钟频率意味着每秒执行更多的代码行, 这意味着对市场事件的反应更快. 但是, 时钟频率只是最佳设置的一个方面. 诸如 CPU 架构、主板和芯片组设计、内存配置文件以及网卡特性等因素都可能对战略的整体性能产生重大影响. 现代 CPU 架构允许每个 CPU 芯片有多个处理内核②, 每个内核可以有多个硬件线程. 核心和硬件线程之间的区别在于每个核心具有独立的硬件资源, 而物理线程几乎共享增值核心的资源, 例如实际寄存器.

① NaN 表示 "非数字", 并且是一个没有定义过的值, 它来源于一些类似于 0/0 或者 $\sqrt{-1}$ 的数字运算. 另外, 需要注意的是 IEEE 754 格式并非只使用 $\beta = 2$ 的情况, 而且还会选择 p 与整数 $e_{\min}$ 和 $e_{\max}$ (当 $e_{\min} < 0 < e_{\max}$), 以至于所有代表性数字都能满足 $e \in [e_{\min}, e_{\max}]$, 数字上溢和下溢则对应着 $e > e_{\max}$ 和 $e < e_{\min}$. 近似误差的公差决定了 p, $e_{\min}$ 和 $e_{\max}$ 的选择.

② 制备是以晶圆片的形式产生集成电路完成的. 每个晶圆片进一步切成小块, 以便每个块包含电路的一个副本, 而这每个晶圆片则称为处理器芯片. 该单元 "芯片" 来自晶片切割工艺.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

9. **延迟降低技术: 缓存优化和向量化** 随着 CPUs 执行效率和时钟速度的提高, 交易应用程序的瓶颈往往是内存访问. 因此, 对于实现低延迟应用程序, 了解内存体系结构 (尤其是在内存缓存中) 是必不可少的. 高速缓存存储器, 也称为 CPU 存储器, 是一种随机存取存储器 (RAM), 计算机微处理器可以比普通 RAM 更快地存取它. 速度的差异源于 (a) 缓存内存内容比常规内存中的数据有更高的优先级由微处理器进行处理,(b) 缓存内存是由半导体 SRAM (静态 RAM) 来实现, 其访问速度比 DRAM (动态 RAM) 来处理常规内存要快. 此外, 由于 SRAM 比 DRAM 昂贵得多, 所以在大多数服务器上, 缓存内存相比常规内存要小得多, 并且经常会以等级 1 (缩写为 L1) 的形式拥有最高的访问权限且等级 2 拥有次级优先访问权, 等等. 更多关于电脑内存结构的细节请看 Drepper (2007). 缓存优化使用软件工程技术来改进空间和时间局部性, 在较高的级别上, 这意味着频繁地连续访问的数据, 但整体应该在内存中. 它还涉及利用缓存内存以实现低延迟的其他方法. 特别是指出缓存是一种稀缺资源. Loveless 等 (2013) 建议在运行时通过使用在线 One-pass 算法 (如在 3.4 节的补充材料 9 里, 该标量形式的 Kalman 线性矩阵代替了补充的 Kalman 增益矩阵) 来只保留相对少量的 L1 缓存中的数据和参数.

   向量化是代码优化的一种形式, 或由开发人员手动使用称为 “内部函数” 的内核内置函数完成, 或由编译器自动完成. 它利用单指令多数据 (SIMD) 操作, 通过在静态寄存器上使用向量寄存器, 这样可显著提升性能. 英特尔的最新流式 SIMD 扩展 (SSE), SSE4 能够在一个操作中处理 128 位. 向量化默认情况下通过编译标志 `-O3` 来启用, 并且可以通过标志 `-ftree-vectorize` 和 `-fno-tree-vectorize` 明确地打开或关闭. 有关可以或不可以优化的函数的编译器向量化选项和示例, 请参阅 `https://gcc.gnu.org/projects/tree-ssa/vectorization.html`.

10. **静态与动态调度** 在面向对象编程 (OOP) 中, 我们通常使用虚函数实现多态性. 这使得编译后的代码能够通过动态绑定或运行时绑定在派生类中执行适当的函数. 在低延迟应用程序中, 这种分派方法有许多缺点, 例如无法使用内联、每个虚拟方法调用都有额外的间接处理、虚拟表指针每个对象有额外的 8 字节开销, 这些开销可能导致缓存失败. 补救方法是使用静态分派来模仿模板设计模式如 CRTP(奇怪的重复模板模式) 的动态对等项.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

11. **熔断机制的历史** 1987 年 10 月 19 日黑色星期一, 全球股市崩溃, 在很短的时间内就出现了巨额价值损失. 道琼斯工业平均指数 (DJIA) 下跌 508 点至 1738.74 点, 此后便出现了熔断机制. 熔断机制的目的是减少市场波动和大规模的恐慌性抛售, 给交易者时间来重新思考其交易.

    纽交所的熔断机制在 1987 年开始执行, 其特点如下:

    - 在每个季度初, 纽交所设置了三个熔断器, 通常为 S&P 500 的本季度第一个月平均收盘价的 7% (1 级), 13% (2 级) 和 20% (3 级), 四舍五入到最接近的 50 点间隔.
    - 1 级和 2 级下跌可导致 15 分钟的交易暂停, 除非其发生在下午 3:25 之后, 此时没有交易暂停适用.
    - 3 级下跌将导致当日停止交易.

    在 2010 年 6 月 11 日, 即在 2010 年 5 月 6 日发生暴跌事件之后, 美国证券交易委员会要求所有美国证券交易所为所有 S&P 500 指数股票实施额外的安全级熔断器. 特别地, 不使用季度平均收盘价作为基准, 新的股票熔断器将基准定为前五分钟的交易价格. 如果股票价格在五分钟内超过 10% 的涨跌幅度, 那么与这个股票相关的所有交易 (包括在美国其他交易所的交易以及其衍生品交易) 都将暂停至少五分钟. 在那之后, 如果受影响的股权存在严重的不平衡, 发布暂停的交易所可以考虑延长这一暂停. 如果暂停时间超过 10 分钟, 其他交易所可以自由恢复交易.

    截至 2010 年 9 月 13 日, 安全级别的熔断器首次覆盖 Russell 1000 股指和 344 只交易最为活跃的 ETF. 自 2011 年 6 月以来, 它已经覆盖了所有的 NMS 证券. 2012 年 5 月 31 日, 美国证券交易委员会甚至通过使用上下限制机制来改善安全级别的熔断器, 通过防止大型突然价格变动触发上市股票证券交易来应对市场波动. 美国证券交易委员会同一天也修改了市场广度熔断器规则, 并在闪电崩溃后实施了新的市场熔断器. 同一天, 美国证券交易委员会修订了市场广度熔断器规则, 其基于前一天标普 500 指数的收盘价设定触发器. 新的触发器相比过去基于季度数据的触发器来说, 对市场情况更加敏感. 这一修订在 2013 年 2 月实施.

12. **高频交易中的闪崩和运行风险** 闪崩是在很短的时间内迅速发生的证券价格大幅下跌. 2010 年的 "闪崩" 事件发生在 5 月 6 日, 当时由于担心希腊债务危机, 美国股市呈下滑态势. 下午 2 点 42 分, 道琼斯工业平均指

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

数 (DJIA) 当天下跌超过 300 点, 5 分钟内迅速下跌超过 600 点, 截至下午 2 点 47 分, 全天下跌近 1000 点. 20 分钟后, 道指重获 600 点大部分回落. 2010 年 9 月 30 日, 美国证券交易委员会 (SEC) 和美国商品期货交易委员会 (CFTC) 发布了一份联合报告 “关于 2010 年 5 月 6 日的市场事件的调查结果”, 以确定导致闪崩的原因. 该报告描述了一家大型共同基金公司如何销售异常多的 E-Mini S&P 股指期货, 使得所有买单都耗尽, 从而高频交易者大举抛售, 导致 E-Mini S&P 500 的价格从下午 2 点 41 分到 2 点 44 分下跌了 3%. 与此同时, 购买 E-Mini S&P 500 的跨市套利者同时在股票市场上出售了相当数量的股票, 也使 SPY 的价格下跌了约 3%. 报告还指出,CME 的停止逻辑功能使得 E-Mini 的交易暂停了 5 秒钟, 价格稳定, 不久后, E-Mini 开始复苏, 其次是 SPY. 在这份长达 104 页的报告发布后的一天内, 一些评论家指出, 责备闪崩是由于 (来自 Waddel & Reed) 单一订单触发的是不对的, 特别是 CME 反对报告对什么导致了闪崩的原因解释. 联邦储备委员会学术合作者的经济学家发表了一篇工作论文 (Kirilenko 等, 2015), 其中包含对闪崩中交易数据的回顾和实证分析. 正如美国证券交易委员会的报告所指出的那样, 根据 “烫手山芋交易” 数据, 报告称, 由于高频率公司迅速获得了这些数据, 从而他们在价格迅速下滑的时候平仓了. 2015 年 4 月 21 日, 美国司法部对高频交易者提出了 “22 项刑事指控, 包括欺诈和市场操纵”, 据称这些高频交易商曾经利用欺骗算法为 E-Mini S&P 500 股指期货提交了几千个订单, 相当于大约 “2 亿美元的筹码认为市场将下跌”, 并在取消前修改它们达 19000 次; 这一行为至少是造成订单失衡的重要原因, 而这又是导致闪崩的原因之一.

2013 年 10 月, 东南亚最大的证券交易所新加坡交易所 (SGX) 发生了一次闪崩, 蒸发了 69 亿美元市值, 其中三只股票 BLUM,ACAP 和 LionGold Corp 的价值损失至少达 87%.

2014 年 10 月 15 日, 美国债券市场出现了闪崩, 其 10 年期债券收益率由上午 9 点至 9 点 36 分从前日收盘价 2.2% 至 1.9%, 但在 15 分钟内反弹到 2% 以上. 纽约联邦储备银行、美国财政部、美国证券交易委员会和美国商品期货交易委员会于 2015 年 7 月发表的联合报告表示, 闪崩不能归因于一个具体的原因, 而是由于一些因素, 包括顺序的下降订单簿深度, 数量空前的空头头寸, 以及在该事件窗口部分期间的自我交易水平的提高. 该报告将自我交易定义为同一实体采取买卖双方交易的交易.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

2015 年 8 月 24 日, 道指在头 5 分钟的交易中下跌了 1100 点. 8 月 23 日星期天晚上, 亚洲股市大幅下挫引发了美国和欧洲股指期货的下跌, 美国股指期货在美国开盘前下跌达 7%.

因此, 在高频交易中, 对于闪崩事件的潜在的巨大损失的风险管理尤为重要. 如果不能防止操作风险中在 8.5.1 节中提到的软件错误风险, 可能会导致灾难性的损失. 一个典型的例子就是 Knight Capital 集团 2012 年的软件错误, 该集团的最大业务是在美国股票和期权市场上进行交易, 同时也从事机构销售和电子交易执行. 2011 年, 公司价值 15 亿美元, 员工约 1450 人. 2012 年 8 月 1 日, 由于技术人员忘记将新的 `RLP` (零售流动性程序) 代码复制到八台服务器中的一台, 以实现股票指令的自动路由, 因此它在 2012 年 8 月 1 日部署了未经测试的软件, 该软件包含一个过时的函数 `PowerPeg`. `RLP` 改变了以前用于激活 `PowerPeg` 的标志, 该标志旨在于变动价格. 因此, 发送给该服务器的订单触发了有缺陷的 `PowerPeg`, 造成股价大幅波动, 并在 45 分钟内在 154 只股票中执行了 400 万次执行. Knight Capital 在税前亏损 4.4 亿美元后, 其股价下跌超过 70%.8 月 5 日, 它从几个主要投资者手中筹集了大约 4 亿美元的资金, 并继续运营, 然后被全球电子贸易公司 (Getco LLC) 收购, 后者组建了新公司 KCG Holdings, Inc.

2010 年的闪崩引起了公众对高频交易 (HFT) 的兴趣以及 HFT 的众多神话和争论. 期货工业协会主要贸易商小组在 2015 年将这些内容的一些争议和进一步的讨论进行了总结. 监管部门对这些争议的回应和改善股市结构的工作仍在进行中; 美国证券交易委员会主席 Mary Jo White 的演讲以及行业代表和监管机构的评论于 2014 年 7 月 9 日发布在哈佛法学院公司治理与金融监管论坛上 (`http://corpgov.law.harvard.edu`).

8.7 习题

纳斯达克将由 NASDAQ OMX 集团开发的直接外向市场数据的协议 TotalView-ITCH, 用于其软件和市场数据反馈服务的 FPGA 硬件版本.[①]与 FIX 不同, ITCH 消息字段是固定长度, 不带分隔符或标记. 消息的开始和结束位置由相对于消息开始的偏移量和字段的长度来表示. 此功能允许解析器将消

① `http://www.nasdaqtrader.com/content/technicalsupport/specifications/dataproducts/NQTVITCHspecification.pdf`

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

息快速直接转入结构, 从而可以快速查找消息中的特定字段. TotalView-ITCH 采用标准 ITCH 格式, 具有足够订单深度的订单级数据, 并使用一系列订单消息从初始化到执行来跟踪客户订单的生命周期. 回馈是由一系列按顺序排列的消息组成, 每个消息的长度根据消息类型而定. 纳斯达克通过美国新泽西州 Carteret 的美国主数据中心设施来播报市场数据.

表 8.7 和表 8.8 分别显示 ADD 和 EXECUTE 命令消息的规范. 所有整数字段都是位元组由大到小的二进制编码的数字, 所有的 Alpha 字段都是左对齐的和右对齐的 ASCII 字段. 价格是提供了相应精度的整数字段. 表 8.9 表示一个买方订单以价格 114.500 美元的价格买入 500 股的增加订单信息的 ITCH 表示形式.

表 8.7 ITCH5.0 添加订单消息规范. 此消息将为纳斯达克系统接受的未归类订单而生成. 添加订单消息的固定长度为 36 字节

名称	偏移	长度	值	消息
消息类型	0	1	“A”	加入订单——无 MPID 归因消息
股票定位	1	2	整数	定位标识证券
追踪号码	3	2	整数	纳斯达克交易所内部追踪号码
时间戳	5	6	整数	午夜开始的纳秒级
订单号	11	8	整数	在收到新订单时分配给新订单的唯一引用编号
购买/销售指标	19	1	α	加入订单的形式. “B”= 买单; “S”= 卖单
股数	20	4	整数	加入账目的订单中总股数
股票代号	24	8	α	股票象征, 右边填满了空格
价格	32	4	价格 (4)	新订单的显示价格

表 8.8 ITCH5.0 订单执行消息规范. 只要书上的订单全部或部分执行, 就会发送此消息

名称	偏移	长度	值	消息
消息类型	0	1	“A”	加入订单——无 MPID 归因消息
股票定位	1	2	整数	纳斯达克交易所定位标识证券
追踪号码	3	2	整数	内部追踪号码
时间戳	5	6	整数	午夜开始的纳秒级
订单号	11	8	整数	订单参考数字与执行订单相结合
执行股数	19	4	α	执行股数
匹配号码	23	8	整数	纳斯达克产生每日特别的匹配数字给执行

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

表 8.9 ITCH5.0 添加订单消息. “A” 表示增加消息, “B” 表示买单, “S” 表示卖单; “股票定位” 采取整数值, 是识别证券的定位代码; “追踪号码” 是指纳斯达克 OMX 内部追踪号码; “时间戳” 是 0 点以来的纳秒数

添加订单消息	
消息类型	A
股票定位	12345
追踪号码	12233
时间戳	1222444555
订单号	3321123456
购买/销售指标	B
股数	500
股票代号	TEST
价格	114.500

```
0000 41 30 39 2F C9 00 00 48 DD 06 0B 00 00 C5 F4 52 A|12345|12233|1222444555|3321123456
0001 80 42 00 00 01 F4 54 45 53 54 20 20 20 20 00 11 B|500|TEST
0002 78 A8                                          114.500
```

1. 估计以下传输方式在芝加哥 Aurora 和新泽西 Secaucus 之间的传输时间: (i) 铜纤, (ii) 光纤, (iii) 微波, (iv) 毫波.

2. 在 C++ 中, 编译器以行为索引翻译一个二维数组; 也就是下面这个矩阵,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

被存储在连续的内存中, 排序为 $1, 2, 3, \ldots, 8, 9$. 编写一个 C++ 类, 矩阵乘子, 它具有以下接口:

```
template<class T>
class MatrixMultiplier {
public:
  // initialize m_matrix
  MatrixMultiplier(int a_row,int a_col);

  // return the sum of elements by looping
```

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)


```
  // over first the columns then the rows
  T rowMajorSum();

  // return the sum of elements by looping
  // over first the rows then the columns
  T colMajorSum();
private:
  // two dimensional array of
  // a_row - by - a_col
  T** m_matrix;
};
```

首先实现成员函数, 实例化这个类并创建一个大小为 10×50 的矩阵, 然后调用两个成员函数并计算它们的执行时间. 之后重复更大的尺寸, 增加 10%, 达到 1000×5000. 绘制你的结果. 您的结果是否由于数据类型 `int32_t, int64_t` 和 `double` 而变得显著不同? 注意: 你需要实现一种方法, 以确保在每个测试之间, 所有数据缓存都已适当无效.

3. 令 $x \in \mathbb{R}$ 并且 $\hat{x}$ 是其以精度 p 和基数 β 在公式 (8.1) 中 $d_0 \neq 0$ 的浮点数表示形式. 假设 $\hat{x}$ 以环绕的形式被取得. 令 $\delta = |x - \hat{x}|$ 为近似误差的绝对值. 对于 $x \neq 0$, 其对应的相关误差被定义为 $\varepsilon = \delta/|x|$.

   (a) 在 $b > 0$ 的情况下, 所有在区间 $[\hat{x} - b, \hat{x} + b)$ 的实数共享同一表达形式. 从而, $\delta \leqslant \beta^{e-p+1}/2$ 且 $\varepsilon \leqslant \beta^{1-p}/2$. 这些上界什么时候会收紧?[①]

   (b) 令 x 和 y 为两个正实数, 并且 $x > y$. 令 p 位数的 $\hat{x}$ 和 $\hat{y}$ 的差为 $x \ominus y$.

      (i) 令这种差分算子的绝对值误差为 $\delta_\ominus = |x - y - (x \ominus y)|$ 和 $\varepsilon_\ominus = \delta_\ominus |x - y|$. 证明: 存在两个实数 ε_x 和 ε_y, 满足 $\delta_\ominus = x\varepsilon_x - y\varepsilon_y$, 其中 $|\varepsilon_x| \leqslant \beta^{1-p}/2$, $|\varepsilon_y| \leqslant \beta^{1-p}/2$. 因此证明 $\varepsilon_\ominus$ 可以跟 $\beta - 1$ 一样大.

      (ii) 证明 $|\varepsilon_\ominus| \leqslant \beta^{1-p}/2$, 如果差分运算 $\ominus$ 使用一个保护数位, 也就是说 $\hat{x} - \hat{y}$ 是基于 $p+1$ 位数计算的.

① β^{e-p+1} 叫作最后一个单元或者精度最低的单元, 通常用 ulp 表示. $\beta^{1-p}/2$ 叫作单位舍入或者相对精确, 用 rp 表示. 注意, 尽管 rp 和 ulp 都是表示变量的函数, p 和 β, 在 (8.1) 中 ulp 同时也取决于指数 e, 这使得它还是实数估计值的一个函数.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

4. 写出 32 位 IEEE 754 浮点计数法: 牛顿引力常数 $G = 6.6784 \times 10^{-11}$, 普朗克常数 $h = 6.626069 \times 10^{-34}$, 元电荷 $e = 1.602176565 \times 10^{-19}$. 检查这些数字是否会有可能存在补充材料 7 中提到的下溢风险.

5. 芝加哥商品交易所 (CME) 将增量市场数据向下反馈给每个市场数据组.[①]这种双重性是为了尽量减少 UDP 数据包丢失的影响. 这两个频道在 $\pm 2\mu s$ 具有时差的区域内同时播报. 编写一个接收两个反馈的算法, 并将原始数据以最快的速度输出到接收反馈的客户端. 这通常被称为反馈仲裁. 注: 可参阅 CME 消息规范了解可用的市场数据反馈属性.

① CME 还提供了两个市场复苏反馈, 以传播市场数据快照.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

附录 A　鞅　理　论

本附录介绍了我们在本书不同章节使用的离散时间和连续时间鞅的基本理论. 读者需要有一些关于鞅的强定律、二次变化、半鞅和随机积分方面的基本背景.

A.1　离散时间鞅

定义 X 为概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量, 满足 $E|X| < \infty$. 定义 $\mathcal{G} \subset \mathcal{F}$ 为一个 σ 域. 随机变量 Y 称为 X 在给定 $\mathcal{G}$ 的条件期望, 记为 $E(X|\mathcal{G})$, 若它满足以下两个条件:

1. Y 为 $\mathcal{G}$ 可测,

2. 对于所有的 $A \in \mathcal{G}$, $\int_A X dP = \int_A Y dP$.

条件期望即为 Radon-Nikodym 导数 $d\nu/dP$, 其中 $\nu(A) = \int_A X dP$, $A \in \mathcal{G}$. 因此, 除了 P 空集外, 条件期望值是唯一的. 在 $Y = I(B)$ 的特殊情况下, 可以得到给定 $\mathcal{G}$ 时 B 的条件概率函数 $P(B|\mathcal{G})$. 滤波是指 σ 域 $\mathcal{F}_n \subset \mathcal{F}$ 上的一个非减序列 (即 $\mathcal{F}_n \subset \mathcal{F}_{n+1}$). 关于滤波 $\{\mathcal{F}_n\}$, 若 $\{N = n\} \in \mathcal{F}_n$ (对于任意的 n), 则非负整数随机变量 N 称为停时.

与停时 N 相关的是 σ 域

$$\mathcal{F}_N = \{A \in \mathcal{F} : A \cap \{N = n\}\} \in \mathcal{F}_n, \text{ 对于 } n \geqslant 1\}. \tag{A.1}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

当满足如下条件时, 随机变量序列 X_n 称为一致可积鞅:

$$\sup_{n\geqslant 1} E\left\{|X_n| I(|X_n|\geqslant a)\right\}\to 0 \quad \text{当 } a\to\infty. \tag{A.2}$$

一致可积性是鞅理论中的一个重要概念. 鞅理论的一些基本结论包括最优停时定理和鞅收敛定理, 以及经典鞅不等式. 对于序列 $\{X_n\}$, 考虑过滤器 $\{\mathcal{F}_n\}$, 如果 X_n 在 $\mathcal{F}_n$ 上可测且 $E(X_n|\mathcal{F}_{n-1})=X_{n-1}$ 对于任意的 n 均成立, 则称 $\{X_n\}$ 是一个鞅过程. 当上式中的 "=" 替代为 $\geqslant$ (或 $\leqslant$) 时, 鞅过程称为下鞅 (或上鞅) 过程. 当满足对于所有的 n, $E(\varepsilon_n|\mathcal{F}_{n-1})=0$ 几乎处处成立时, 序列 $\{\varepsilon_n\}$ 称为 $\{\mathcal{F}_n\}$ 上的鞅差分过程序列.

定理 1 定义 $\{X_n,\mathcal{F}_n,n\geqslant 1\}$ 为一个下鞅过程, 且 $M\leqslant N$ 为 ($\{\mathcal{F}_n\}$ 上的) 停时. 当 $\{X_{N\wedge n},n\geqslant 1\}$ 一致可积时, $E(X_N|\mathcal{F}_M)\geqslant X_M$ 几乎处处成立, 因此有 $EX_N\geqslant EX_M$.

定理 2 定义 $\{X_n,\mathcal{F}_n,n\geqslant 1\}$ 为一个下鞅过程. 当满足 $\sup\limits_{n\geqslant 1}E(X_n^+)<\infty$ 时, X_n 几乎处处收敛到极限 X_∞, 且有 $E|X_\infty|<\infty$.

定理 3 定义 $\{X_n,\mathcal{F}_n,n\geqslant 1\}$ 为一个下鞅过程. 则对于任意的 $\lambda>0$,

$$\lambda P\left\{\max_{1\leqslant i\leqslant n} X_i>\lambda\right\}\leqslant E\left\{X_n I\left(\max_{1\leqslant i\leqslant n} X_i>\lambda\right)\right\}\leqslant EX_n^+,$$

$$\lambda P\left\{\min_{1\leqslant i\leqslant n} X_i<-\lambda\right\}\leqslant EX_n^+-EX_1.$$

此外, 对于任意的 $p>1$,

$$E\left(\max_{1\leqslant i\leqslant n} X_i^+\right)^p\leqslant\left(\frac{p}{p-1}\right)^p E(X_n^+)^p.$$

定理 4 定义 $\{M_n=\sum_{i=1}^n d_i,\mathcal{F}_n,n\geqslant 1\}$ 为一个鞅过程. 则存在有限的正数 a_p,b_p, 其仅与 p 相关, 且满足

$$a_p E\left(\sum_{i=1}^n d_i^2\right)^{p/2}\leqslant E\max_{j\leqslant n}|M_j|^p\leqslant b_p E\left(\sum_{i=1}^n d_i^2\right)^{p/2},\quad \text{对于所有 } p\geqslant 1.$$

当满足对于任意的 n, $EM_n^2<\infty$ 时, 鞅过程 $(M_n,\mathcal{F}_n,n\geqslant 1)$ 称为平方可积的. 对于滤波 $\{\mathcal{F}_n\}$ 上的随机序列 $\{M_n\}$ 来说, 在信息集 $\{\mathcal{F}_n\}$ 下, 只存在有

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

限个停时 τ_m 能够满足 $\lim\limits_{m\to\infty}\tau_m=\infty$ 处处成立, 且 $\{M_{\tau_m\wedge n},\mathcal{F}_n,n\geqslant 1\}$ 对于任意 $m\geqslant 1$ 成立时, $\{\mathcal{F}_n\}$ 称为局部平方可积鞅. 接下来, 我们对于 Azuma 不等式 (定理 5)、局部可积或平方可积鞅的收敛定理和强定律 (定理 6) 进行介绍. 正如 2.6.3 节的第二段中指出, 定理 6 是 Lai, Wei 和 Ying 定义的随机回归模型理论中的重要方法. 正如 7.8 节补充材料 3 中指出, 定理 5 在 7.6 节中提到的基于暗池的删失数据的 ε 最优配置准则计算中起到重要作用.

定理 5 (Azuma, 1967) 定义 $\{M_n=\sum_{i=1}^n d_i,\mathcal{F}_n,n\geqslant 1\}$ 为一个局部可积鞅, 满足存在非随机常数 $a_i<b_i$, 使得 $a_i\leqslant d_i\leqslant b_i$ 对所有的 $i\geqslant 1$ 成立. 则对于所有的 $x\geqslant 0$,

$$P(|M_n|\geqslant x)\leqslant 2\exp\left(-\frac{2x^2}{\sum_{i=1}^n(b_i-a_i)^2}\right).$$

定理 6 定义 $\{\varepsilon_n\}$ 为一个递增 σ 域 $\{\mathcal{F}_n\}$ 上的鞅差分序列, 满足 $\sup\limits_n E\{\varepsilon_n^2|\mathcal{F}_{n-1}\}<\infty$ 几乎处处成立. 对于每个 n, 定义 u_n 为 $\mathcal{F}_{n-1}$ 可测的随机变量. 则有

$$\sum_1^n u_i\varepsilon_i \text{ 在 } \left\{\sum_1^\infty u_i^2<\infty\right\} \text{ 上几乎处处收敛,} \tag{A.3}$$

对于任意 $\eta>1/2$,

$$\left(\sum_1^n u_i\varepsilon_i\right)\Big/\left\{\left(\sum_1^n u_i^2\right)^{1/2}\left[\log\left(\sum_1^n u_i^2\right)\right]^\eta\right\}\to 0 \quad \text{在 } \left\{\sum_1^\infty u_i^2=\infty\right\} \text{ 上几乎处处成立,} \tag{A.4}$$

因此,

$$\sum_1^n u_i\varepsilon_i=o\left(\sum_1^n u_i^2\right)+O(1) \tag{A.5}$$

以概率 1 成立. 另外,

$$\sum_1^\infty |u_i|\varepsilon_i^2<\infty \text{ 在 } \left\{\sum_1^\infty |u_i|<\infty\right\} \text{ 上几乎处处成立,} \tag{A.6}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

对于任意的 $\rho > 1$,

$$\left(\sum_1^n |u_i|\varepsilon_i^2\right) \Big/ \left(\sum_1^n |u_i|\right)^\rho \to 0 \text{ 在 } \left\{\sum_1^\infty |u_i| = \infty\right\} \text{ 上几乎处处成立.} \tag{A.7}$$

当 $\sup_n E\{|\varepsilon_n|^\alpha|\mathcal{F}_{n-1}\} < \infty$ 对于特定的 $\alpha > 2$ 几乎处处成立时, (A.7) 式可以进一步写成

$$\limsup_{n\to\infty} \left(\sum_1^n |u_i|\varepsilon_i^2\right) \Big/ \left(\sum_1^n |u_i|\right) < \infty$$
$$\text{在 } \left\{\sup_i |u_i| < \infty, \sum_1^\infty |u_i| = \infty\right\} \text{ 上几乎处处成立.} \tag{A.8}$$

A.2 连续时间鞅

当样本是几乎处处右连续时, A.1 节中的基本的鞅理论可以很容易地扩展到连续时间鞅/下鞅/上鞅过程中. 特别地, 这些过程存在左极限, 且为 *càdlàg* 过程. 此处我们对于一些结论进行了归纳. 连续时间鞅的一些具体介绍可以参见 Elliott (1982), Karatzas 和 Shreve (1991) 以及 Revuz 和 Yor (1999). 当满足 $\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{\epsilon>0} \mathcal{F}_{t+\epsilon}$ 时, 滤波 $(\mathcal{F}_t)$ 过程为*右连续*. 若 $\mathcal{F}_0$ 包含 $\mathcal{F}$ 中所有的 P 空集 (概率为 0 测度), 滤波 $(\mathcal{F}_t)$ 过程为*完全的*. 接下来的论述中, 我们假设 $\{X_t, t \geqslant 0\}$ 并且适应于右连续且完全的滤波 $\{\mathcal{F}_t\}$. 当 $\{X_t, t \geqslant 0\}$ 在 $\mathcal{F}$ 上可测时, 称 $\{X_t\}$ 在滤波 $\{\mathcal{F}_t\}$ 上是*适应的*. 当 $\Omega \times [0,\infty)$ 上的 σ 域由一个在 $(0,\infty)$ 左连续的适应过程空间生成时, 该 σ 域为*可预测的* σ 域. 当 $\{X_t\}$ 在从 $\Omega \times [0,\infty)$ 到 $\mathbb{R}$ 上的映射 $(\omega, t) \mapsto X_t(\omega)$ 在可预测的 σ 域上可测时, $\{X_t\}$ 是一个可预测的过程. 定义在 $[0,\infty]$ 的变量 T, (i) 是一个 ($\{\mathcal{F}_t\}$ 上的) *停时过程*, 如果满足对于任意 $t \geqslant 0$, $\{T \leqslant t\} \in \mathcal{F}_t$; (ii) 是一个*最优时间*, 如果满足对于任意 $t > 0$, $\{T < t\} \in \mathcal{F}_t$; (iii) 是一个可预测时间, 如果存在一个递增的停时序列 T_n, 满足在 $\{T > 0\}$ 上 $T_n < T$, 且 $\lim_{n\to\infty} T_n = T$ 几乎处处成立.

随机序列 $X = \{X_t, t \geqslant 0\}$ 称为*可测的*, 当满足从 $\Omega \times [0,\infty)$ 到 $\mathbb{R}$ 上的映射 $(\omega, t) \mapsto X_t(\omega)$ 是可测的. 对于一个可测序列, 存在一个可预测过程 $Y = \{Y_t, t \geqslant 0\}$, 满足对于任意的可预测时间 T,

$$E\{X_T I(T < \infty)|\mathcal{F}_{T-}\} = Y_T I(T < \infty) \quad \text{几乎处处成立.} \tag{A.9}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

Y 过程是唯一存在的, 称为 X 的可预测投影, 记作 X^{Π}. 假设 $A=\{A_t, t\geqslant 0\}$ 是一个非减、可测序列, 满足 $E(A_\infty)<\infty$, 则存在一个唯一的非减可测序列 $A^{(p)}$ 满足对于所有的有界可测过程 $\{X_t, t\geqslant 0\}$,

$$E(X_t^{\Pi}A_t)=E(X_tA_t^{(p)}). \tag{A.10}$$

$A^{(p)}$ 过程称为 A 的可预测对偶投影, 具体可参见 Elliott (1982, 第 72 页).

Doob-Meyer 分解及其应用 定义 $\mathcal{T}_a$ 为一类停时, 满足对于所有的 $T\in\mathcal{T}_a$, $P(T\leqslant a)=1$. 当一个在滤波 $\{\mathcal{F}_t\}$ 上适应的右连续过程 $\{X_t, t\geqslant 0\}$, 满足对于所有的 $a>0$, $\{X_T, T\in\mathcal{T}_a\}$ 是一致可积时, 该过程即属于 DL 类. 当 $\{X_t, \mathcal{F}_t, t\geqslant 0\}$ 是一个非负的连续的下鞅过程时, 该过程即属于 DL 类. 根据 *Doob-Meyer 分解定理*, 当右连续下鞅过程 $\{X_t, \mathcal{F}_t, t\geqslant 0\}$ 属于 DL 类时, 只可以分解为

$$X_t=M_t+A_t, \tag{A.11}$$

其中 $\{M_t, \mathcal{F}_t, t\geqslant 0\}$ 是一个右连续鞅, 满足 $M_0=0$ 且 A_t 可预测、非减且右连续. 另外, 该分解是唯一的. 若存在另一个分解 $X_t=M_t'+A_t'$, 则 $P\{M_t=M_t', A_t=A_t',$ 对于所有的 $t\}=1$. *Doob-Meyer 分解的过程* A_t 称为下鞅 $\{X_t, \mathcal{F}_t, t\geqslant 0\}$ 的补偿.

假设 $\{M_t, \mathcal{F}_t, t\geqslant 0\}$ 是一个右连续的、平方可积的鞅过程, 即 $EM_t^2<\infty$ 对于所有的 t 成立. 由于 M_t^2 是右连续的非负下鞅过程 (根据 Jensen 不等式), 则其可以进行 Doob-Meyer 分解, 其补偿称为可预测的变分过程并表示为 $\langle M\rangle_t$, 即 $M_t^2-\langle M\rangle_t$ 是一个鞅. 如果 $\{N_t, \mathcal{F}_t, t\geqslant 0\}$ 是另一个连续的平方可积的鞅过程, 那么 $(M_t+N_t)^2-\langle M+N\rangle_t$ 和 $(M_t-N_t)^2-\langle M-N\rangle_t$ 是鞅过程, 且可预测的变分过程 $\langle M, N\rangle_t$ 定义如下:

$$\langle M, N\rangle_t=\frac{1}{4}\left\{\langle M+N\rangle_t-\langle M-N\rangle_t\right\}, \quad t\geqslant 0. \tag{A.12}$$

定义 $\mathcal{M}$ 为满足 $M_0=0$ 的所有右连续、平方可积的鞅过程 M 的线性空间集. 对于 $(\Omega, \mathcal{F}, P)$ 上的两个过程 X 和 Y, 当满足 $P(X_t=Y_t,$ 对于所有 $t\geqslant 0)=1$ 时, 称 X 和 Y 不可区分. 定义 $\mathcal{M}$ 的范式为

$$\|M\|=\sum_{n=1}^{\infty}\frac{\min\left(\sqrt{EM_n^2}, 1\right)}{2^n}. \tag{A.13}$$

则可以定义 $\mathcal{M}$ 上的度量 $\rho(M, M')=\|M-M'\|$ 以判断两个序列是否可区分. 对于 $\mathcal{M}$ 的子集 $\mathcal{H}$, 当其为度量空间上的闭集, 且满足如下两个条件时, $\mathcal{M}$ 为一个稳定空间:

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

1. $M \in \mathcal{H} \Rightarrow M^T \in \mathcal{H}$ 对于所有的停时 T 都成立, 其中 $M_t^T = M_{T\wedge t}$;

2. $M \in \mathcal{H}$ 和 $A \in \mathcal{F}_0 \Rightarrow MI(A) \in \mathcal{H}$.

对于 $\mathcal{M}$ 上的两个鞅过程 M, N, 当满足 $\langle M, N\rangle_t = 0$ 对于所有的 $t \geqslant 0$ 几乎处处成立时, 或等价地, 当 $\{M_tN_t, \mathcal{F}_t, t \geqslant 0\}$ 是一个鞅过程时, 两个鞅过程正交. 当 $\mathcal{H}$ 是 $\mathcal{M}$ 上的一个稳定子空间时, 有

$$\mathcal{H}^{\perp} := \{N \in \mathcal{M} : N \text{ 正交于 } M \text{ 对于所有的 } M \in \mathcal{H} \text{ 成立}\}. \tag{A.14}$$

此外, 每一个 $X \in \mathcal{M}$ 具有唯一的 (可识别性) 分解

$$X = M + N, \quad \text{其中 } M \in \mathcal{H} \text{ 且 } N \in \mathcal{H}^{\perp}. \tag{A.15}$$

接下来, 我们将 M 进一步分解为不连续的跳跃过程和一个 "连续的部分" M^c, 其中 $M \in \mathcal{M}$ (或者一个更一般的 M, 定义为 $\mathcal{M}$ 的元素的几乎处处极限). 连续的部分与 (A.15) 的分解相关, 且 $\mathcal{H} = \mathcal{M}^c$, 其中

$$\mathcal{M}^c = \{M \in \mathcal{M} : M \text{ 的样本路径连续}\}. \tag{A.16}$$

$\mathcal{M}^c$ 是 $\mathcal{M}$ 上的一个稳定的子空间, 因此由 (A.15) 可知, 每一个 $M \in \mathcal{M}$ 存在一个唯一的分解

$$M = M^c + M^d, \quad M^c \in \mathcal{M}^c \text{ 且 } M^d \in (\mathcal{M}^c)^{\perp}. \tag{A.17}$$

M^c 称为 M 的连续的部分, M^d 称为 M 的不连续的部分. M^c 和 M^d 是正交鞅.

对于 $M \in \mathcal{M}$ 和 $t > 0$, 定义 Π 为 $[0, t]$ 上的一个分割 $0 = t_0 < t_1 < \cdots < t_k = t$. 则 $\|\Pi\| := \max\limits_{1\leqslant i\leqslant k} |t_i - t_{i-1}| \to 0$, $\sum_{i=1}^{k}(M_{t_i} - M_{t_{i-1}})^2$ 依概率收敛, 极限值记为 $[M]_t$. 随机变量 $[M]_t$ 是 M 上的二次变分过程. 对于 $M \in \mathcal{M}^c$, $[M] = \langle M\rangle$. 更一般地, 对于 $M \in \mathcal{M}$,

$$[M]_t = \langle M^c\rangle_t + \sum_{0<s\leqslant t} (\triangle M_s)^2, \tag{A.18}$$

其中 $\triangle M_s = M_s - M_{s-}$, 注意到 $M_{s-} = \lim\limits_{u\uparrow s} M_u$ 只在 M 是 càdlàg 时存在. $\mathcal{M}$ 上的 M 和 N 的平方协变差定义为 $[M, N]_t = \{[M + N] - [M - N]\}/4$. (A.18)可以进一步拓展为

$$[M, N]_t = \langle M^c, N^c\rangle_t + \sum_{0<s\leqslant t} (\triangle M_s)(\triangle N_s). \tag{A.19}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

infoway.io

实时金融行情数据接口

低延迟 • 高可用 • 全覆盖 **专为量化交易设计的行情API**

全市场覆盖 一个接口查所有

股票API
涵盖A股、港股、美股等多个市场。

外汇API
包含56组主要货币对，高频更新。

加密货币API
超过100个币种，全天24小时实时报价。

商品期货API
贵金属、能源天然气、期货指数等

Infoway API是一款专门为量化交易设计的行情接口。支持逐笔成交/订单簿快照等颗粒度，兼容RESTful API和WebSocket，满足高频量化、资产管理系统、行情终端等场景需求。

全域市场一网打尽

每日处理4300万条市场数据流，开发者无需对接多个数据源即可构建多资产策略系统。

简单易接入

我们的API设计注重易用性，支持多种编程语言，确保无论您使用什么开发环境，都能快速将实时行情数据集成到您的应用中。

实时行情接口：www.infoway.io

免费试用！

注册后自动获得试用资格，无需实名认证。（**本活动长期有效**）

联系我们

官网地址：https://infoway.io
客服邮箱：hi@infoway.io



实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

上面的可积性假设可以放宽到*局部可积条件*. 当存在停时序列 T_n, 满足 $\{M_{T_n\wedge t}, \mathcal{F}_t, t \geqslant 0\}$ 是一个鞅过程 (或者是平方可积的鞅过程, 或有界鞅过程) 时, $\{M_t, \mathcal{F}_t, t \geqslant 0\}$ 称为*局部鞅* (或*局部平方可积鞅*, 或*局部有界鞅*). 在此情况下, 我们可以对于局部平均可积鞅, 重新定义 $\langle M\rangle_t$, $\langle M, N\rangle_t$, $[M]_t$, $[M, N]_t$, M^c 和 M^d. 此外, 连续局部鞅 M_t 可以表示为时变的布朗运动如下:

$$M_t = W_{\langle M\rangle_t}. \tag{A.20}$$

当 V 是一个在有限区间内有限变分的适应过程时, 其对偶可预测的投影过程 $V^{(p)}$ (参见 (A.10)) 本质上是唯一在有限区间内有限变分的可预测过程, 满足 $V - V^{(p)}$ 是一个局部鞅; 参见 Elliott (1982, 第 121 页).

随机积分和半鞅的 Itô 公式 定义 $\mathcal{M}^{\text{loc}}$ 为右连续、局部平方可积鞅的类, 满足 $M_0 = 0$. 令 M, N 属于 $\mathcal{M}^{\text{loc}}$, H, K 是两个可测的过程. 在 $0 \leqslant s \leqslant t$ 时, 令 $\langle M, N\rangle_{s,t} = \langle M, N\rangle_t - \langle M, N\rangle_s$. 注意到

$$\langle M + rN\rangle_{s,t} = \langle M, M\rangle_{s,t} + 2r\langle M, N\rangle_{s,t} + r^2\langle N, N\rangle_{s,t}$$

是一个 r 的非负二次函数. 因此

$$|\langle M, N\rangle_{s,t}| \leqslant \{\langle M, M\rangle_{s,t}\}^{1/2} \{\langle N, N\rangle_{s,t}\}^{1/2}. \tag{A.21}$$

因此, 我们可以用一个求和来对 Lebesgue-Stieltjes 积分的下界进行近似. 根据 (A.21), 可得如下 *Kunita-Watanabe 不等式*:

$$\int_0^t |H_s||K_s| d\overline{\langle M, N\rangle}_s \leqslant \left(\int_0^t H_s^2 d\langle M\rangle_s\right)^{1/2} \left(\int_0^t K_s^2 d\langle N\rangle_s\right)^{1/2}, \tag{A.22}$$

其中 $\bar{\zeta}_t$ 表示 ζ 在 $[0, t]$ 区间的总变化.

接下来, 我们定义随机积分 $\int_0^t X_s dY_s$, 其中 $X = \{X_s, 0 \leqslant s \leqslant t\}$, $Y = \{Y_s, 0 \leqslant s \leqslant t\}$. 当 Y 在 $[0, t]$ 上满足有界变差时, 该积分可以看作一个 $[0, t]$ 上依路径的 Lebesgue-Stieltjes 积分. 当 Y 是一个右连续、平方可积的鞅, X 是一个可预测过程, 满足 $\int_0^t X_s^2 d\langle Y\rangle_s < \infty$ 几乎处处成立时, $\int_0^t X_s dY_s$ 可以定义积分的 (依概率) 极限 (可以简化为求和), 该积分的积分项为阶梯函数, 且在 L_2 上收敛于 X. 对于一个适应于 $\{\mathcal{F}_t, t \geqslant 0\}$ 的过程 Y, 若它可以分解为 $Y_t = Y_0 + M_t + V_t$, 其中 M 是一个局部可积的连续鞅过程, V_t 是一个在有限区间内有限变分的适应 càdlàg 过程, 且 $M_0 = V_0 = 0$, Y 可以称为*半鞅*. 因此, 我们可以定义 $\int_0^t X_s dY_s$, 其中 X 是一个可预测的局部有界过程, Y 是一个半鞅. 另外, $\left\{\int_0^t X_s dY_s, t \geqslant 0\right\}$ 也是一个半鞅.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

定理 7 (半鞅的 Itô 公式) 定义 $X(t)=(X_1(t),\ldots,X_m(t))$, $t\geqslant 0$ 为一个向量过程, X_i 是一个半鞅过程, 可以分解为 $X_i(0)+M_i(t)+V_i(t)$. 定义 $f:[0,\infty)\times\mathbb{R}^m\to\mathbb{R}$ 属于 $C^{1,2}$ 类 (即 $f(t,x)$ 对 x 两次连续可微, 对 t 连续可微). 则 $\{f(t,X(t)),t\geqslant 0\}$ 是一个半鞅, 且有

$$\begin{aligned}f\left(t,X(t)\right)=f\left(0,X(0)\right)&+\sum_{i=1}^{m}\int_{0+}^{t}\frac{\partial}{\partial x_i}f\left(s,X(s-)\right)dX_i(s)\\&+\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\int_{0+}^{t}\frac{\partial^2}{\partial x_i\partial x_j}f\left(s,X(s-)\right)d\langle M_i^c,M_j^c\rangle_s\\&+\sum_{0<s\leqslant t}\Bigg\{f\left(s,X(s)\right)-f\left(s,X(s-)\right)\\&-\sum_{i=1}^{m}(\partial/\partial x_i)f\left(s,X(s-)\right)\triangle X_i(s)\Bigg\}.\end{aligned}$$

定理 7 的一个重要推论是半鞅 X 和 Y 的乘积法则: XY 也是一个半鞅, 且

$$d(X_tY_t)=X_{t-}dY_t+Y_{t-}dX_t+d[X,Y]_t. \tag{A.23}$$

特别地, 对于一个局部可积的、连续的鞅过程 M, 假定 (A.23) 中 $X=Y=M$, 则有

$$[M]_t=M_t^2-2\int_0^t M_{s-}dM_s. \tag{A.24}$$

由 (A.24) 可知 $M^2-[M]$ 是一个鞅过程, 但是 $[M]$ 是不可预测的. 由于 $M^2-\langle M\rangle$ 是一个鞅过程, 而且 $\langle M\rangle$ 是可以预测的, 则 $\langle M\rangle_t$ 是非减过程 $[M]_t$ 的补偿.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

附录 B　马尔可夫链

对于离散状态空间 $\mathbb{S}$ 的随机过程 $(X_t)_{t\geqslant 0}$, 当满足对于所有 $s,t\geqslant 0$, $i,j\in\mathbb{S}$, 以下条件成立时, $(X_t)_{t\geqslant 0}$ 称为连续时间马尔可夫链 (CTMC):

$$P\left(X_{s+t}=j|X_s=i,\mathcal{F}_s^X\right)=P\left(X_{s+t}=j|X_s=i\right)=P_{ij}(t),$$

其中 $\mathcal{F}^X$ 是由 $(X_t)_{t\geqslant 0}$ 产生 σ 域. 在本节中, 我们将介绍 CTMC 的生成元和离散嵌入链, 以及离散马尔可夫链的势论和无限维马尔可夫决策理论中的 Bellman 方程. 本附录对于 5.5.3 节和 6.5.2 节的 LOB 建模和订单配置提供了基本的数学基础.

B.1　CTMC 生成元 $\boldsymbol{Q}$

一个 CTMC 可以由其生成元 $\boldsymbol{Q}=(q_{ij})_{i,j\in\mathbb{S}}$ 刻画, 其中

$$P(X_{t+h}=j|X_t=i)=\delta_{ij}+q_{ij}h+o(h),\quad q_{ii}=-\sum_{j\neq i}q_{ij}<0,\quad q_{ij}\geqslant 0.$$

当 $i=j$ (或 $i\neq j$) 时, $\delta_{ij}=1$ (或 0), 并且 q_{ij} 称为从状态 i 到 j 的转移强度.

嵌入链　对于离散时间马尔可夫链 $(\tilde{X})_{n=0,1,\ldots}$, 如果其转移矩阵 $\boldsymbol{P}=(p_{ij})_{i,j\in\mathbb{S}}$ 满足 $i\neq j$ 时, $p_{ij}=\dfrac{q_{ij}}{\sum_{k\neq i}q_{ik}}$, 而 $p_{ii}=0$, 则 $(\tilde{X})$ 称为与 $(X_t)_{t\geqslant 0}$ 相关联的嵌入链. 事实上, $\tilde{X}_1=X_{\tau_1}$, $\tilde{X}_2=X_{\tau_2},\ldots$, 其中 $\tau_i=\min\{t>\tau_{i-1}:X_t\neq X_{\tau_{i-1}}\}$, $\tau_0=0$. 如果 X_t 是遍历的, 则它与其嵌入链一样, 具有平稳的分布.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

B.2 马尔可夫链的势论

假设 $X=(X_n)_{n\geqslant 1}$ 是可数状态空间 $\mathbb{S}$ 中的一个马尔可夫链, 其转换矩阵为 $P=(P_{ij})_{i,j\in\mathbb{S}}$. 令 $D\subset\mathbb{S}$, ∂D 为 D 的边界. 令 $\tau=\inf\{n: X_n\notin D\}=\inf\{n: X_n\in\partial D\}$. 给定一个非负运行成本函数 $V_c=(V_c(s))_{s\in D}\geqslant 0$ 和一个非负的末期成本 $V_T=(V_T(s))_{s\in\partial D}\geqslant 0$ 时, 定义从状态 s 到 ∂D 的预期成本为

$$\Phi(s)=\mathbb{E}\left[\sum_{n=1}^{\tau-1}V_c(X_n)+V_T(X_\tau)I_{\{\tau<\infty\}}|X_0=s\right]. \tag{B.1}$$

则 Φ 服从如下线性方程组:

$$\begin{cases}\Phi=P\Phi+V_c, & \text{在 } D \text{ 中},\\ \Phi=V_T, & \text{在 } \partial D \text{ 上}.\end{cases}$$

B.3 马尔可夫决策理论

假设 $X=(X_n)_{n\geqslant 1}$ 是有限状态空间上的马尔可夫链 $\mathbb{S}$, 其转移矩阵为 P. 假设 $\mathbb{A}=\mathbb{C}\cup\mathbb{T}$ 是一个有限集, 其中 $\mathbb{C}$ 是 (在末期行动之前) 一系列持续的行动, 而 $\mathbb{T}$ 是末期行动集. 定义 $A:\mathbb{S}\mapsto 2^{\mathbb{A}}$ 为将非空行为集合 $A(s)$ 关联到每个状态 $s\in\mathbb{S}$ 的函数. 此处 $2^{\mathbb{A}}$ 是包含 $\mathbb{A}$ 的所有子集的幂集. 显然, $C(s)=A(s)\cap\mathbb{C}$ 并且 $T(s)=A(s)\cap\mathbb{T}$. 对每个 $s\in\mathbb{S}$ 和 $a\in C(s)$, 当选择行动 a 时, 从 s 到 s' 的转移概率表示为 $P_{ss'}(a)$. 另外, 假设以下的正则性条件成立:

- $V_c(s,a)$ 是有界且非负的; 它是状态 s 下 $C(s)$ 中行动 a 继续的成本.
- $V_T(s,a)$ 是有界且非负的; 它是状态 s 下 $T(s)$ 中行动 a 终止的成本.
- $\alpha=(\alpha_0,\alpha_1,\dots)$ 是一系列行动, 其中 $\alpha_n:\mathbb{S}^{n+1}\to\mathbb{A}$. 因此 $\alpha_n(s_0,\dots,s_n)\in A(s_n)$ 对每个 $n\geqslant 0$ 和 $(s_0,\dots,s_n)\in\mathbb{S}^{n+1}$ 成立.
- $V(s,\alpha)$ 是从 $X_0=s$ 开始的预期成本, 服从策略 α 直到终止.

如果对于任何策略 α, 如下条件成立, 则一个策略 α^* 被称为最佳策略:

$$V(s,\alpha^*)\leqslant V(s,\alpha).$$

最优期望成本 $V^*(s)$ 定义如下:

$$V^*(s)=\inf_{\alpha}V(s,\alpha).$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

Bellman 方程 $V^*(s)$ 即为如下 *Bellman* 方程的解:

$$V(s) = \min \left[\min_{\alpha \in C(s)} V_c(s, \alpha) + \sum_{s' \in \mathbb{S}} P_{ss'}(\alpha) V(s'), \min_{\alpha \in T(s)} V_T(s, \alpha) \right].$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

附录 C　双随机自激点过程

在本附录中, 我们对于 Hawkes 过程的背景与实现方法进行介绍, Hawkes 过程在 4.6.2 节和 5.3 节中被用于对交易间隔和 LOB 对手方的市场订单的到达过程进行建模. 首先, 我们将到达时间 $t_1, t_2, \dots$ 与 4.6.2 节中提到的如下过程相关联:

$$N_t = \sum_{i=1}^{\infty} I_{\{t_i<t\}} = \max\{i : t_i \leqslant t\}. \tag{C.1}$$

(C.1) 中的第一个等式说明 N_t 是一个*计数过程* (即指示函数之和), 第二个等式说明该过程也是一个*更新过程*, 其与到达时间间隔 $t_i - t_{i-1}$ 的到达理论相关. C.1 节在附录 A 的基础上, 介绍了计数过程的鞅理论. C.2 节利用该理论建立了双随机点过程模型, 4.6.2 节的自激励点过程 (SEPP) 是其特例. C.3 节介绍了 SEPP 过程的似然推断, 如满足 (4.51) 中的 $\phi(x) = x$ 的情况或 5.3 节中多元设定的情况. C.4 节介绍了利用蒙特卡罗方法来对于点过程进行模拟的方法.

C.1　鞅理论和多元计算过程的补偿

基于 A.2 节的连续时间鞅理论, *多元计算过程* $\boldsymbol{N} = (N_1, \dots, N_k)$ 是一个适应于滤波 $\{\mathcal{F}_t\}$ 的 k 维 *càdlàg* 过程, 满足 (a) $\boldsymbol{N}(0) = \mathbf{0}$; (b) N_i 是一个非减的分段常数, 且阶跃幅度为 1; (c) $\boldsymbol{N}$ 的任意两个组成不会同时发生阶跃. 若 $\mathcal{F}_t$ 由 $\{\boldsymbol{N}_s, s \leqslant t\}$ 生成, 则该滤波称为 "自激" 滤波. 由 (c) 可知, $N_\cdot = \sum_{i=1}^{k} N_i$ 是一个计数过程, 且存在一系列的停时 $0 < t_1 < t_2 < \cdots$, 定义 $J_1, J_2, \dots$ 取值

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

为 $\{1,2,\ldots,k\}$. 则 $N.$ 为在时间 t_i 时发生跳跃, 且 J_i 为此次跳跃的表示.

首先, 考虑 $k=1$ 时的情形. 由于 $N(t)$ 是一个右连续非减序列, 且其适应于滤波 $\{\mathcal{F}_t\}$, 因此该过程是一个下鞅. 根据 Doob-Meyeer 分解 (A.11), $N(t)$ 可以分解为 $N(t)=M_t+A_t$, 其中 M_t 是一个鞅过程, 满足 $M_0=0$, A_t 是一个可预测的非减右连续的补偿. 假设 A_t 是一个连续过程, 且其导数在 Lebesgue 测度下几乎处处连续. 则由于 N_t 可以写成一系列的指示函数 $I_{\{t_i<t\}}$ 之和, 则可知 A_t 的导数等于 (4.50) 中定义的 λ_t 几乎处处成立. 因此, $A_t=\int_0^t\lambda(s)ds$, 记为 $\Lambda(t)$, 4.6.2 节中称此为聚合强度, 且表明 $N(t)-\Lambda(t)$ 是一个鞅过程.

接下来, 我们讨论一般化的 k 维过程, 且假设 (4.50) 式中当 N 替代为 $\boldsymbol{N}$ 时, 其极限几乎处处存在. 定义极限为 $\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\ldots,\lambda_k)$. 则对于 $i=1,\ldots,k$, N_i 是一个计数过程, 其补偿为 $\Lambda_i(t)=\int_0^t\lambda_i(s)ds$. 定义 $M_i(t)=N_i(t)-\Lambda_i(t)$ 为 Doob-Meyer 分解中的鞅过程. 根据本节的第一段中的 (b) 和 (c) 可知, $\langle M_i\rangle_t=\Lambda_i(t)$, 且对于任意 $i\neq j$, $\langle M_i,M_j\rangle_t=0$. 更一般地, 在不假设 N_i 的补偿 Λ_i 绝对连续的情况下, 可得

$$\begin{aligned}\langle M_i\rangle_t&=\Lambda_i(t)=\int_0^t\Delta\Lambda_i(s)d\Lambda_i(s),\\ \langle M_i,M_j\rangle_t&=-\int_0^t\Delta\Lambda_i(s)d\Lambda_j(s),\ \text{对于 } i\neq j,\end{aligned}\tag{C.2}$$

对于一个非减函数 A, $\Delta A(s)=A(s)-A(s-)$, $A(s-)=\lim\limits_{t\uparrow s}A(t)$. 尽管 $\boldsymbol{N}$ 包含跳跃, $\boldsymbol{\Lambda}$ 可以是完全连续的. 一个典型的例子是泊松点过程, 其具有服从泊松分布的独立增量, 即 $k=1$ 时, $N(t)-N(s)$ 独立于 $\mathcal{F}_s, 0\leqslant s\leqslant t$, 且 $P(N(t)-N(s)=n|\mathcal{F}_s)=e^{-\lambda(t-s)}\dfrac{[\lambda(t-s)]^n}{n!}$. 此时 (4.50) 成立, 且 $\lambda(t)\equiv\lambda$, 则有 $\Lambda(t)=\lambda t$. Watanabe (1964) 对上述泊松过程进行了如下的鞅刻画: 在 $\lambda>0$ 时, 当且仅当 $N(t)-\lambda t$ 是一个 $\{\mathcal{F}_t\}$ 上的鞅过程时, $N(t)$ 是一个强度为 λ 的泊松过程.

C.2 双随机点过程模型

由于泊松过程 $N(t), t\geqslant 0$ 的到达时间 $t_1<t_2<\cdots$ 是 $(0,\infty)$ 上的随机点, 因此泊松过程也称为*泊松点过程*. 此时到达时间的间隔服从独立同分布指数分布. 对于一个一般的计数过程来说, 到达时间的间隔不一定具有相同的分布, 此时该过程也称为*非同质泊松分布*. 当其强度过程 $\lambda(t)$ 也是随机过程时, 点过程称为*双随机过程*. 在对双随机过程进行建模时, 需要首先对于

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$\lambda(t)$ 进行建模, 因为当跳跃发生时, 跳跃幅度需要为 1. (4.51) 即为这样的模型, 它是 (4.52) 中 Hawkes 过程的一个特例, 它满足随机微分方程 $d\lambda(t) = \beta[\mu - \lambda(t)]dt + \alpha dN(t)$. 当 $\alpha/\beta < 1$ 时, Hawkes 过程 (4.52) 是平稳的, 且其稳定状态期望值为 $\mu/(1-\alpha/\beta)$. 另外, 其积分强度为

$$\Lambda(t) = \int_0^t \lambda(u)du = \mu t - \frac{\alpha}{\beta}\sum_{i:t_i<t}(e^{-\beta(t-t_i)} - 1),$$

该方程可以利用强度过程的分段积分来求得.

多元情况的拓展 多元双随机计数过程 $\boldsymbol{N} = (N_1, \ldots, N_k)$ 可以利用如下强度过程来进行建模:

$$\lambda_i(t) = \phi_i\left(\mu_i(t) + \sum_{j=1}^{k}\int_{u<t} h_{ij}(t-u)dN_j(u)\right), \quad i = 1, \ldots, k, \tag{C.3}$$

其中 λ_i 不仅通过 h_{ii} 与 $\boldsymbol{N}$ 的第 i 个组成部分的历史事件相关, 当 $j \neq i$ 时, 也通过 h_{ij} 与 $\boldsymbol{N}$ 的其他部分的历史事件相关. 多元模型的一个特例是 k 维 Hawkes 过程, 其强度过程为

$$\lambda_i(t) = \mu_i(t) + \sum_{j=1}^{k}\alpha_{ij}\int_{u<t} e^{-\beta_{ij}(t-u)}dN_j(u), \quad i = 1, \ldots, k, \tag{C.4}$$

其平稳条件是 $\|\boldsymbol{A}\| < 1$, 其中 $\boldsymbol{A}$ 是一个 $k \times k$ 矩阵, 其第 (i,j) 个元素为 $\alpha_{ij}/\beta_{ij}, 1 \leqslant i, j \leqslant k$.

C.3 点过程模型的似然估计

定义一个单变量点过程模型 $N(t)$ 的事件时间 (或到达时间) 为 $t_1 < t_2 < \cdots < t_n$, 其强度函数为 $\lambda_\theta(t)$. 由于 $N(t_n) = n$, 可得似然函数为

$$L(\theta) = \exp\left(-\int_0^{t_n}\lambda_\theta(u)du\right)\prod_{j=1}^{n}\lambda_\theta(t_j),$$

因此对数似然函数 $\mathcal{L}(\theta) = \log L$ 为

$$\mathcal{L}(\theta) = -\int_0^{t_n}\lambda_\theta(u)du + \int_0^{t_n}\log\lambda_\theta(u)dN(u); \tag{C.5}$$

具体可参见 Daley 和 Vere-Jones (2003, 命题 7.2 III). 对于具有参数 $\theta = (\mu, \alpha, \beta)$ 的指数核函数的 Hawkes 过程, 对数似然函数不是凹的, 而对数似然

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

函数 (C.5), 由参数为 λ_t 的 (4.52) 给出, 且可简化为

$$\begin{aligned}\mathcal{L}(\theta) = & -\int_0^{t_n} \underbrace{\left(\mu + \int_{u<t} \alpha e^{-\beta(t-u)} dN(u)\right)}_{\text{关于 } \beta \text{ 的凹函数}} dt \\ & + \int_0^{t_n} \underbrace{\log\left(\mu + \int_{u<t} \alpha e^{-\beta(t-u)} dN(u)\right)}_{\text{关于 } \beta \text{ 的凸函数}} dN(u). \end{aligned} \tag{C.6}$$

为计算 MLE, Ogata (1978) 将 (C.6) 改写为

$$\mathcal{L}(\theta) = -\mu t_n - \frac{\alpha}{\beta}\sum_{i=1}^{n}\left(1 - e^{-\beta(t_n - \tau_i)}\right) + \sum_{i=1}^{n}\log\left[\mu + \alpha A_i(\beta)\right], \tag{C.7}$$

其中当 $i \geqslant 2$ 时,

$$A_i(\beta) = \sum_{j:t_j<t_i} e^{-\beta(t_i - t_j)}, A_1(\beta) = 0.$$

梯度矩阵的各项定义为

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mu} &= -t_n + \sum_{i=1}^{n}\left(\frac{1}{\mu + \alpha A_i(\beta)}\right), \\ \frac{\partial \mathcal{L}}{\partial \alpha} &= \frac{1}{\beta}\sum_{i=1}^{n}\left(e^{-\beta(t_n - t_i)} - 1\right) + \sum_{i=1}^{n}\frac{A_i(\beta)}{\mu + A_i(\beta)} \\ \frac{\partial \mathcal{L}}{\partial \beta} &= -\alpha\sum_{i=1}^{n}\left[\frac{1}{\beta}(t_n - t_i)e^{-\beta(t_n - t_i)} + \frac{1}{\beta^2}e^{-\beta(t_n - t_i)}\right] + \sum_{i=1}^{n}\frac{B_i(\beta)}{\mu + \alpha A_i(\beta)}\end{aligned}$$

其中 $B_i(\beta) = \sum_{j:t_j<t_i}(t_i - t_j)e^{-\beta(t_i - t_j)}, i \geqslant 2$, 且 $B_1(\beta) = 0$. Hessian 矩阵的各项定义为

$$\begin{aligned}\frac{\partial^2 \mathcal{L}}{\partial \mu^2} &= \sum_{i=1}^{n}\left(\frac{-1}{(\mu + \alpha A_i(\beta))^2}\right), \quad \frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = -\sum_{i=1}^{n}\left(\frac{A_i(\beta)}{\mu + \alpha A_i(\beta)}\right)^2, \\ \frac{\partial^2 \mathcal{L}}{\partial \mu \partial \alpha} &= \sum_{i=1}^{n}\left(\frac{-A_i(\beta)}{(\mu + \alpha A_i(\beta))^2}\right), \quad \frac{\partial^2 \mathcal{L}}{\partial \mu \partial \beta} = \sum_{i=1}^{n}\left(\frac{\alpha B_i(\beta)}{(\mu + \alpha A_i(\beta))^2}\right), \\ \frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \beta} = & -\sum_{i=1}^{n}\left[\frac{1}{\beta}(t_n - t_i)e^{-\beta(t_n - t_i)} + \frac{1}{\beta^2}\left(e^{-\beta(t_n - t_i)} - 1\right)\right] \\ & + \sum_{i=1}^{n}\left[\frac{-B_i(\beta)}{\mu + \alpha A_i(\beta)} + \frac{\alpha A_i(\beta) B_i(\beta)}{(\mu + \alpha A_i(\beta))^2}\right],\end{aligned}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

$$\frac{\partial^2 \mathcal{L}}{\partial \beta^2} = \alpha \sum_{i=1}^{n} \left[\frac{1}{\beta}(t_n - t_i)^2 e^{-\beta(t_n - t_i)} + \frac{2}{\beta^2}(t_n - t_i) e^{-\beta(t_n - t_i)} + \frac{2}{\beta^2}\left(e^{-\beta(t_n - t_i)} - 1\right) \right] + \sum_{i=1}^{n} \left[\frac{C_i(\beta)}{\mu + \alpha A_i(\beta)} - \left(\frac{\alpha B_i(\beta)}{\mu + \alpha A_i(\beta)} \right)^2 \right].$$

其中

$$C_i(\beta) = \sum_{j: t_j < t_i} (t_i - t_j)^2 e^{-\beta(t_i - t_j)}, i \geqslant 2, \text{ 且 } C_1(\beta) = 0.$$

注意到 $A_i(\beta)$ 可以由 $A_i(\beta) = e^{-\beta(t_i - t_{i-1})}[1 + A_{i-1}(\beta)]$ 递归求得. 类似地, $B_i(\beta), C_i(\beta)$ 也可以递归求得.

处理非凹问题的一种简单方法是首先根据对强度衰减的估计确定 β 的先验值. 当 β 确定时, 对数似然函数是退化的参数向量的凹函数, 此时对单变量参数 β 进行优化是相对直观的. Ozaki (1979) 介绍了最大化 $\mathcal{L}$ 的一些其他方法, 包括 Davidon-Fletcher-Powell 方法.

多变量情形拓展 (C.5) 在多变量情形的拓展可以参见 Daley 和 Vere-Jones (2003, 命题 7.3 III). 在 5.3.1 节中提到的具有指数核函数的二元 Hawkes 过程的对数似然函数为 $\mathcal{L}(\theta_1, \theta_2) = \mathcal{L}(\theta_1) + \mathcal{L}(\theta_2)$, 这也是 (5.4) 的基础, 且与 $\mathcal{L}(\theta_2)$ 的公式类似.

MLE 的渐近正态性和 GLR 的 χ^2 近似 在正则性条件下, Ogata (1978) 求解得到似然推论的经典渐近理论 (参见 Lai 和 Xing (2008) 的 2.4 节), 该结论也适用于平稳 Hawkes 过程, 如下所示:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N\left(0, \boldsymbol{I}^{-1}(\theta_0)\right), \quad 2\left[\mathcal{L}(\hat{\theta}) - \mathcal{L}(\theta_0)\right] \xrightarrow{\mathcal{D}} \chi_d^2, \tag{C.8}$$

其中 $\xrightarrow{\mathcal{D}}$ 表示依分布收敛, θ_0 是参数的真实值, $\hat{\theta}$ 是 MLE, $\boldsymbol{I}(\theta_0)$ 是 Fisher 信息矩阵. 由于数据生成机制是一个计数过程, 而不是传统的独立同分布过程, 此时鞅中心极限定理 (参见附录 D) 和鞅强定律 (参见附录 A.1) 可用于证明 (C.8), 而不是在传统设定下使用 MLE 的渐近理论和 GLR 统计量. 在时间序列设定下, 尽管 Fisher 信息矩阵仍然可以基于 $\mathcal{L}(\theta)$ 的 Hessian 矩阵求得 (参见 Lai 和 Xing (2008), 第 61 页), 但此时 $\boldsymbol{I}(\theta_0)$ 很难求得, 且 θ_0 在 (C.8) 中是一个未知参数. 因此, 在进行似然估计时, 通常会基于大数定律, 将 (C.8) 中的 $\boldsymbol{I}(\theta_0)$ 替换为观测到的 Fisher 信息矩阵 $\boldsymbol{\nabla}^2 \mathcal{L}(\hat{\theta})$. 具体可参见 Lai 和 Xing (2008), 第 58, 328 和 329 页.

软件实施 4.6.2 节中提到, 有两个 R 语言程序包可以用于拟合单变量点

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

过程, ppstat 可以对于

$$\lambda_\theta(t) = \phi\left(\boldsymbol{\alpha}^\top \boldsymbol{u}_t + \sum_{i:t_i<t} h_\beta(t-t_i)\right)$$

形式的强度函数进行拟合, 其中 $\boldsymbol{u}_t$ 是时间 t 的外生共生变量矩阵, $h_{\boldsymbol{\beta}}(s) = \boldsymbol{\beta}^\top \boldsymbol{B}(s)$, $\boldsymbol{B}(s)$ 是给定的基准函数向量. 该函数使用 Riemann 求和来对于对数似然函数 (C.5) 中的积分项 $-\int_0^{t_n} \lambda_\theta(s)ds$ 估计. 程序包 NHPoisson 则假设强度函数服从如下对数线性过程:

$$\log(\lambda_{\boldsymbol{\beta}}(t)) = \boldsymbol{\beta}^\top \boldsymbol{x}_t, \tag{C.9}$$

其中 $\boldsymbol{x}_t$ 包含 $t_i I_{\{t_i<t\}}$ 和外生共生变量 $\boldsymbol{u}_t$. 该程序包的输出项包含 $\boldsymbol{\beta}$ 的 MLE, 观测到的信息矩阵的逆, 以及不同时间的强度的置信区间. 该程序包也可以进行变量选择和残差分析.

C.4 双重随机 SEPP 的模拟

Lewis 的 Thinning 方法 SEPP 可以用 Lewis 和 Shedler (1979) 提出的 *Thinning* 算法来模拟, 该算法最初被用来模拟一个非均匀的泊松过程. 事实上, 基于 Lewis 的 Thinning 算法是来源于拒绝抽样技术[①], Ogata (1981) 在此基础上, 修改了 Lewis 的算法, 用来模拟相应的强度函数是随机的情况下的 SEPP. 该算法流程具体如下:

1. 假设事件时间集合 $\{t_i : i = 0, \ldots, j\}$ 已经被模拟, 且 $t_0 = 0$. 令 $s = t_j$ 和 $\delta > 0$ 是一个小的正数, 来模拟 t_{j+1}.

2. 令 $\lambda_* = \max_{t\in(s,s+\delta)} \lambda(t)$.

3. 从参数为 λ_* 的指数分布中抽样 ξ.

4. 若 $\dfrac{\lambda(s+\xi)}{\lambda_*} < 1$, 则进行第 5 步. 否则, $(s, s+\delta)$ 中没有事情发生; 令 $s := s + \delta$, 然后执行第 2 步.

5. 从 $[0, 1]$ 上的均匀分布中抽取 U.

① 在模拟领域的文献中通常讨论拒绝抽样的理论; 有关介绍请参阅 Ross (2015) 的 5.2 节.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

6. 若 $\dfrac{\lambda(s+\xi)}{\lambda_*} \geqslant U$, 则 $t_{j+1} = s + \xi$; 设 $j := j + 1$, 然后执行第 1 步. 否则执行第 3 步.

如果 $\lambda(t)$ 是单调递减 (发生事件时除外), 则 $\lambda_* = \lambda(s)$, 此时 δ 可以任意选取. 然而, 在条件强度 $\lambda(t)$ 单调递增时 (发生事件时除外), δ 的选择是至关重要的. 如果 δ 太小, 那么 λ_* 将会相对较小, 并且 ξ 可能非常大, 以至于变得大于 $s + \delta$. 这导致许多小的时间间隔, 包括一个事件的可能性很低. 如果 δ 太大, 那么 λ_* 将会相对较大, ξ 可能相当大; 因此, 许多潜在的 "事件" 将被削减. 根据 Harte (2010), `PtProcess` 中的模拟算法设置 δ 等于 δ 的指数分布的 70 分位, 其比率等于 $\lambda(s)$. 而且, 由于 Thinning 算法是一个拒绝采样过程, 所以其效率也由步骤 3 和 6 之间的平均迭代次数来决定, 以成功模拟 t_{j+1}.

上述 Thinning 算法可以基于 R 语言中的 `PtProcess` 的程序包实现. Harte (2010) 介绍了 `PtProcess` 在统计推断中的其他应用. 在 R 语言的 `SAPP` 软件包中实现了 SEPP 的改进模拟过程, 该软件包是点过程统计分析的缩写. SAPP 是基于 Ogata 等 (2006) 的 Fortran 77 程序构建的, 以研究地震活动. 另一个 R 语言程序包 `hawkes` 实现了 SEPP 的多变量形式估计. 而对于 4.6.2 节中提到的包含时变的基准强度函数的 Hawkes 的模拟则可以利用 R 语言程序包 `IHSEP` 实现.[①]

Hawkes 过程模拟的最新进展 Møller 和 Rasmussen (2005) 指出, Thinning 算法会受到由 $(s, s+\delta)$ 中的事件和 $(-\infty, s]$ 中的事件的依赖性造成的 "边缘" 效应的影响. 论文提出了一个基于使用具有分支结构的聚类的抽样方案 (即以独立同分布的方式对整个过程进行抽样). 另外, 为了消除边缘效应, 论文中的算法从 $t = -\infty$ 开始, 并且需要一定的平稳条件; 另见 Møller 和 Rasmussen (2006). 另一个改进是由 Giesecke 和 Kim (2007) 和 Giesecke 等 (2011) 提出的, 论文直接通过基本的分析分布函数对持续时间 $t_i - t_{i-1}$ 进行采样. 该方法成功地避免了产生强度路径以及引入离散偏差. 然而, 由于该算法利用 Brent 方法评估分布函数的逆, 因此可能是计算密集型的. Dassios 和 Zhao (2013) 还通过采用基于群集的 Hawkes 过程的定义, 而不是基于 (4.50) 中所述的基于强度的定义, 推导得到另一种精确的模拟算法. 该定义允许使用 Hawkes 过程的隐式分支结构, 以便通过泊松过程模拟的递归实现其模拟; 具体可参见 Dassios 和 Zhao (2013).

① `IHSEP` 是非齐次自激过程的缩写, 由第一作者 Chen 和 Hall (2013) 贡献. 除了模型仿真之外, `IHSEP` 还执行似然推理, 计算 SEPPVB 类的平均强度过程并分析点过程残差.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

附录 D　排队系统的弱收敛和极限定理

经典的大数定律 (SLLN) 和中心极限定理 (CLT) 对于独立同分布序列 $\{X_i\}_{i\geqslant 1}$ 的第 n 个求和项 $\dfrac{S_n}{n}$ 和 $S_n - n\mu/\sqrt{\sigma^2 n}$ 的收敛性质进行了描述, 其中 $S_n = \sum_{i=1}^{n} X_i$, $E[X_i] = \mu$, $\mathrm{Var}(X_i) = \sigma^2$. 一个随机过程的极限取决于其收敛性, 且取决于整个序列的求和而不是 CLT 中的第 n 个部分和. 如果随机过程可以看作是随机函数, 那么随机过程的极限可以被转换成函数空间中概率测度的收敛性, 接下来我们对此详细介绍.

首先, 定义 X 为一个概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 到一个可测空间 $(A, \mathcal{A})$ 的映射. 如果对于 $\forall B \in \mathcal{A}$, 满足 $X^{-1}B \in \mathcal{F}$, 则 X 称为一个*随机元素*. 另外, $\mathbb{P} = \mathbb{P}X^{-1}$ 是 X 的关于概率的概率测度 $(A, \mathcal{A})$. 定义 A 为度量空间, 而 $\mathcal{A}$ 为其 Borel σ 域 (由开集生成), 定义 $\{X_n\}_{n\geqslant 1}$ 为取值在 A 的随机元素序列. 对于 $\{X_n\}_{n\geqslant 1}$, 若满足对于任意的 A 上的有界、连续实值函数 f, $Ef(X_n) \to Ef(X)$ 成立, 则 $\{X_n\}_{n\geqslant 1}$ 弱收敛于 X, 记为 $X_n \Rightarrow X$.

接下来, 我们考虑了随机过程的样本路径的函数空间, 重点讨论 càdlàg. 一个在 $[0,1]$ 右连续且有左极限的函数称为 càdlàg. 记 $D[0,1]$ 为从 $[0,1]$ 映射到 $\mathbb{R}$ 的所有 càdlàg 函数的集合. 定义 Λ 为从 $[0,1]$ 映射到 $[0,1]$ 的严格递增函数类. 对于 $\lambda \in \Lambda$, 定义

$$||\lambda|| = \sup_{s\neq t} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right|. \tag{D.1}$$

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

则 $D[0,1]$ 上的 Skorokhod 度量为

$$d(x_1,x_2) \equiv \inf_{\lambda\in\Lambda}\{||x_1\circ\lambda - x_2|| \vee ||\lambda - e||\},$$

其中 $a\vee b = \max\{a,b\}$, $\circ$ 表示函数的复合. 定义 $\mathcal{B}$ 是基于 Skorokhod 度量 $D[0,1]$ 的所有 Borel σ 域的集合. 则一个 càdlàg 路径的随机过程 X_t, $0\leqslant t\leqslant 1$ 即为在 $D[0,1]$ 取值的随机元素. 因此, 我们可以定义 Skorokhod 度量 $D[0,1]$ 上的取值在 $[0,1]$ 的 càdlàg 随机过程的弱收敛性. $D([0,1],\mathbb{R})$ 空间可以进行两方面的一般化. 首先, 函数的取值范围可以从 $\mathbb{R}$ 拓展到 $\mathbb{R}^k$; 其次, 函数的域可以从 $[0,1]$ 拓展到 $[0,\infty)$. 更详细的讨论参见 Whitt (2002).

在本附录中, 我们描述了关于随机过程极限的一些基本结果, 从而为 5.5.1, 5.5.2 和 C.3 节提供了一些基本的背景资料.

D.1 Donsker 定理及其扩展

Donsker 定理是随机过程极限的一个基本定理, 也被称为函数中心极限定理 (FCLT). 该定理讨论了独立同分布变量的部分和到函数空间 D 上的布朗运动的收敛性质. 由于极限只取决于 X_i 的前两阶矩, 该定理也被称为 *Donsker 不变性原理*.

定理 8 定义 $\{X_n\}_{n\geqslant 1}$ 为一个独立同分布的随机变量序列, 满足 $\mathbb{E}[X_1]=\mu$ 和 $\mathrm{Var}(X_1)=\sigma^2$, 定义第 n 个归一化的部分和 S_n 为

$$S_n(t) \equiv n^{-1/2}\left(\sum_{i=1}^{\lfloor nt\rfloor} X_i - \mu nt\right). \tag{D.2}$$

则当 $n\to\infty$,

$$S \Longrightarrow \sigma B, \quad \text{在 } D[0,\infty) \text{ 中}, \tag{D.3}$$

此处使用 Skorokhod 度量, 其中 B 是一个标准布朗运动, 而 $\lfloor nt\rfloor$ 表示小于或等于 nt 最大整数.

Donsker 定理可以进行多方面拓展, 包括基于 Cramér-Wold 策略的多维拓展, 以及将独立同分布条件放宽为弱依赖条件. 一个典型的弱依赖条件是一个双侧平稳序列的均匀混合条件. 另一个典型的条件是 FCLT 的均匀混合条件, 包括遍历性和鞅属性. 更多的细节参照 Bradley (2005), Billingsley

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

(1968), Jacod 和 Shiryaev (1987) 以及 Whitt (2002). 对于一个双边平稳序列 $\{X_n\}_{-\infty<n<\infty}$, 当满足 $\sum_{n=1}^{\infty}\rho_n<\infty$ 时, 它即为均匀混合条件, 其中

$$\rho_n\equiv\sup\big\{|E[XY]|:X\in\mathcal{F}_k,E(X)=0,E(X^2)\leqslant 1,Y\in\mathcal{G}_{n+k},E(Y)=0,E(Y^2)\leqslant 1\big\},$$

其中 $\mathcal{F}_n=\sigma(X_k,k\leqslant n)$ (由 $\{X_k,k\leqslant n\}$ 生成的 σ 域), $\mathcal{G}_n=\sigma(X_k,k\geqslant n)$ (由 $\{X_k,k\geqslant n\}$ 生成的 σ 域).

定理 9 如果 $\{X_n\}_{-\infty<n<\infty}$ 是一个满足均匀的混合条件的双边平稳序列, 且

$$E(X_n)=m<\infty,\quad E(X_n^2)<\infty, \tag{D.4}$$

则 $\sigma^2=\mathrm{Var}X_n+2\sum_{k=1}^{\infty}\mathrm{Cov}(X_1,X_{1+k})<\infty$ 且 (D.3) 成立. 此处 σ^2 称为渐近方差.

另一个拓展是鞅 FLCT, 其在随机过程的统计推断中有着重要作用, 参见 Durrett (2010).

定理 10 假设 $\{X_{n,m},\mathcal{F}_{n,m},1\leqslant m\leqslant n\}$ 是一个鞅差分序列. 定义 $S_{n,k}=\sum_{m=1}^{k}X_{n,m}$, $V_{n,k}=\sum_{m=1}^{k}E(X_{n,m}^2|\mathcal{F}_{n,m-1})$. 假设当 $n\to\infty$ 时, (i) $V_{n,\lfloor nt\rfloor}\to t$ 对于所有的 $0\leqslant t\leqslant 1$ 依概率成立, 且 (ii) $\sum_{m=1}^{k}E(X_{n,m}^2I_{\{|X_{n,m}|>\delta\}}|\mathcal{F}_{n,m-1})\to 0$ 对于所有的 $\delta>0$ 都依概率成立. 定义 $S_n(t)=S_{n,\lfloor nt\rfloor}$, 则 (D.3) 式成立, 且 $\sigma=1$.

定理 10 中的条件 (ii) 通常被称为条件 *Lindeberg*. 当 $X_{n,1},\ldots,X_{n,m}$ 为独立零均值变量, 满足 $\sum_{m=1}^{k}E(X_{n,m}^2)=1$, 它可以简化为经典的 Lindeberg 条件如下:

$$\lim_{n\to\infty}\sum_{m=1}^{k}E(X_{n,m}^2I_{\{|X_{n,m}|>\delta\}})=0,\ \text{对于所有}\ \delta>0, \tag{D.5}$$

该条件也是 $S_{n,m}$ 的极限分布为正态分布的充要条件.

D.2 排队系统和极限定理

排队论利用随机方法来分析拥挤和排队延误等问题, 研究各种物理量包括排队等候时间、到达过程、服务过程、服务器的数量、系统的数量、“客户” 的数量 (客户可能是人、数据包、汽车等).

CTMC 的 M/M/1 队列 M/M/1 队列是最简单的队列系统, 包含 (a) 一个服务器和一种任务; (b) 请求的到达服从一个强度为 λ 的泊松过程; (c) 服

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

务时间呈参数为 μ 的指数分布. 此处, 服务器根据先到先服务的策略, 从队列的前端为客户服务. 一旦服务完成, 客户就会离开队列, 而缓冲区的大小是无限的, 因此系统中客户的数量是没有限制的. 接下来, 定义 $C(t)$ 为系统中在时刻 t 等待的客户数量. 不难看出, 它是一个连续时间的 Markov 链, 生成元矩阵 Q 定义如下:

$$Q=\begin{pmatrix} -\lambda & \lambda & & & & \\ \mu & -(\lambda+\mu) & \lambda & & & \\ & \mu & -(\lambda+\mu) & \lambda & & \\ & & \mu & -(\lambda+\mu) & \lambda & \\ & & & \ddots & \ddots & \ddots \end{pmatrix}. \tag{D.6}$$

队列的流量限制 排队系统在一段时间的演变过程是一个比交易间隔更大的问题. 队列长度可以用简单的随机过程 (如布朗运动) 来描述. 从数学角度来说, 大流量限制是大数定律 (FSLLN) 和函数中心极限定理 (FCLT) 的变形形式, 具体可参见 Iglehart (1973a, b), Borovkov (1976), Harrison (1985) 和 Whitt (2002).

考虑如下的 M/M/1 排队系统, 其到达率为 λ、服务速度为 μ 且在时刻 t 时系统中等待的客户数量 $C(t)$ 为连续时间 Markov 链. 定义 $\rho=\lambda/\mu$. 则当 $\rho\uparrow 1$ 时, 该系统为繁忙情况. 此时, 考虑一个比例化的队列 $C_\rho(\cdot)$ 如下:

$$C_\rho(t)=(1-\rho)C\left(\frac{t}{(1-\rho)^2}\right).$$

在一个适当定义的拓扑空间中, 该队列弱收敛于一个反射布朗运动, 即,

$$C_\rho \Longrightarrow B^R \quad \text{在 } D[0,T] \text{ 中}, \tag{D.7}$$

其包含 Skorokhod 度量, B^R 称为反射布朗运动, 漂移项为 $\lambda-\mu$ 并且方差为 $\lambda+\mu$; 这一繁忙的流量限制可以一般化到包含多个服务器的更复杂的队列系统, 以及用于更一般的到达过程和服务时间的分布. 包含违约的队列是 LOB 分析中的一个非常有趣的问题. 在违约队列中假设客户并没有耐心, 且每个人都以某种概率离开队列. Ward 和 Glynn (2003, 2005) 指出, 在正则性条件下, 违约队列的扩散极限是 Ornstein-Uhlenbeck 过程.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

参考文献

F. Abergel and A. Jedidi. Long-time behavior of a Hawkes process-based limit order book. *SIAM J. Finan. Math.*, 6(1): 1026–1043, 2015.

Y. Achdou, F. J. Buera, J.-M. Lasry, P.-L. Lions, and B. Moll. Partial differential equation models in macroeconomics. *Phil. Trans. Roy. Soc. A*, 372 (2028): 20130397, 2014.

H.-J. Ahn, K.-H. Bae, and K. Chan. Limit orders, depth, and volatility: Evidence from the stock exchange of Hong Kong. *J. Finance*, 56(2): 767–788, 2001.

S. Ahuja, G. Papanicolaou, D. Ren, and F. Yang. Limit order trading with a mean-reverting reference price. Working paper, Department of Mathematics, Stanford University, 2016.

Y. Aït-Sahalia. Disentangling diffusion from jumps. *J. Finan. Econom.*, 74(3): 487–528, 2004.

Y. Aït-Sahalia and J. Jacod. Testing whether jumps have finite or infinite activity. *Ann. Stat.*, 39(3): 1689–1719, 2011.

Y. Aït-Sahalia and J. Jacod. Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. *J. Economic Literature*, 50(4): 1007–1050, 2012.

Y. Aït-Sahalia and J. Jacod. *High-Frequency Financial Econometrics.* Princeton University Press, Princeton, NJ, 2015.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

Y. Aït-Sahalia, P. A. Mykland, and L. Zhang. How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*, 18: 351–416, 2005.

Y. Aït-Sahalia, J. Fan, and D. Xiu. High-frequency covariance estimates with noisy and asynchronous financial data. *J. Amer. Statist. Assoc.*, 105(492): 1504–1517, 2010.

J. Akahori, N.-L. Liu, M. E. Mancino, and Y. Yasuda. The Fourier estimation method with positive semi-definite estimators. *ArXiv e-prints*, 2014.

M. Akian, J. L. Menaldi, and A. Sulem. On an investment-consumption model with transaction costs. *SIAM J. Control Optimiz.*, 34(1): 329–364, 1996.

S. S. Alexander. Price movements in speculative markets: Trends or random walks. *Industrial Management Review*, 2(2): 7–26, 1961.

A. Alfonsi and P. Blanc. Dynamic optimal execution in a mixed-marketimpact hawkes price model. *ArXiv e-prints*, 2014.

A. Alfonsi, A. Fruth, and A. Schied. Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2): 143–157, 2010.

P. H. Algoet and T. M. Cover. Asymptotic optimality and asymptotic equipartition properties of log-optimum investment. *Ann. Probab.*, 16(2): 876–898, 1988.

R. Almgren and N. Chriss. Value under liquidation. *Risk*, 12(12): 61–63, 1999.

R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *J. Risk*, 3: 5–40, 2001.

R. Almgren, C. Thum, E. Hauptmann, and H. Li. Direct estimation of equity market impact. *Risk*, 3(2): 5–39, 2005.

T. G. Andersen and T. Bollerslev. Intraday periodicity and volatility persistence in financial markets. *J. Empirical Finance*, 4: 115–158, 1997.

C. S. Asness, T. J. Moskowitz, and L. H. Pedersen. Value and momentum everywhere. *J. Finance*, 68(3): 929–985, 2013.

K. J. Åström. Theory and applications of adaptive control survey. *Automatica*, 19(5): 471–486, 1983.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

M. Avellaneda and S. Stoikov. High-frequency trading in a limit order book. *Quantitative Finance*, 8(3): 217–224, 2008.

K. Azuma. Weighted sums of certain dependent random variables. *Tôhoku Math. J.* (2), 19: 357–367, 1967.

L. Bachelier. *Théorie de la Spéculation*. Gauthier-Villars, 1900.

L. Bachelier. *La Spéculation et le Calcul des Probabilités*. Gauthier-Villars, 1938.

K. Back and S. Baruch. Strategic liquidity provision in limit order markets. *Econometrica*, 81(1): 363–392, 2013.

Z. Bai. Methodologies in spectral analysis of large dimensional random matrices, a review. *Statistica Sinica*, pages 611–662, 1999.

Z. Bai and J. W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.*, 26(1): 316–345, 1998.

C. A. Ball and W. N. Torous. A simplified jump process for common stock returns. *J. Financial and Quantitative Analysis*, 18(1): 53–65, 1983.

F. M. Bandi and J. R. Russell. Separating microstructure noise from volatility. *J. Finan. Econom.*, 79(3): 655–692, 2006.

N. Barberis, M. Huang, and T. Santos. Prospect theory and asset prices. *Quarterly J. Economics*, 116: 1–53, 2001.

M. Barigozzi and M. Hallin. Generalized dynamic factor models and volatilities: Recovering the market volatility shocks. *The Econometrics Journal*, 19(1): C33–C60, 2016.

O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, 76: 1481–1536, 2008.

O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econometr.*, 162(2): 149–169, 2011.

A. G. Barto, R. S. Sutton, and C. W. Anderson. Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Trans. Systems, Man and Cybernetics*, SMC-13(5): 834–846, 1983.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

L. Bauwens and N. Hautsch. Modelling financial high frequency data using point processes. In *Handbook of Financial Time Series*, pages 953–979. Springer, Berlin Heidelberg, 2009.

D. S. Bayard. A forward method for optimal stochastic nonlinear and adaptive control. *IEEE Trans. Automat. Contr.*, 36(9): 1046–1053, 1991.

D. Becherer, T. Bilarev, and P. Frentrup. Multiplicative limit order markets with transient impact and zero spread. *ArXiv e-prints*, 2015.

R. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1st edition, 1957.

S. Benartzi and R. H. Thaler. Myopic loss aversion and the equity premium puzzle. *Quarterly J. Economics*, 110(1): 73–92, 1995.

V. E. Benes, L. A. Shepp, and H. S. Witsenhausen. Some solvable stochastic control problems. *Stochastics*, 4(1): 39–83, 1980.

A. Bensoussan and J.-L. Lions. *Impulse Control and Quasi-Variational Inequalities*. Dunod, 1982.

B. S. Bernanke and A. S. Blinder. The federal funds rate and the channels of monetary transmission. *The American Economic Review*, pages 901–921, 1992.

B. S. Bernanke, J. Boivin, and P. Eliasz. Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach. *The Quarterly journal of economics*, 120(1): 387–422, 2005.

D. Bertsimas and A. W. Lo. Optimal control of execution costs. *J. Financial Markets*, 1(1): 1–50, 1998.

D. Bertsimas, A. W. Lo, and P. Hummel. Optimal control of execution costs for portfolios. *Computing in Science & Engineering*, 1(6): 40–53, 1999.

D. Bertsimas, V. Gupta, and I. C. Paschalidis. Inverse optimization: A new perspective on the Black-Litterman model. *Operations Research*, 60(6): 1389–1403, 2012.

J. Besag. Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B*, 36: 192–236, 1974. With discussion by D. R. Cox, A. G. Hawkes, P. Clifford, P. Whittle, K. Ord, R. Mead, J. M. Hammersley, and M. S. Bartlett and with a reply by the author.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

J. Besag. On the statistical analysis of dirty pictures. *J. Roy. Statist. Soc. Ser. B*, 48(3): 259–302, 1986.

M. J. Best and R. R. Grauer. On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies*, 4(2): 315–342, 1991.

B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the Paris Bourse. *J. Finance*, 50(5): 1655–1689, 1995.

P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 1968.

F. Black and R. B. Litterman. Asset allocation: Combining investor views with market equilibrium. *J. Fixed Income*, 1(2): 7–18, 1991.

J. Blanchet and X. Chen. Continuous-time Modeling of Bid-Ask Spread and Price Dynamics in Limit Order Books. *ArXiv e-prints*, 2013.

D. Bloembergen, K. Tuyls, D. Hennes, and M. Kaisers. Evolutionary dynamics of multi-agent learning: A survey. *J. Artificial Intelligence Research*, pages 659–697, 2015.

J. R. Blum. Multidimensional stochastic approximation methods. *Ann. Math. Stat.*, 25: 737–744, 1954.

O. Bondarenko. Statistical arbitrage and securities prices. *Review of Financial Studies*, 16(3): 875–919, 2003.

A. A. Borovkov. *Stochastic Processes in Queueing Theory*. Springer, 1976.

J.-P. Bouchaud, M. Mézard, and M. Potters. Statistical properties of stock order books: Empirical results and models. *Quantitative Finance*, 2(4): 251–256, 2002.

J.-P. Bouchaud, Y. Gefen, M. Potters, and M. Wyart. Fluctuations and response in financial markets: The subtle nature of 'random' price changes. *Quantitative Finance*, 4(2): 176–190, 2004.

G. E. P. Box. Science and statistics. *J. Amer. Statist. Assoc.*, 71(356): 791–799, 1976.

G. E. P. Box and N. R. Draper. *Empirical Model Building and Response Surfaces*. Wiley, Hoboken, NJ, 1987.

G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1970.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

G. E. P. Box and D. A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Assoc.*, 65(332): 1509–1526, 1970.

R. C. Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2: 107–144, 2005.

L. Breiman. Optimal gambling systems for favorable games. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 65–78, Berkeley, 1961. University of California Press.

L. Breiman. Statistical modeling: The two cultures. *Statist. Sci.*, 16(3): 199–231, 2001. With invited comments and a rejoinder by the author.

M. J. Brennan. The role of learning in dynamic portfolio decisions. *European Finance Review*, 1(3): 295–306, 1998.

M. Britten-Jones. The sampling error in estimates of mean-variance efficient portfolio weights. *J. Finance*, 54(2): 655–671, 1999.

M. Broadie. Computing efficient frontiers using estimated parameters. *Ann. Oper. Res.*, 45(1): 21–58, 1993.

W. Brock, J. Lakonishok, and B. LeBaron. Simple technical trading rules and the stochastic properties of stock returns. *J. Finance*, 47(5): 1731–1764, 1992.

P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting*. Springer, New York, 2nd edition, 2002.

T. Cai, A. Munk, and J. Schmidt-Hieber. Sharp minimax estimation of the variance of Brownian motion corrupted with Gaussian noise. *Statistica Sinica*, 20(3): 1011–1024, 2010.

J. Y. Campbell, A. W. Lo, and A. C. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ, 1997.

N. Canner, N. G. Mankiw, and D. N. Weil. An asset allocation puzzle. *The American Economic Review*, 87(1): 181–191, 1997.

R. Carmona and F. Delarue. Probabilistic analysis of mean-field games. *SIAM J. Control Optimiz.*, 51(4): 2705–2734, 2013.

R. Carmona and D. Lacker. A probabilistic weak formulation of mean field games and applications. *Ann. Appl. Probab.*, 25(3): 1189–1231, 2015.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

P. Carr and L. Wu. Time-changed Lévy processes and option pricing. *J. Finan. Econom.*, 71(1): 113–141, 2004.

Á. Cartea, S. Jaimungal, and J. Ricci. Buy low, sell high: A high frequency trading perspective. *SIAM J. Finan. Math.*, 5(1): 415–444, 2014.

G. Chamberlain and M. Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51: 1281–1304, 1983.

H. P. Chan and T. L. Lai. A general theory of particle filters in hidden Markov models and some applications. *Ann. Stat.*, 41(6): 2877–2904, 2013.

H. P. Chan and T. L. Lai. MCMC with sequential state substitutions: Theory and applications. Working paper, Department of Statistics, Stanford University, 2016.

F. Chen and P. Hall. Inference for a nonstationary self-exciting point process with an application in ultra-high frequency financial data modeling. *J. Appl. Probab.*, 50(4): 1006–1024, 2013.

H.-S. Chen, D. G. Simpson, and Z. Ying. Infill asymptotics for a stochastic process model with measurement error. *Statistica Sinica*, 10(1): 141–156, 2000.

X. Chen and M. Dai. Characterization of optimal strategy for multiasset investment and consumption with transaction costs. *SIAM J. Finan. Math.*, 4(1): 857–883, 2013.

V. K. Chopra, C. R. Hensel, and A. L. Turner. Massaging mean-variance inputs: Returns from alternative global investment strategies in the 1980s. *Management Science*, 39(7): 845–855, 1993.

P. K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1): 135–155, 1973.

R. Cont. Statistical modeling of high-frequency financial data. *Signal Processing Magazine, IEEE*, 28(5): 16–25, 2011.

R. Cont and A. de Larrard. Order book dynamics in liquid markets: Limit theorems and diffusion approximations, 2012.

R. Cont and A. de Larrard. Price dynamics in a markovian limit order market. *SIAM J. Finan. Math.*, 4(1): 1–25, 2013.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

R. Cont and A. Kukanov. Optimal order placement in limit order markets, 2016. forthcoming.

T. M. Cover. Universal data compression and portfolio selection. In *Proceedings of the 37th Annual Symposium on Foundations of Computer Science*, pages 534–538. IEEE Computer Society, 1996.

T. M. Cover and E. Ordentlich. Universal portfolios with side information. *IEEE Trans. Information Theory*, 42(2): 348–363, 1996.

A. Cowles. Can stock market forecasters forecast? *Econometrica*, 1(3): 309–324, 1933.

A. Cowles. Stock market forecasting. *Econometrica*, 12(3/4): 206–214, 1944.

D. R. Cox. Role of models in statistical analysis. *Statist. Sci.*, 5(2): 169–174, 1990.

D. R. Cox and N. Reid. A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91(3): 729–737, 2004.

J. C. Cox and C.-F. Huang. Optimal consumption and portfolio policies when asset prices follow a diffusion process. *J. Economic Theory*, 49(1): 33–83, 1989.

J. Cvitanić, A. Lazrak, L. Martellini, and F. Zapatero. Dynamic portfolio choice with parameter uncertainty and the economic value of analysts' recommendations. *Review of Financial Studies*, 19(4): 1113–1156, 2006.

M. Dai and Y. Zhong. Penalty methods for continuous-time portfolio selection with proportional transaction costs. *J. Computational Finance*, 13(3): 1–31, 2010.

D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes Vol. I. Springer-Verlag*, New York, 2nd edition, 2003.

A. Dassios and H. Zhao. Exact simulation of Hawkes process with exponentially decaying intensity. *Electron. Commun. Probab.*, 18(62): 1–13, 2013.

M. A. Davis, X. Guo, and G. L. Wu. Impulse controls for multi-dimensional jump diffusions. *SIAM J. Control Optimiz.*, 48(8): 5276–5293, 2010.

M. H. Davis and A. R. Norman. Portfolio selection with transaction costs. *Mathematics of Operations Research*, 15(4): 676–713, 1990.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

M. H. A. Davis, V. G. Panas, and T. Zariphopoulou. European option pricing with transaction costs. *SIAM J. Control Optim.*, 31(2): 470–493, 1993.

V. H. de la Peña, T. L. Lai, and Q.-M. Shao. *Self-Normalized Processes: Limit Theory and Statistical Applications*. Probability and Its Applications. Springer-Verlag, Berlin Heidelberg, 2009.

G. Demange, D. Gale, and M. Sotomayor. Multi-item auctions. *J. Political Economy*, 94(4): 863–872, 1986.

M. A. Dempster, I. V. Evstigneev, and K. R. Schenk-Hoppé. Volatility-induced financial growth. *Quantitative Finance*, 7(2): 151–160, 2007.

A. S. Deng, T. L. Lai, and H. Xing. *Data Science and Decision Analytics in Finance, Information Technology, and Manufacturing*. Wiley, Hoboken NJ, 2017.

J. B. Detemple. Asset pricing in a production economy with incomplete information. *J. Finance*, 41(2): 383–391, 1986.

D. diBartolomeo. Smarter rebalancing: Using single period optimization in a multi-period world. Technical report, Northfield Research, 2012.

M. U. Dothan and D. Feldman. Equilibrium interest rates and multiperiod bonds in a partially observable economy. *J. Finance*, 41(2): 369–382, 1986.

D. N. Dreman. *Contrarian Investment Strategy: The Psychology of Stock Market Success*. Random House, Inc., 1979.

D. N. Dreman. *Contrarian Investment Strategies: The Psychological Edge*. Simon and Schuster Co., 2012.

U. Drepper. What every programmer should know about memory. *Red Hat, Inc.*, 2007.

D. Duffie. *Dark Markets: Asset Pricing and Information Transmission in Over-the-Counter Markets*. Princeton University Press, 2012.

R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, New York, NY, USA, 4th edition, 2010.

P. Dütting, T. Roughgarden, and I. Talgam-Cohen. Modularity and greed in double auctions. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, pages 241–258, 2014.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

E. Eberlein and S. Raible. Term structure models driven by general Lévy processes. *Mathematical Finance*, 9(1): 31–53, 1999.

E. Eberlein, U. Keller, and K. Prause. New insights into smile, mispricing, and value at risk: The hyperbolic model. *J. Business*, 71(3): 371–405, 1998.

B. Edelman, M. Ostrovsky, M. Schwarz, and others. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review*, 97(1): 242–259, 2007.

N. El Karoui. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Stat.*, pages 2757–2790, 2008.

R. J. Elliott. *Stochastic Calculus and Applications*, volume 18. Springer-Verlag, New York, 1982.

R. F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 45: 987–1007, 1982.

R. F. Engle. The econometrics of ultra-high frequency data. *Econometrica*, 68(1): 1–22, 2000.

R. F. Engle and G. M. Gallo. A multiple indicators model for volatility using intra-daily data. *J. Econometrics*, 131(12): 3–27, 2006.

R. F. Engle and J. R. Russell. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, 66(5): 1127–1163, 1998.

R. F. Engle, J. Mezrich, and L. You. Optimal asset allocation. In *Smith-Barney Market Commentary*, 1998. January 28th issue.

R. F. Engle, R. Ferstenberg, and J. Russell. Measuring and modeling execution cost and risk. *J. Portfolio Management*, 38(2): 14, 2012.

T. W. Epps. Comovements in stock prices in the very short run. *J. Amer. Statist. Assoc.*, 74(366): 291–298, 1979.

E. F. Fama. Mandelbrot and the stable Paretian hypothesis. *J. Business*, 36: 420–429, 1963.

E. F. Fama. Efficient capital markets: A review of theory and empirical work. *J. Finance*, 25(2): 383–417, 1970.

E. F. Fama and M. E. Blume. Filter rules and stock-market trading. *J. Business*, 39(1): 226–241, 1966.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *J. Finan. Econom.*, 33(1): 3–56, 1993.

E. F. Fama and K. R. French. A five-factor asset pricing model. *J. Finan. Econom.*, 116(1): 1–22, 2015.

E. F. Fama and R. Roll. Parameter estimates for symmetric stable distributions. *J. Amer. Statist. Soc.*, 66(334): 331–338, 1971.

J. Fan and Y. Wang. Spot volatility estimation for high-frequency data. *Statistics and its Interface*, 1(2): 279–288, 2008.

J. Fan, Y. Li, and K. Yu. Vast volatility matrix estimation using highfrequency data for portfolio selection. *J. Amer. Statist. Assoc.*, 107(497): 412–428, 2012.

J. Fan, Y. Liao, and M. Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *J. Roy. Statist. Soc. Ser. B*, 75(4): 603–680, 2013.

J. D. Farmer, P. Patelli, and I. I. Zovko. The predictive power of zero intelligence in financial markets. *Proc. Nat. Acad. Sci. USA*, 102(6): 2254–2259, 2005.

R. Fernholz and B. Shay. Stochastic portfolio theory and stock market equilibrium. *J. Finance*, 37(2): 615–624, 1982.

L. Fisher. Some new stock-market indexes. *J. Business*, 39(1): 191–225, 1966.

T. Fletcher, Z. Hussain, and J. Shawe-Taylor. Multiple kernel learning on the limit order book. In J. S.-T. Tom Diethe, Nello Cristianini, editor, *Journal of Machine Learning Research Workshop and Conference Proceedings*, volume 11, pages 167–74, 2010.

P. Fodra and M. Labadie. High-frequency market-making with inventory constraints and directional bets. *ArXiv e-prints*, 2012.

A. Földes and L. Rejtö. Strong uniform consistency for nonparametric survival curve estimators from randomly censored data. *Ann. Statist.*, 9(1): 122–129, 1981.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: One-sided estimation and forecasting. *J. Amer. Statist. Assoc.*, 100 (471): 830–840, 2005.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

P. A. Forsyth, J. S. Kennedy, S. Tse, and H. Windcliff. Optimal trade execution: A mean quadratic variation approach. *J. Econom. Dynam. Contr.*, 36(12): 1971–1991, 2012.

G. M. Frankfurter, H. E. Phillips, and J. P. Seagle. Performance of the Sharpe portfolio selection model: A comparison. *J. Financial and Quantitative Analysis*, 11(02): 195–204, 1976.

R. S. Freeman. *Introduction to Financial Technology*. Complete Technology Guides for Financial Services. Elsevier, 2006.

D. Fudenberg, M. Mobius, and A. Szeidl. Existence of equilibrium in large double auctions. *J. Economic Theory*, 133(1): 550–567, 2007.

K. Ganchev, Y. Nevmyvaka, M. Kearns, and J. W. Vaughan. Censored exploration and the dark pool problem. *Commun. ACM*, 53(5): 99–107, 2010.

N. Gârleanu and L. H. Pedersen. Dynamic trading with predictable returns and transaction costs. *J. Finance*, 68(6): 2309–2340, 2013.

H. Gatfaoui. Deviation from normality and Sharpe ratio behavior: A brief simulation study. *Investment Management and Financial Innovations*, 7 (4): 95–107, 2010.

J. Gatheral. No-dynamic-arbitrage and market impact. *Quantitative Finance*, 10(7): 749–759, 2010.

J. Gatheral and R. C. Oomen. Zero-intelligence realized variance estimation. *Finance and Stochastics*, 14(2): 249–283, 2010.

J. Gatheral and A. Schied. Optimal trade execution under geometric brownian motion in the almgren and chriss framework. *Int. J. Theoret. Appl. Finance*, 14(03): 353–368, 2011.

J. Gatheral and A. Schied. Dynamical models of market impact and algorithms for order execution. In J. A. L. Jean-Pierre Fouque, editor, *Handbook on Systemic Risk*, pages 579–599. Cambridge University Press, New York, 2013.

R. Ge and C. Huang. A continous approach to nonlinear integer programming. *Applied Mathematics and Computation*, 34: 39–60, 1989.

I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Dover, 2000.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

R. Gencay. Optimization of technical trading strategies and the profitability in security markets. *Economics Letters*, 59(2): 249–254, 1998.

G. Gennotte. Optimal portfolio choice under incomplete information. *J. Finance*, 41(3): 733–746, 1986.

G. Gennotte and A. Jung. Investment strategies under transaction costs: The finite horizon case. *Management Science*, 40(3): 385–404, 1994.

K. Giesecke and B. Kim. Estimating tranche spreads by loss process simulation. In *Proceedings of the 39th Conference on Winter Simulation*, pages 967–975. IEEE Press, 2007.

K. Giesecke, B. Kim, and S. Zhu. Monte Carlo algorithms for default timing problems. *Management Science*, 57(12): 2115–2129, 2011.

L. R. Glosten and P. R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *J. Finan. Econom.*, 14(1): 71–100, 1985.

M. Gorham and N. R. Singh. *Electronic Exchanges: The Global Transformation from Pits to Bits*. Elsevier, 2009.

M. D. Gould, M. A. Porter, S. Williams, M. McDonald, D. J. Fenn, and S. D. Howison. Limit order books. *Quantitative Finance*, 13(11): 1709–1742, 2013.

B. Graham and D. L. Dodd. *Security Analysis*. McGraw-Hill Professional, New York, 1934.

R. C. Grinold. Implementation efficiency. *Finan. Analysts J.*, 61(5): 52–64, 2005.

R. C. Grinold and R. N. Kahn. *Active Portfolio Management*. McGraw-Hill, New York, 2nd edition, 2000.

M. G. Gu and T. L. Lai. Functional laws of the iterated logarithm for the product-limit estimator of a distribution function under random censorship or truncation. *Ann. Probab.*, 18(1): 160–189, 1990.

O. Guéant, J.-M. Lasry, and P.-L. Lions. Mean field games and applications. In *Paris-Princeton Lectures on Mathematical Finance 2010*, pages 205–266. Springer, 2011.

O. Guéant, C.-A. Lehalle, and J. Fernandez-Tapia. Optimal portfolio liquidation with limit orders. *SIAM J. Finan. Math.*, 3(1): 740–764, 2012.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

O. Guéant, C.-A. Lehalle, and J. Fernandez-Tapia. Dealing with the inventory risk: A solution to the market making problem. *Math. Finan. Econom.*, 7(4): 477–507, 2013.

F. Guilbaud and H. Pham. Optimal high-frequency trading with limit and market orders. *Quantitative Finance*, 13(1): 79–94, 2013.

D. M. Guillaume, M. M. Dacorogna, R. R. Davé, U. A. Müller, R. B. Olsen, and O. V. Pictet. From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics*, 1(2): 95–129, 1997.

X. Guo and M. Zervos. Optimal execution with multiplicative price impact. *SIAM J. Finan. Math.*, 6(1): 281–306, 2015.

X. Guo, A. de Larrard, and Z. Ruan. Optimal placement in a limit order book: An analytical approach. Preprint, 2013.

X. Guo, Z. Ruan, and L. Zhu. Dynamics of order positions and related queues in a limit order book. Preprint, 2015.

J. Han, T. L. Lai, and V. Spivakovsky. Approximate policy optimization and adaptive control in regression models. *Computational Economics*, 27(4): 433–452, 2006.

P. R. Hansen. A test for superior predictive ability. *J. Business & Economic Statistics*, 23(4): 365–380, 2005.

P. R. Hansen and A. Lunde. Realized variance and market microstructure noise. *J. Business & Economic Statistics*, 24(2): 127–161, 2006.

P. R. Hansen, J. Large, and A. Lunde. Moving average-based estimators of integrated variance. *Econometric Reviews*, 27(1–3): 79–111, 2008.

P. R. Hansen, Z. Huang, and H. H. Shek. Realized GARCH: A joint model for returns and realized measures of volatility. *J. Appl. Economet.*, 27(6): 877–906, 2012.

J. M. Harrison. *Brownian Motion and Stochastic Flow Systems*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1985.

D. Harte. PtProcess: An R package for modelling marked point processes indexed by time. *J. Statist. Software*, 35(1): 1–32, 2010.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

J. Hasbrouck. Measuring the information content of stock trades. *J. Finance*, 46(1): 179–207, 1991.

J. Hasbrouck. Modelling market microstructure time series. In G. S. Maddala and C. R. Rao, editors, *Handbook of Statistics*, 14: 647–692. North-Holland, 1996.

J. Hasbrouck. Trading fast and slow: Security market events in real time. Technical report, NYU Working Paper No. FIN-99-012, 1999.

J. Hasbrouck and G. Sofianos. The trades of market makers: An empirical analysis of NYSE specialists. *J. Finance*, 48(5): 1565–1593, 1993.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, New York, 2nd edition, 2009.

N. Hautsch. *Econometrics of Financial High-Frequency Data*. Springer-Verlag, Berlin Heidelberg, 2012.

A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1): 83–90, 1971.

T. Hayashi and N. Yoshida. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2): 359–379, 2005.

T. Hayashi and N. Yoshida. Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes. *Ann. Inst. Statist. Math.*, 60(2): 367–406, 2008.

R. Haynes and J. S. Roberts. Automated trading in futures markets. Technical report, White Papers of the U.S. Commodity Futures Trading Commission, 2015.

D. Helbing and A. Kirman. Rethinking economics using complexity theory. *real-world economics review*, no. 64: 23–51, 2013.

T. Hendershott and M. S. Seasholes. Market maker inventories and stock prices. *Amer. Econom. Review*, 97(2): 210–214, 2007.

R. D. Henriksson and R. C. Merton. On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *J. Business*, 54(4): 513–533, 1981.

E. Hillebrand. Neglecting parameter changes in GARCH models. *J. Economet.*, 129(1): 121–138, 2005.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

T. Ho and H. R. Stoll. Optimal dealer pricing under transactions and return uncertainty. *J. Finan. Econom.*, 9(1): 47–73, 1981.

M. Hoffmann, A. Munk, and J. Schmidt-Hieber. Adaptive wavelet estimation of the diffusion coefficient under additive error measurements. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48: 1186–1216. Institut Henri Poincaré, 2012.

B. Hollifield, R. A. Miller, P. Sandås, and J. Slive. Estimating the gains from trade in limit-order markets. *J. Finance*, 61(6): 2753–2804, 2006.

C. H. Hommes. Modeling the stylized facts in finance through simple nonlinear adaptive systems. *Proc. Nat. Acad. Sci. USA*, 99(Supplement 3): 7221–7228, 2002.

U. Horst and D. Kreher. A weak law of large numbers for a limit order book model with fully state dependent order dynamics. *ArXiv e-prints*, 2015.

U. Horst and M. Paulsen. A law of large numbers for limit order books. *ArXiv e-prints*, 2015.

K. Huang, D. Simchi-Levi, and M. Song. Optimal market-making with risk aversion. *Operations Research*, 60(3): 541–565, 2012.

M. Huang, P. E. Caines, and R. P. Malhamé. Large-population cost-coupled LQG problems with non-uniform agents: Individual-mass behavior and decentralized ε-Nash equilibria. *IEEE Trans. Automat. Contr.*, 52(9): 1560–1571, 2007.

W. Huang, C.-A. Lehalle, and M. Rosenbaum. Simulating and analyzing order book data: The queue-reactive model. *J. Amer. Statist. Assoc.*, 110(509): 107–122, 2015.

G. Huberman and W. Stanzl. Price manipulation and quasi-arbitrage. *Econometrica*, 72(4): 1247–1275, 2004.

H. Hult and J. Kiessling. *Algorithmic trading with Markov chains*. PhD thesis, Stockholm University, Sweden, 2010.

D. L. Iglehart. Weak convergence in queueing theory. *Advan. Appl. Probab.*, 5(3): 570–594, 1973a.

D. L. Iglehart. Weak convergence of compound stochastic process, I. *Stochastic Processes and their Applications*, 1(1): 11–31, 1973b.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

C.-K. Ing and T. L. Lai. A stepwise regression method and consistent model selection for high-dimensional sparse linear models. *Statistica Sinica*, pages 1473–1513, 2011.

J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*, volume 288 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin Heidelberg, 1987.

J. Jacod, Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter. Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and Their Applications*, 119(7): 2249–2276, 2009.

N. Jegadeesh and S. Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *J. Finance*, 48(1): 65–91, 1993.

H. Jin and X. Y. Zhou. Behavioral portfolio selection in continuous time. *Mathematical Finance*, 18(3): 385–426, 2008.

J. D. Jobson and B. Korkie. Estimation for Markowitz efficient portfolios. *J. Amer. Statist. Assoc.*, 75(371): 544–554, 1980.

I. M. Johnstone and A. Y. Lu. On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.*, 104(486): 682–693, 2009.

P. Jorion. Bayes-Stein estimation for portfolio analysis. *J. Financial and Quantitative Analysis*, 21(3): 279–292, 1986.

F. Jovanovic and P. Le Gall. Does God practice a random walk? The 'financial physics' of a nineteenth-century forerunner, Jules Regnault. *European Journal for the History of Economic Thought*, 8(3): 323–362, 2001.

L. P. Kadanoff. More is the same; phase transitions and mean field theories. *J. Statist. Physics*, 137(5): 777–797, 2009.

L. P. Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement learning: A survey. *J. Artificial Intelligence Research*, 4: 237–285, 1996.

D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2): 263–291, 1979.

I. Karatzas. A class of singular stochastic control problems. *Advan. in Appl. Probab.*, 15(2): 225–254, 1983.

I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, 2nd edition, 1991.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

I. Karatzas, J. P. Lehoczky, and S. E. Shreve. Optimal portfolio and consumption decisions for a "small investor" on a finite horizon. *SIAM J. Control Optimiz.*, 25(6): 1557–1586, 1987.

M. Kearns and Y. Nevmyvaka. Machine learning for market microstructure and high frequency trading. In D. Easley, M. de Prado, and M.O'Hara, editors, *High Frequency Trading—New Realities for Traders*. RiskBooks, 2013.

J. Kelly. A new interpretation of information rate. *Bell System Technical Journal*, 35(4): 917–926, 1956.

M. Kendall. The analysis of economic time-series. Part I: Prices. *J. Roy. Statist. Soc. Ser. A*, 116(1): 11–25, 1953.

J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *Ann. Math. Stat.*, 23: 462–466, 1952.

A. A. Kirilenko, A. S. Kyle, M. Samadi, and T. Tuzun. The Flash Crash: The impact of high-frequency trading on an electronic market, 2015.

V. Krishna. *Auction Theory*. Academic Press, New York, 2009.

M. Kritzman, S. Myrgren, and S. Page. Portfolio rebalancing: A test of the Markowitz-Van Dijk heuristic. MIT Sloan Research Paper, 2007.

Y. Kroll, H. Levy, and H. M. Markowitz. Mean-variance versus direct utility maximization. *J. Finance*, 39(1): 47–61, 1984.

K. F. Kroner and J. Sultan. Time-varying distributions and dynamic hedging with foreign currency futures. *J. Financial and Quantitative Analysis*, 28(4): 535–551, 1993.

L. Kruk. Functional limit theorems for a simple auction. *Mathematics of Operations Research*, 28(4): 716–751, 2003.

P. R. Kumar and P. Varaiya. *Stochastic Systems: Estimation, Identification, and Adaptive Control*. Prentice Hall, Englewood Cliffs, NJ, 1986.

H. Kushner. Necessary conditions for continuous parameter stochastic optimization problems. *SIAM J. Control*, 10(3): 550–565, 1972.

A. S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6): 1315–1335, 1985.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

T. L. Lai. Sequential changepoint detection in quality control and dynamical systems. *J. Roy. Statist. Soc. Ser. B*, 57(4): 613–658, 1995. With discussion and a reply by the author.

T. L. Lai. Stochastic approximation. *Ann. Stat.*, 31(2): 391–406, 2003.

T. L. Lai and V. Bukkapatanam. Adaptive filtering, nonlinear state-space models, and applications to finance and econometrics. In Y. Zeng and S. Wu, editors, *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance, 1: 3–22. Springer, New York, 2013.

T. L. Lai and P. Gao. Approximate dynamic programming and multi-period mean-variance portfolio rebalancing with transactions costs. Working paper, Department of Statistics, Stanford University, 2016a.

T. L. Lai and P. Gao. Transaction costs and stochastic control in dynamic portfolio management. Working paper, Department of Statistics, Stanford University, 2016b.

T. L. Lai and J. Lim. Asymptotically efficient parameter estimation in hidden Markov spatio-temporal random fields. *Statist. Sinica*, 25(1): 403–421, 2015.

T. L. Lai and T. W. Lim. Option hedging theory under transaction costs. *J. Econom. Dynam. Control*, 33(12): 1945–1961, 2009.

T. L. Lai and K. W. Tsang. Post-selection multiple testing and a new approach to test-based variable selection. Working paper, Department of Statistics, Stanford University, 2016.

T. L. Lai and C. Z. Wei. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Stat.*, 10(1): 154–166, 1982.

T. L. Lai and C. Z. Wei. Extended least squares and their applications to adaptive control and prediction in linear systems. *IEEE Trans. Automat. Control*, 31(10): 898–906, 1986.

T. L. Lai and S. P.-S. Wong. Valuation of American options via basis functions. *IEEE Trans. Automat. Contr.*, 49(3): 374–385, 2004.

T. L. Lai and S. P.-S. Wong. Combining domain knowledge and statistical models in time series analysis. In *Time Series and Related Topics*, vol-

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

ume 52 of *IMS Lecture Notes Monogr. Ser.*, pages 193–209. Inst. Math. Statist., Beachwood, OH, 2006.

T. L. Lai and H. Xing. *Statistical Models and Methods for Financial Markets*. Springer Texts in Statistics. Springer, New York, 2008.

T. L. Lai and H. Xing. Stochastic change-point ARX-GARCH models and their applications to econometric time series. *Statist. Sinica*, 23(4): 1573–1594, 2013.

T. L. Lai and H. Xing. *Risk Analytics and Management in Finance and Insurance*. Chapman & Hall/CRC, 2017. forthcoming.

T. L. Lai and Z. Ying. Estimating a distribution function with truncated and censored data. *Ann. Stat.*, 19(1): 417–442, 1991a.

T. L. Lai and Z. Ying. Recursive identification and adaptive prediction in linear stochastic systems. *SIAM J. Control Optim.*, 29(5): 1061–1090, 1991b.

T. L. Lai, M.-C. Shih, and S. P.-S. Wong. Flexible modeling via a hybrid estimation scheme in generalized mixed models for longitudinal data. *Biometrics*, 62(1): 159–167, 317–318, 2006.

T. L. Lai, S. T. Gross, and D. B. Shen. Evaluating probability forecasts. *Ann. Statist.*, 39(5): 2356–2382, 2011a.

T. L. Lai, H. Xing, and Z. Chen. Mean-variance portfolio optimization when means and covariances are unknown. *Ann. Appl. Statist.*, 5(2A): 798–823, 2011b.

T. L. Lai, P. Gao, and L. Xu. Filtering approaches to dynamic portfolio selection in the presence of parameter uncertainty. Working paper, Department of Statistics, Stanford University, 2016a.

T. L. Lai, K. W. Tsang, and H. Yuan. Gradient boosting for high-dimensional nonlinear regression. Working paper, Department of Statistics, Stanford University, 2016b.

K. Lam and H. Yam. Cusum techniques for technical trading in financial markets. *Financial Engineering and the Japanese Markets*, 4(3): 257–274, 1997.

C. G. Lamoureux and W. D. Lastrapes. Persistence in variance, structural change, and the GARCH model. *J. Business & Economic Statistics*, 8(2):

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

225–234, 1990.

S. Laruelle, C.-A. Lehalle, and G. Pages. Optimal split of orders across liquidity pools: A stochastic algorithm approach. *SIAM J. Finan. Math.*, 2(1): 1042–1076, 2011.

J.-M. Lasry and P.-L. Lions. Mean field games. *Japan. J. Math.*, 2(1): 229–260, 2007.

H. A. Latane. Criteria for choice among risky ventures. *J. Political Economy*, 67(2): 144–155, 1959.

G. Latouche. Level-independent quasi-birth-and-death processes. In J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, and J. C. Smith, editors, *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Inc., 2010.

P. Le Gall. *A History of Econometrics in France: From Nature to Models*. Routledge, 2007.

S. Le Roy. Efficient capital markets and martingales. *J. Econom. Literature*, 27: 1583–1621, 1989.

O. Ledoit and M. Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empirical Finance*, 10(5): 603–621, 2003.

O. Ledoit and M. Wolf. Honey, I shrunk the sample covariance matrix. *J. Portfolio Management*, 30(4): 110–119, 2004.

O. Ledoit and M. Wolf. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Stat.*, 40(2): 1024–1060, 2012.

C. M. Lee, B. Mucklow, and M. J. Ready. Spreads, depths, and the impact of earnings information: An intraday analysis. *Review of Financial Studies*, 6(2): 345–374, 1993.

S.-W. Lee and B. E. Hansen. Asymptotic theory for the GARCH(1, 1) quasi-maximum likelihood estimator. *Econometric Theory*, 10(1): 29–52, 1994.

B. N. Lehmann. Fads, martingales, and market efficiency. *Quarterly J. Econ.*, 105(1): 1–28, 1990a.

E. L. Lehmann. Model specification: The views of Fisher and Neyman, and later developments. *Statist. Sci.*, 5(2): 160–168, 1990b.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

H. Leland. Optimal portfolio implementation with transactions costs and capital gains taxes. Working paper, Haas School of Business, University of California Berkeley, 2000.

D. Lepingle. La variation d'ordre p des semi-martingales. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 36(4): 295–316, 1976.

H. Levy and H. M. Markowitz. Approximating expected utility by a function of mean and variance. *The American Economic Review*, 69(3): 308–317, 1979.

J. Lewellen. Momentum and autocorrelation in stock returns. *Review of Financial Studies*, 15(2): 533–564, 2002.

P. A. Lewis and G. S. Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3): 403–413, 1979.

D. Li and W.-L. Ng. Optimal dynamic portfolio selection: Multiperiod mean-variance formulation. *Mathematical Finance*, 10(3): 387–406, 2000.

P.-L. Lions and J.-M. Lasry. Large investor trading impacts on volatility. In *Annales de l'institut Henri Poincaré (C) Analyse non linéaire*, volume 24, pages 311–323. Gauthier-Villars, 2007.

R. B. Litterman. The active risk puzzle. *J. Portfolio Management*, 30(5): 88–93, 2004.

R. B. Litterman. Beyond active alpha. In *CFA Institute Conference Proceedings Quarterly*, volume 25, pages 14–21. CFA Institute, 2008.

H. Liu. Optimal consumption and investment with transaction costs and multiple risky assets. *J. Finance*, 59(1): 289–338, 2004.

N.-L. Liu and H.-L. Ngo. Approximation of eigenvalues of spot cross volatility matrix with a view toward principal component analysis. *ArXiv e-prints*, 2014.

G. M. Ljung and G. E. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2): 297–303, 1978.

A. W. Lo. The adaptive markets hypothesis. *J. Portfolio Management*, 30(5): 15–29, 2004.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

A. W. Lo and J. Hasanhodzic. *The Evolution of Technical Analysis-Financial Prediction From Babylonian Tablets to Bloomberg Terminals.* Bloomberg Press, 2010.

A. W. Lo and A. C. MacKinlay. When are contrarian profits due to stock market overreaction? *Review of Financial Studies*, 3(2): 175–205, 1990.

A. W. Lo and A. C. MacKinlay. *A Non-Random Walk Down Wall Street.* Princeton University Press, 1999.

A. W. Lo and M. T. Mueller. Warning: Physics envy may be hazardous to your wealth! *J. Invest. Manage.*, 8(2): 13–63, 2010.

A. W. Lo, H. Mamaysky, and J. Wang. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *J. Finance*, 55(4): 1705–1770, 2000.

London Stock Exchange Group. Rebuild Order Book Service, February 2011.

F. A. Longstaff and E. S. Schwartz. Valuing American options by simulation: A simple least-squares approach. *Review of Financial Studies*, 14(1): 113–147, 2001.

J. Loveless, S. Stoikov, and R. Waeber. Online algorithms in high-frequency trading. *Communications of the ACM*, 56(10): 50–56, 2013.

D. G. Luenberger. *Investment Science.* Oxford University Press, Oxford, 2nd edition, 2013.

J. Ma. On the principle of smooth fit for a class of singular stochastic control problems for diffusions. *SIAM J. Control Optimiz.*, 30(4): 975–999, 1992.

D. B. Madan and E. Seneta. The variance gamma (V. G.) model for share market returns. *J. Business*, 63(4): 511–524, 1990.

D. B. Madan, P. P. Carr, and E. C. Chang. The variance gamma process and option pricing. *European Finance Review*, 2(1): 79–105, 1998.

A. Madhavan and S. Smidt. An analysis of changes in specialist inventories and quotations. *J. Finance*, 48(5):1595–1628, 1993.

M. J. Magill and G. M. Constantinides. Portfolio selection with transactions costs. *J. Economic Theory*, 13(2): 245–263, 1976.

B. G. Malkiel. *A Random Walk Down Wall Street: The Time-tested Strategy for Successful Investing.* W. W. Norton & Company, 2003.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

P. Malliavin and M. E. Mancino. Fourier series method for measurement of multivariate volatilities. *Finance and Stochastics*, 6(1): 49–61, 2002.

P. Malliavin and M. E. Mancino. A Fourier transform method for nonparametric estimation of multivariate volatility. *Ann. Stat.*, 37(4): 1983–2010, 2009.

M. E. Mancino and M. C. Recchioni. Fourier spot volatility estimator: Asymptotic normality and efficiency with liquid and illiquid high-frequency data. *PloS One*, 10(9): e0139041, 2015.

M. E. Mancino and S. Sanfelici. Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise. *Computational Statistics & Data Analysis*, 52(6): 2966–2989, 2008.

M. E. Mancino and S. Sanfelici. Estimating covariance via Fourier method in the presence of asynchronous trading and microstructure noise. *J. Finan. Economet.*, 9(2): 367–408, 2011a.

M. E. Mancino and S. Sanfelici. Multivariate volatility estimation with high-frequency data using Fourier method. In I. Florescu and F. Viens, editors, *Handbook of Modeling High-Frequency Data in Finance*, pages 243–294. John Wiley & Sons, Inc., New York, 2011b.

B. Mandelbrot. The variation of certain speculative prices. *J. Business*, 36(4): 394–419, 1963.

B. Mandelbrot and H. M. Taylor. On the distribution of stock price differences. *Operations Research*, 15(6): 1057–1062, 1967.

V. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics*, 1(4): 457–483, 1967.

H. M. Markowitz and E. L. van Dijk. Single-period mean-variance analysis in a changing world (corrected). *Finan. Analysts J.*, 59(2): 30–44, 2003.

R. Martin, S. Rachev, and F. Siboulet. Phi-alpha optimal portfolio and extreme risk management. In *The Best of Wilmott* 1: 223–248. Wiley, Hoboken, NJ, 2005.

N. Matloff. *Parallel Computing for Data Science: With Examples in R, C++ and CUDA*. Chapman and Hall/CRC, 2015.

R. P. McAfee. A dominant strategy double auction. *J. Economic Theory*, 56(2): 434–450, 1992.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

J. H. McCulloch. Simple consistent estimators of stable distribution parameters. *Communications in Statistics–Simulation and Computation*, 15(4): 1109–1136, 1986.

R. E. McCulloch and R. S. Tsay. Nonlinearity in high-frequency financial data and hierarchical models. *Studies in Nonlinear Dynamics & Econometrics*, 5(1): 1–18, 2001.

R. C. Merton. Lifetime portfolio selection under uncertainty: The continuoustime case. *The Review of Economics and Statistics*, 51(3): 247–257, 1969.

R. C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *J. Economic Theory*, 3(4): 373–413, 1971.

R. C. Merton. Option pricing when underlying stock returns are discontinuous. *J. Finan. Econom.*, 3(1–2): 125–144, 1976.

R. C. Merton. *Continuous-Time Finance.* Macroeconomics and Finance Series. Basil Blackwell, 1990.

A. Meucci. *Risk and Asset Allocation.* Springer, 2005.

A. Meucci. Black–Litterman approach. In R. Cont, editor, *Encyclopedia of Quantitative Finance.* John Wiley & Sons, Ltd, 2010.

S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* Communications and Control Engineering Series. Springer-Verlag, 1993.

R. O. Michaud. *Efficient Asset Management.* Harvard Business School Press, 1989.

P. R. Milgrom. *Putting Auction Theory to Work.* Cambridge University Press, 2004.

M. H. Miller, J. Muthuswamy, and R. E. Whaley. Mean reversion of Standard & Poor's 500 Index basis changes: Arbitrage-induced or statistical illusion? *J. Finance*, 49(2): 479–513, 1994.

C. C. Moallemi and K. Yuan. The value of queue position in a limit order book. Working paper, 2015.

C. C. Moallemi, B. Park, and B. Van Roy. The execution game. Preprint, 2009.

F. Modigliani and L. Modigliani. Risk-adjusted performance. *J. Portfolio Management*, 23(2): 45–54, 1997.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

J. Møller and J. G. Rasmussen. Perfect simulation of Hawkes processes. *Advan. Appl. Probab.*, 37(3): 629–646, 2005.

J. Møller and J. G. Rasmussen. Approximate simulation of Hawkes processes. *Methodology and Computing in Applied Probability*, 8(1): 53–64, 2006.

T. J. Moskowitz, Y. H. Ooi, and L. H. Pedersen. Time series momentum. *J. Finan. Econom.*, 104(2): 228–250, 2012.

A. J. Motyer and P. G. Taylor. Decay rates for quasi-birth-and-death processes with countably many phases and tridiagonal block generators. *Adv. Appl. Probab.*, 38(2): 522–544, 2006.

A. Munk, J. Schmidt-Hieber, et al. Nonparametric estimation of the volatility function in a high-frequency model corrupted by noise. *Electron. J. Stat.*, 4: 781–821, 2010.

W. Murray and K.-M. Ng. An algorithm for nonlinear optimization problems withbinary variables. *Computational Optimization and Applications*, 47(2): 257–288, 2008.

W. Murray and U. V. Shanbhag. A local relaxation method for nonlinear facility location problems. *Multiscale Optimization Methods and Applications*, 82: 173–204, 2006.

W. Murray and U. V. Shanbhag. A local relaxation approach for the siting of electrical substation. *Computational Optimization and Applications*, 38(3): 299–303, 2007.

W. Murray and H. Shek. A local relaxation method for the cardinality constrained portfolio optimization problem. *Computational Optimization and Applications*, 53(3): 681–709, 2012.

J. Muthuswamy, S. Sarkar, A. Low, and E. Terry. Time variation in the correlation structure of exchange rates: High-frequency analyses. *J. Futures Markets*, 21(2): 127–144, 2001.

R. B. Myerson and M. A. Satterthwaite. Efficient mechanisms for bilateral trading. *J. Economic Theory*, 29(2): 265–281, 1983.

A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM J. Optimiz.*, 19(4): 1574–1609, 2009.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

V. Niederhoffer and M. F. M. Osborne. Market making and reversal on the stock exchanges. *J. Amer. Statist. Assoc.*, 15(61): 897–916, 1966.

N. Nisan and A. Ronen. Computationally feasible VCG mechanisms. *J. Artif. Intell. Res.*, 29: 19–47, 2007.

A. A. Obizhaeva and J. Wang. Optimal trading strategy and supply/demand dynamics. *J. Financial Markets*, 16(1): 1–32, 2013.

Y. Ogata. The asymptotic behavour of maximum likelihood estimators for stationary point processes. *Ann. Inst. Statist. Math.*, 30: 243–261, 1978.

Y. Ogata. On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1): 23–31, 1981.

Y. Ogata, K. Katsura, and J. Zhuang. TIMSAC84: Statistical analysis of series of events (TIMSAC84-SASE) version 2. *Computer Science Monographs*, no. 32, 2006.

B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications. Springer*, New York, 6th edition, 2003.

A. Onatski. Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4): 1004–1016, 2010.

A. Onatski. Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2): 244–258, 2012.

E. Ordentlich and T. M. Cover. The cost of achieving the best portfolio in hindsight. *Mathematics of Operations Research*, 23(4): 960–982, 1998.

M. F. M. Osborne. Brownian motion in the stock market. *Operations Research*, 7: 145–173, 1959.

T. Ozaki. Maximum likelihood estimation of hawkes' self-exciting point processes. *Ann. Inst. Statist. Math.*, 31(1): 145–155, 1979.

E. Page. Continuous inspection schemes. *Biometrika*, 41(1/2): 100–115, 1954.

J. Palczewski, R. Poulsen, K. R. Schenk-Hoppé, and H. Wang. Dynamic portfolio optimization with transaction costs and state-dependent drift. *European J. Oper. Res.*, 243(3): 921–931, 2015.

S. Peng. A general stochastic maximum principle for optimal control problems. *SIAM J. Control Optimiz.*, 28(4): 966–979, 1990.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

A. F. Perold and W. F. Sharpe. Dynamic strategies for asset allocation. *Finan. Analysts J.*, 51(1): 149–160, 1995.

M. H. Pesaran and A. Timmermann. A simple nonparametric test of predictive performance. *J. Business & Economic Statistics*, 10(4): 461–465, 1992.

H. Pham. On some recent aspects of stochastic control and their applications. *Probab. Surveys*, 2: 506–549, 2005.

S. Pliska. *Introduction to Mathematical Finance.* Blackwell Publishers, Oxford, 1997.

M. Potters and J.-P. Bouchaud. More statistical properties of order books and price impact. *Physica A: Statistical Mechanics and its Applications*, 324(1): 133–140, 2003.

W. B. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality.* John Wiley & Sons, Inc., 2007.

W. B. Powell and I. O. Ryzhov. Optimal learning and approximate dynamic programming. In D. L. Frank L. Lewis, editor, *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*, IEEE Press Series on Computational Intelligence, pages 410–431. Wiley-IEEE Press, 2013.

D. J. Powers. Price Reporting and Dissemination; The Placement and Execution of Orders. In E. A. Gaumnitz, editor, *Futures Trading Seminar: A Commodity Marketing Forum for College Teachers of Economics*, volume III. Mimir Publishers, Inc., 1966.

S. Predoiu, G. Shaikhet, and S. Shreve. Optimal execution in a general onesided limit-order book. *SIAM J. Finan. Math.*, 2(1): 183–212, 2011.

S. J. Press. A compound events model for security prices. *J. Business*, 40(3): 317–335, 1967.

Principal Traders Group of The Futures Industry Association. Debunking the myths of high-frequency trading: Opinions published in response to Michael Lewis' Flash Boys and recent comments by NY Attorney General Eric Schneiderman, 2015.

S. Rachev and S. Mittnik. *Stable Paretian Models in Finance.* Wiley, Hoboken, NJ, 2000.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

J. Regnault. *Calcul des Chances et Philosophie de la Bourse.* Mallet-Bachelier and Castel, 1863.

W. Ren. A mean-field LOB model. Ph.D. thesis, Chapter 3, Department of Mathematics, Stanford University, 2016.

D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, Berlin, 3rd edition, 1999.

M. K. Richter and K.-C. Wong. Non-computability of competitive equilibrium. *Econom. Theory*, 14(1): 1–27, 1999.

H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Statist.*, pages 400–407, 1951.

R. Roll. A simple implicit measure of the effective bid-ask spread in an efficient market. *J. Finance*, 39: 1127–1139, 1984.

J. P. Romano and M. Wolf. Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4): 1237–1282, 2005.

S. A. Ross. The arbitrage theory of capital asset pricing. *J. Economic Theory*, 13(3): 341–360, 1976.

S. M. Ross. *Simulation.* Academic Press, Inc., Orlando, FL, 5th edition, 2015.

I. Rosu. A dynamic model of the limit order book. *Review of Financial Studies*, 22: 4601–4641, 2009.

A. Roth and M. A. O. Sotomayor. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis.* Econometric Society Monographs. Cambridge University Press, Cambridge, 1990.

T. H. Rydberg and N. Shephard. Dynamics of trade-by-trade price movements: Decomposition and models. *J. Finan. Economet.*, 1(1): 2–25, 2003.

T. Sabel, J. Schmidt-Hieber, and A. Munk. Spot volatility estimation for highfrequency data: Adaptive estimation in practice. In *Modeling and Stochastic Learning for Forecasting in High Dimensions*, pages 213–241. Springer, 2015.

P. A. Samuelson. Proof that properly anticipated prices uctuate randomly. *Industrial Management Review*, 6(2): 41–49, 1965.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

P. A. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics*, 51(3): 239–246, 1969.

P. A. Samuelson. Mathematics of speculative price. *SIAM Review*, 15(1): 1–42, 1973.

A. Schied and T. Schöneborn. Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets. *Finance and Stochastics*, 13(2): 181–204, 2009.

A. Schied, T. Schöneborn, and M. Tehranchi. Optimal basket liquidation for cara investors is deterministic. *Appl. Mathemat. Finance*, 17(6): 471–489, 2010.

M. Scholes and J. Williams. Estimating betas from nonsynchronous data. *J. Finan. Econom.*, 5(3): 309–327, 1977.

H. H. Shek. *Statistical and Algorithm Aspects of Optimal Portfolios.* PhD thesis, Stanford University Institute of Computational and Mathematical Engineering, 2010.

N. Shephard and K. Sheppard. Realising the future: Forecasting with highfrequency-based volatility (HEAVY) models. *J. Appl. Economet.*, 25(2): 197–231, 2010.

Y. Shetty and S. Jayaswal. *Practical .NET for Financial Markets.* Apress, 2006.

R. J. Shiller. *Irrational Exuberance.* Princeton University Press, 3rd edition, 2015.

D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587): 484–489, 2016.

Y. Simaan. Estimation risk in portfolio selection: The mean variance model versus the mean absolute deviation model. *Management Science*, 43(10): 1437–1446, 1997.

C. A. Sims. Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review*, 36(5): 975–1000, 1992.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

E. Smith, J. Doyne Farmer, L. Gillemot, and S. Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative Finance*, 3: 481–514, 2003.

M. L. Stein. A comparison of generalized cross validation and modified maximum likelihood for estimating the parameters of a stochastic process. *Ann. Statist.*, 18(3): 1139–1157, 1990.

M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging.* Springer Series in Statistics. Springer, New York, 1999.

J. H. Stock and M. W. Watson. Forecasting using principal components from a large number of predictors. *J. Amer. Statist. Assoc.*, 97(460): 1167–1179, 2002.

J. H. Stock and M. W. Watson. Dynamic factor models. *Oxford handbook of economic forecasting*, 1: 35–59, 2011.

S. Stoikov and M. Saglam. Option market making under inventory risk. *Review of Derivatives Research*, 12: 55–79, 2009.

R. Sullivan, A. Timmermann, and H. White. Data-snooping, technical trading rule performance, and the bootstrap. *J. Finance*, 54(5): 1647–1691, 1999.

R. S. Sutton. Learning to predict by the methods of temporal differences. *Machine Learning*, 3(1): 9–44, 1988.

V. N. Temlyakov. Weak greedy algorithms. *Advances in Computational Mathematics*, 12(2): 213–227, 2000.

R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, pages 267–288, 1996.

J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, 53(12): 4655–4666, 2007.

R. Tsay. *Analysis of Financial Time Series.* Wiley, New Jersey, 3rd edition, 2010.

J. N. Tsitsiklis and B. Van Roy. Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing highdimensional financial derivatives. *IEEE Trans. Automat. Contr.*, 44(10): 1840–1851, 1999.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

G. Tsoukalas, J. Wang, and K. Giesecke. Dynamic portfolio execution, 2014.

H. R. Varian. Position auctions. *Int. J. Indust. Organiz.*, 25(6): 1163–1178, 2007.

W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *J. Finance*, 16(1): 8–37, 1961.

V. Voev and A. Lunde. Integrated covariance estimation using high-frequency data in the presence of noise. *J. Finan. Economet.*, 5(1): 68–104, 2007.

A. R. Ward and P. W. Glynn. A diffusion approximation for a markovian queue with reneging. *Queueing Systems*, 43(1–2): 103–128, 2003.

A. R. Ward and P. W. Glynn. A diffusion approximation for a GI/GI/1 queue with balking or reneging. *Queueing Systems*, 50(4): 371–400, 2005.

S. Watanabe. On discontinuous additive functionals and Lévy measures of a Markov process. *Japanese J. Math.*, 34: 53–70, 1964.

C. J. Watkins and P. Dayan. Q-learning. *Machine Learning*, 8(3–4): 279–292, 1992.

D. Weaver. Minimum obligations of market makers. Technical report, UK Government Foresight Project, Future of Computer Trading in Financial Markets, 2012. Economic Impact Assessment EIA7.

P. Weber and B. Rosenow. Order book approach to price impact. *Quantitative Finance*, 5(4): 357–364, 2005.

P. Weiss. L'hypothèse du champ moléculaire et la propriété ferromagnétique. *J. Phys. Theor. Appl.*, 6(1): 661–690, 1907.

H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1): 1–25, 1982.

H. White. A reality check for data snooping. *Econometrica*, 68(5): 1097–1126, 2000.

W. Whitt. *Stochastic-process Limits: An Introduction to Stochastic-process Limits and Their Application to Queues.* Springer, 2002.

E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, pages 548–564, 1955.

J. T. Williams. Capital asset prices with heterogeneous beliefs. *J. Finan. Econom.*, 5(2): 219–239, 1977.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

A. S. Willsky and H. L. Jones. A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Trans. Automat. Contr.*, 21(1): 108–112, 1976.

M. Wooldridge. *An Introduction to Multiagent Systems.* John Wiley & Sons, 2009.

Y. Xia. Learning about predictability: The effects of parameter uncertainty on dynamic asset allocation. *J. Finance*, 56(1): 205–246, 2001.

D. Xiu. Quasi-maximum likelihood estimation of volatility with highfrequency data. *J. Economet.*, 159: 235–250, 2010.

V. I. Zakamouline. A unified approach to portfolio optimization with linear transaction costs. *Mathematical Methods of Operations Research*, 62(2): 319–343, 2005.

L. Zhang. Estimating covariation: Epps effect, microstructure noise. *J. Econometrics*, 160(1): 33–47, 2011.

L. Zhang, P. A. Mykland, and Y. Aït-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high frequency data. *J. Amer. Statist. Assoc.*, 100(472): 1394–1411, 2005.

B. Zheng, E. Moulines, and F. Abergel. Price jump prediction in a limit order book. *J. Math. Finance*, 3(2): 242–255, 2013.

B. Zhou. High-frequency data and volatility in foreign exchange rates. *J. Business & Econom. Stat.*, 14(1): 45–52, 1996.

X. Y. Zhou and D. Li. Continuous-time mean-variance portfolio selection: A stochastic LQ framework. *Appl. Math. Optimiz.*, 42(1): 19–33, 2000.

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *J. Roy. Statist. Soc. Ser.* B, 67(2): 301–320, 2005.

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

索　　引

ϵ 最优策略, 182
ε 最优分配规则, 224
CFTC (商品期货交易委员会), 258
CRRA 函数, 66
CUSUM 变点检测规则, 41
Dirichlet 权重投资组合, 92
Epps 效应, 107
Euler-Lagrange 方程, 176
Fama-French 三因子模型和五因子模型, 48
FIX/FAST (FIX Adapted for STreaming) 协议, 244
Fourier-Fejér 反演公式, 113
Hamilton-Jacobi-Bellman (HJB) 方程, 84
Hawkes 过程, 118
HEAVY 模型, 124
Ho 和 Stoll 模型, 210
Java, 248
Jensen 指数, 9
Kalman-Bucy 滤波, 72
Kaplan-Meier 统计量, 231
Kiefer-Wolfowitz 框架, 233
Kruk 模型, 162
Lasso Logistic 回归方法, 149
Lévy 单纯跳跃过程, 124
LOB, 151
LOB 的多主体模型, 167
LOB 的形状函数, 185
LQG 控制问题, 94
Markowitz 方法, 28
McAfee 机制, 14
Merton 比例, 67
NPEB 法则, 28
NYSE Arca, NYSE MKT, 253
NYSE 交易和报价 (TAQ) 数据库, 116
Ornstein-Uhlenbeck (OU) 过程, 212
QMLE 估计量, 106
RealGARCH 建模, 130
Riccati 方程, 94
S&P 500 指数, 62
Sharpe 比率 (SR), 8
Sortino 比率, 9
SPY, 239

A

按比例的订单位置, 156
按比例的净订单流过程, 156
按比例的最佳买入和卖出序列, 156
按比例分配, 243

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

暗池, 210
凹函数, 66

B

巴兹全球市场 (BATS global markets, 缩写 BATS), 250
半价单顺序, 250
半正定多元已实现核估计量, 132
报价数据, 247
被动委托, 252
标值点过程, 116
冰山订单, 251
波士顿证券交易所, 245
博弈论, 167
博弈论的框架, 13
补偿, 278
捕食交易, 178
不同期限的 “支节”, 239

C

操作风险, 257
策略迭代, 81
策略风险, 259
策略网络, 91
长记忆过程, 168
常规控制, 215
场外交易市场, 237
超常收益, 59
成本最小化, 187
乘法误差模型 (MEM), 123
乘数价格影响, 179
惩罚函数, 76
持仓量, 240
重抽样, 62
从属过程, 31
从属随机过程理论, 107
从众心理, 43
存货风险, 217
错位偏误, 131

D

代价函数, 87
单基金定理, 22
单期均值–方差优化, 64
单因子模型, 24, 58
单元测试, 257
当期收益率, 17
道琼斯工业平均指数 (DJIA), 267
等权重的低效率投资组合, 22
点过程, 116
电子通信网络, 2
跌幅, 41
订单保护, 253
订单管理, 249
订单或报价驱动, 98
订单路径, 209
订单网关, 242
订单位置, 154
动量策略, 43
动态规划, 65
动态投资策略, 58
短 (期) 波动率, 164
短视或确定性等价准则, 86
对冲基金, 1
多播, 244
多重检验, 44
多代理模型, 169
多期投资组合管理, 64
多项式时间, 64
多因子定价模型, 23
多元已实现核, 110

E

二次变差, 97

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

二次规划, 27
二进制编码 (SBE), 244

F

反卷积, 128
方差压缩投资组合, 61
方向性策略, 8
非不贸易的区域, 230
非参数回归, 41
非回归测试, 257
非凸优化问题, 64
非线性滤波, 96
“非智能” 方法, 168
费城证券交易所 (Philadelphia Stock Exchange, 缩写 PSX), 250
分层, 256
分层算法, 148
分离准则, 94
风险管理, 257
风险偏好系数, 174
风险相关的业绩 (RAP), 9
负一期自回归系数, 99

G

改进 Kaplan-Meier 估计量, 222
概率权重交易量, 141
杠杆效应, 17
高频交易 (HFT), 5
高频做市商, 63
个体风险, 22
更新, 246
更新过程, 156
更新时间, 108
工具, 243
公开喊价, 1
公募基金, 7
共轭梯度法, 77
供给和需求, 12
估计有效前沿, 25
股票型基金, 8
股市崩盘, 172
广义抽样时间, 108
广义似然比率 (GLR), 55
广义线性模型, 119
国家市场体系管理规则 (Reg NMS), 253

H

合规风险, 258
核平滑, 104
互换期权, 239
回归测试, 257
汇总成交量 (TCV), 245
会话, 242
混合型基金, 8
货币型基金, 8

J

机制设计, 13
积分方差, 101
积极型投资组合管理中的 α 和 β, 58
积极型学习, 90
基本面交易员, 172
基本面驱动的交易策略 (FMQ), 5
基金经理, 57
基于订单, 137
基于级别的市场数据, 248
基于价格水平, 137
即时波动率, 97
即时成本, 79
计算学习理论, 150
技术分析, 41
价格–时间优先, 1
价格操纵, 184
价格冲击, 214
价格聚集, 17
价格影响方程, 178

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

尖峰, 31
简化形式模型, 151
降低到最小, 250
交易前和交易后的风险管理, 11
交易收益, 174
交易所交易基金 (ETF), 8
交易所佣金, 62
交易通过, 253
交易员, 57
交易之间间隔, 117
截面动量, 42
金融行业监管机构 FINRA, 254
金融信息交换 (FIX) 协议, 4
近邻松弛二次规划, 79
经济物理, 167
精确时间协议, 248
竞争均衡, 12
局部松弛方法, 78
决策树, 150
均衡价格, 12
均衡投资组合, 24
均值回归比率, 160

K

开盘至收盘收益率, 130
可销售限价订单, 138
刻度时钟, 213
跨市场的扫描订单 (ISO), 253
宽平稳模型, 35
扩散极限, 151

L

拉格朗日乘子, 26
拉格朗日方法, 226
理论决策, 65
利差, 179
连续成本函数, 84
连续基点, 260
连续时间半鞅理论, 101
连续时间马尔可夫过程, 81
连续时间马尔可夫链, 152
连续收益率, 17
量化基金, 7
量化投资组合管理, 57
量化研究者, 57
零和交易, 184
零膨胀, 232
流动极限, 156
流动性溢价, 5
伦敦证券交易所 (LSE), 145

M

马尔可夫控制准则, 82
卖空规则, 255
脉冲控制, 213
美国证券交易委员会 (SEC), 2
(美元) 期限, 259
蒙特卡罗 Rollout 算法, 91
蒙特卡罗树状搜索, 91
孟买股票交易所 (BSE), 216
幂定律, 232
名义价值, 136
模型风险, 259

N

纳什均衡, 13
纳斯达克交易所 (NASDAQ), 2
内幕消息, 59
拟合优度, 148
拟生灭过程 (QBD), 159
逆 Wishart 分布, 23
逆交易量权重公平价值 (IWFV) 价格, 142
逆向投资策略, 43
纽约商品交易所 (NYMEX), 3

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

纽约证券交易所 (NYSE), 3

O

欧洲期货交易所, 244

P

帕累托分布, 30
帕累托形式的随机游走, 30
排队论, 163
配对交易策略, 42
匹配引擎, 243
漂移项, 72
频率数据, 247
平滑拟合准则, 182
平均场论, 169
平稳转移概率, 80

Q

期望效用, 65
欺骗, 256
奇异型随机控制问题, 68
奇异值分解, 73
前笔交易, 108
前进法/后退法逐步方法, 150
前景理论, 43
嵌入式积极型投资组合, 61
清算成本, 174, 177
清算和结算, 246
清算所, 62
清算问题, 174
全国最优买卖报价 (NBBO), 217
全局平滑算法, 76
全期望公式, 79
全球宏观策略, 5, 8

R

认识偏差, 43
熔断机制, 11
软件测试, 257

S

筛选规则, 41
商业现货 (COTS) 操作, 3
上限, 239
社会选择方程, 13
深度学习, 91
神经网络, 151
神经元动态规划, 90
时间标记同步, 248
市场操纵, 178
市场的 β, 60
市场风险, 173
市场化订单, 143
市场投资组合, 60
市场微观结构, 63
市场影响, 62
市场约束, 16
市场指令, 136
事件驱动型策略, 8
适应性控制, 87
收费结构, 245
收盘, 250
收盘至收盘的收益率, 130
受保护报价, 253
数据驱动型的量化交易策略, 17
双随机点过程, 143
瞬时波动率, 112
瞬时协方差, 115
瞬时协方差的傅里叶估计量, 131
似然方程, 36
松弛变量, 77
随机逼近 (SA), 232
随机控制, 65
随机漫步, 16
随机森林, 150
随机优化, 173

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

随机最大化准则, 171
缩减技术, 23

T

贪心算法, 221
弹性速率, 186
探索和开发, 224
调整后 Dirichlet 核, 114
同步方法, 111
统计套利, 5
统计套利机会 (SAOs), 7
统一报价系统 (CQS), 253
统一证券行情系统 (CTS), 253
投资组合的权重, 58
投资组合优化, 57
图形化用户界面 (GUI) 系统, 3

W

瓦尔拉斯 (Walrasian) 均衡, 12
微观结构噪声, 102
维度诅咒, 70
稳定遍历过程, 157
稳定过程, 31
稳定帕累托分布, 30
无穷小生成元, 81
无限时间折现问题, 80

X

稀疏抽样估计量, 110
系统暴露, 259
系统性策略, 5
系统性风险, 22
下限, 239
先进先出, 154
现代投资组合理论, 21
现行的最大价格, 41
限价委托市场, 13
限价指令, 136
线性搜索, 78
相对风险厌恶系数, 66
相对价值策略, 8
相对价值交易, 238
向前向后随机微分方程 (FBSDE), 171
消极型管理, 57
消极型学习, 90
协变差, 107
新版本环境, 258
新闻效应, 107
新兴市场, 60
信息比率 (IR), 9
序贯学习, 221
序列反应模型, 158
选取合适的 “特性”, 149

Y

延迟套利, 238
鞅回归, 34
样本外表现, 38
一致性测试, 258
移动平均规则, 42
已实现 GARCH, 122
已实现测度的 GARCH 模型, 123
已实现方差, 97
已实现核, 104
因子选择, 23
印度证券交易委员会 (SEBI), 216
有限窗宽下的 GLR 故障检测, 42
有效价格发现, 1
有效前沿, 22, 23
有效市场假说 (EMH), 6
预定出价, 210
预平均方法, 105
阈值流动性公平价值 (TLFV) 价格, 142
约束优化问题, 226

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

运营经理, 57

Z

暂时性价格影响, 174
增强学习, 90
涨幅, 41
折现系数, 66
证券工业自动化公司 (SIAC), 253
证券信息处理程序 (SIP), 252
支持向量机 (SVM), 150
芝加哥期货交易所 (CBOT), 2
芝加哥商品交易所 (CME), 2
执行落差, 62
直接合同, 240
直通式处理系统 (STP), 4
指定做市商 (DMM), 256
指数效用, 211
智能下单 (SOR), 172
中间价, 213
终端成本, 79
洲际交易所 (ICE), 137
逐笔交易, 98
主机托管, 246
主要市场, 217
状态变量, 210
追踪指数, 57
酌情策略, 5
资本市场的有效性和鞅性, 15
自激励点过程 (SEPP), 116
自身可分解, 30
总体错误率, 44
最小化清算成本, 174
最优订单拆分, 218
最优控制, 65
最优停时, 81
最优执行与配置, 172
左侧的半方差, 9
做市和负 MA(1) 自相关关系, 129
做尾盘, 256

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

应用统计学丛书

书号	书名	著译者
	量化交易：算法、分析、数据、模型和优化	黎子良 等 著 冯玉林、刘庆富 译
	马尔可夫过程及其应用：算法、网络、基因与金融	Étienne Pardoux 著 许明宇 译
	临床试验设计的统计方法	尹国至、石昊伦 著
	随机场：分析与综合（修订扩展版）	Erik Vanmarke 著 陈朝晖、范文亮 译
	统计思维与艺术：统计学入门	Benjamin Yakir 著 徐西勒 译
	诊断医学中的统计学方法（第二版）	侯艳、李康、宇传华、 周晓华 译
	高等统计学概论	赵林城、王占锋 编著
	纵向数据分析方法与应用（英文版）	刘宪
	生物数学模型的统计学基础（第二版）	唐守正、李勇、符利勇 著
	R 软件教程与统计分析：入门到精通	潘东东、李启寨、唐年胜 译
	随机估计及 VDR 检验	杨振海
	随机域中的极值统计学：理论及应用（英文版）	Benjamin Yakir 著
	高等计量经济学基础	缪柏其、叶五一
	金融工程中的蒙特卡罗方法	Paul Glasserman 著 范韶华、孙武军 译
	大维统计分析	白志东、郑术蓉、姜丹丹
	结构方程模型：Mplus 与应用（英文版）	王济川、王小倩 著

实时行情接口：www.infoway.io

实时行情接口：www.infoway.io (覆盖A股、港股、美股、外汇、期货、加密货币)

续表

书号	书名	著译者
	生存分析：模型与应用 (英文版)	刘宪
	MINITAB 软件入门：最易学实用的统计分析教程	吴令云 等 编著
	结构方程模型：方法与应用	王济川、王小倩、姜宝法 著
	结构方程模型：贝叶斯方法	李锡钦 著 蔡敬衡、潘俊豪、周影辉 译
	随机环境中的马尔可夫过程	胡迪鹤 著
	统计诊断	韦博成、林金官、解锋昌 编著
	R 语言与统计分析	汤银才 主编
	属性数据分析引论 (第二版)	Alan Agresti 著 张淑梅、王睿、曾莉 译
	金融市场中的统计模型和方法	黎子良、邢海鹏 著 姚佩佩 译

网上购书： www.hepmall.com.cn, gdjycbs.tmall.com, academic.hep.com.cn, www.amazon.cn, www.dangdang.com

其他订购办法：

各使用单位可向高等教育出版社电子商务部汇款订购。书款通过银行转账，支付成功后请将购买信息发邮件或传真，以便及时发货。购书免邮费，发票随书寄出（大批量订购图书，发票随后寄出）。

单位地址： 北京西城区德外大街4号
电　　话： 010-58581118
传　　真： 010-58581113
电子邮箱： gjdzfwb@pub.hep.cn

通过银行转账：
户　　名： 高等教育出版社有限公司
开 户 行： 交通银行北京马甸支行
银行账号： 110060437018010037603

实时行情接口：www.infoway.io