

WILEY



基于WorldQuant团队多年的成功经验
覆盖Alpha设计的方方面面

金融工程师、量化研究员、投资分析师和宽客的优秀指南

寻找Alpha

量化交易策略

[美] Igor Tulchinsky 等著
杨旭光 周晶 程陈 等译

Finding Alphas

A Quantitative Approach
to Building Trading Strategies

 中国工信出版集团

 电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn



寻找Alpha

量化交易策略

[美] Igor Tulchinsky◎等著
杨旭光 周晶 程陈◎等译

Finding Alphas

A Quantitative Approach
to Building Trading Strategies

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书是世坤投资 (WorldQuant) 公司的众多资深专家, 包括其创始人兼 CEO (Igor Tulchinsky) 的心血之作, 书中对很多能点石成金的交易信号进行了深入分析, 并且提供了可以用来做练习和开发的工具。

本书为读者提供了从数据中找到隐藏信号的方法, 且在其他领域中同样适用。本书是很多不同作者的文章合集, 通过多维度视角分析相似性, 提供了独特的方法来开发 Alpha 信号, 同时涵盖了抽象理论和具体技术等多个方面。书中还提供了关于世坤投资的在线模拟工具 WebSim™ 的实用指导, 从而帮助读者实践 Alpha 信号的开发过程。

通过本书, 读者将学习到信息研究的基本注意事项, 包括基本面分析、统计套利、Alpha 多样性等, 然后可深入研究更高级的领域和更复杂的设计。

Finding Alphas: A Quantitative Approach to Building Trading Strategies, 9781119057864, Igor Tulchinsky et al

Copyright © Igor Tulchinsky et al., WorldQuant Virtual Research Center

All Rights Reserved. Authorised translation from the English language edition published by John Wiley and Sons Ltd. Responsibility for the accuracy of the translation rests solely with PHEI and is not the responsibility of John Wiley and Sons Ltd. Copies of this book sold without a Wiley sticker on the cover are unauthorized and illegal.

本书简体中文版专有翻译出版权由 John Wiley & Sons, Ltd. 授予电子工业出版社。未经许可, 不得以任何方式复制或抄袭本书的任何部分。本书封底贴有 Wiley 防伪标签, 无标签者不得销售。

版权贸易合同登记号 图字: 01-2016-1973

图书在版编目 (CIP) 数据

寻找 Alpha: 量化交易策略 / (美) 伊戈尔·图利钦斯基 (Igor Tulchinsky) 等著; 杨旭光等译. —北京: 电子工业出版社, 2019.8

(量化交易丛书)

书名原文: Finding Alphas: A Quantitative Approach to Building Trading Strategies

ISBN 978-7-121-33136-7

I. ①寻… II. ①伊… ②杨… III. ①对冲基金—投资—经济策略 IV. ①F830.59

中国版本图书馆 CIP 数据核字 (2017) 第 294893 号

责任编辑: 高洪霞

印 刷: 三河市鑫金马印装有限公司

装 订: 三河市鑫金马印装有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编: 100036

开 本: 720×1000 1/16 印张: 11.75 字数: 211 千字

版 次: 2019 年 8 月第 1 版

印 次: 2019 年 8 月第 1 次印刷

定 价: 69.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888, 88258888。

质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。本书咨询联系方式: (010) 51260888-819, faq@phei.com.cn。

译者序

谷歌 AlphaGo 2.0 战胜了国际围棋冠军柯洁；高盛将资产在 100 万美元的“穷人”全部扔给机器人投顾服务；一级市场中量化投资平台也已商业化运营；科技和金融市场更加紧密结合，Fintech 的发展如火如荼。

非常有幸生在这个年代，适逢其会经历这些改变；非常有幸在一级市场和二级市场都有从业经历，更能深刻体会这些改变；非常有幸自己的计算机专业背景让我接触了 Alpha 策略，也希望自己能在这个领域有所建树；非常有幸能翻译《寻找 Alpha：量化交易策略》这本书，在翻译过程中揣摩和理解其他宽客们的想法，对我自身也是一种提高！

翻译工作本身由有志于或者正在从事量化交易的团队成员协同合作完成（细节如下），我负责了汇总及后期的校对工作，非常感谢大家的通力合作，“We are a team!”：

我负责翻译第 1 章~第 4 章和第 20 章；

债券市场高级交易员（CFA&FRM）程陈先生负责翻译第 5 章~第 10 章；

中交三航院邮轮港口技术研究中心副主任魏冰女士负责翻译第 11 章~第

15 章、第 19 章和第 26 章；

半导体行业资深从业者施建安先生负责翻译第 16 章~第 18 章；

半导体行业资深从业者与私募基金经理（上海大学 MBA）汪锡志先生翻译了第 27 章~第 30 章；

棋盘资本陈越洋先生及棋盘资本姜珂歆女士共同翻译了第 21 章~第 24 章；

南京师范大学翻译专业李颖同学翻译了第 25 章、第 31 章~第 34 章。

最后，非常感谢电子工业出版社编辑高洪霞老师的理解和支持。

本书对 Alpha 策略设计从理论到实现上都做了阐述，而不仅仅是从学术角度进行解释，很有实战指导价值，虽然因为市场等原因，我们并不能直接使用 WebSimTM 工具，但是其中的理念仍然值得我们参考。本书更适合量化初学者入门，助其建立系统的体系知识。

希望大家能从本书中获益，有问题可以联系：quantAlpha@163.com。

杨旭光

写于阳澄湖畔

轻松注册成为博文视点社区用户（www.broadview.com.cn），扫码直达本书页面。

- **提交勘误：**您对书中内容的修改意见可在 [提交勘误](#) 处提交，若被采纳，将获赠博文视点社区积分（在您购买电子书时，积分可用来抵扣相应金额）。
- **交流互动：**在页面下方 [读者评论](#) 处留下您的疑问或观点，与其他读者一同学习交流。

页面入口：<http://www.broadview.com.cn/33136>



前 言

本书讲述如何寻找 Alpha，包含了奋斗在量化交易前线的成功宽客们的基于不同视角的系列文章，涵盖了广泛的主题，包括有关 Alpha 存在的理论，以及更具体、更侧重于技术层面的 Alpha 的创建。

第一部分展现了 Alpha 的创建，并简要介绍了 Alpha 的生命周期和止损的观点。

第二部分主要从专业角度介绍 Alpha 的设计，比如信息研究中的注意事项、开发 Alpha 的关键步骤，以及高效 Alpha 的评估和改善。在这部分讨论中，关键的技术点主要在于：换手率、回测、基本面分析、股票量价、统计套利、过度拟合，以及 Alpha 多样性。

第三部分从特定主题入手研究 Alpha 的设计，包括不同资产类型（比如期货和现货）的 Alpha 设计、动量 Alpha 的开发、新闻和社交媒体对于股票收益的影响。

第四部分介绍了基于网页的 Alpha 开发工具 WebSimTM。我们诚邀宽客使用这一免费工具去理解 Alpha 回测（也被视为 Alpha 模拟），并最终创建他们自己的信号。

第五部分为结语。

致 谢

在本书中，我们向读者展现了发现 Alpha 的点石成金术的文章。本书是由 WorldQuant 的创始人、总监、经理、机构内投资经理及量化研究人员共同写的。本书最主要的写作目的有两部分：展现定义 Alpha 的多个视角，以及如何发现 Alpha。在 WorldQuant，我们相信没有哪一个观点是最好的、唯一的答案，一系列的方法总是优于单一的方法。书中也展现了在线金融市场模拟工具：WebSim™，它使用户和咨询顾问可以去创造、测试、模拟和跟踪 Alpha。

在此感谢 Rohit Agarwal、Ionut Aron、Pankaj Bakliwal、Scott Bender 等人，他们在本书出版的过程中做了很多贡献，付出了辛勤的努力和大量的精力。

要特别感谢我们的出版经纪人 Wendy Goldman Rohm，感谢她在早期草稿阶段关键而有前瞻性的批注，以及在整个项目过程中给出的杰出的举证和很多很好的建议；感谢 Werner Coetzee 和 Wiley 公司的团队，感谢他们专业的指导和详细而有用的建议；也十分感谢 Jeffrey Blomberg 和 Kristin Chach，

WorldQuant 最专业的律师团队，感谢他们明智的法律建议和所做的大量工作，以确保我们走在正轨上，而且可以走得更远。还要多谢 Tracy Tseung，感谢她适时的项目管理和编辑协助。

最后，我们要以感恩的心来感谢在 WorldQuant 工作的所有同事的支持和信任，谢谢你们！

声 明

本书的内容仅用于提供信息和教育，不会且不该被视为投资建议。体现出来的观点是多个贡献者的集合，不一定反映 WorldQuant 和 WorldQuant 虚拟实验中心的观点或意见。

关于 WebSim™ 网站

在写这本书的时候，文档里所含的 WebSim™ 的信息与 WebSim™ 网站是对应的。因为网站经常更新，为了防止文档和网站的信息不符，WebSim™ 网站的条款将支配大部分更新的和现有的 WebSim™ 程序。关于 WebSim™ 的最新版和 WebSim™ 所适用的条款，请参阅 <https://websim.worldquantchallenge.com>。

要想获取 WebSim™ 工具的全部功能，或者得到 WebSim™ 支持部门的帮助，需要在 WebSim™ 的官方网站注册。在一定情况下，成功的 Alpha 可能会被 WorldQuant 运营的量化交易投资策略所采纳。

WebSim™ 研究顾问

为了让合格的人士在工作中使用基于网络的模拟平台 WebSim™，WorldQuant 建立了一个研究顾问项目。该项目给予使用者高度自由，他们可以依据自己的身体和智力，去创建他们自己的 Alpha。对于那些正在接受学校教育，或者有志进入金融领域的人，这是一个特别适合的方案。

合格的人士是指那些有很高的量化能力的人，主要来自 STEM（科学、技术、工程或者数学）体系。实际研究内容和专业知识可能会包括统计、金融工程、数学、计算机科学、金融、物理或其他多种 STEM 项目。

所有有关研究顾问项目的信息都可以在 WebSim™官网上找到。

目 录

第一部分 简介

| | |
|--|----|
| 第 1 章 Alpha 设计简介 | 2 |
| 第 2 章 Alpha 的起源——金融价格预测量化模型的生命周期 | 5 |
| 第 3 章 止损 | 10 |

第二部分 设计和评估

| | |
|-------------------------------|----|
| 第 4 章 Alpha 设计 | 16 |
| 第 5 章 如何开发 Alpha: 逻辑实例 | 21 |
| 第 6 章 如何开发 Alpha: 案例研究 | 24 |
| 第 7 章 基本面分析 | 32 |
| 第 8 章 股票价格与交易量 | 36 |
| 第 9 章 换手率 | 38 |
| 第 10 章 回测——信号或者过度拟合 | 41 |
| 第 11 章 Alpha 与风险因子 | 46 |
| 第 12 章 Alpha 与证券组合风险的关系 | 50 |

| | | |
|--------|----------------------|----|
| 第 13 章 | 风险与回撤 | 56 |
| 第 14 章 | 数据与 Alpha 设计 | 62 |
| 第 15 章 | 统计套利、过度拟合与 Alpha 多样性 | 66 |
| 第 16 章 | 提升 Alpha 信号稳定性的技巧 | 69 |
| 第 17 章 | 自动搜寻 Alpha | 72 |
| 第 18 章 | Alpha 研究中的算法和特殊技巧 | 75 |

第三部分 拓展研究

| | | |
|--------|-----------------|-----|
| 第 19 章 | 新闻与社会媒体对股票收益的影响 | 78 |
| 第 20 章 | 由股票期权市场所反映的股票收益 | 83 |
| 第 21 章 | 动量 Alpha 简介 | 90 |
| 第 22 章 | 财务报表分析 | 92 |
| 第 23 章 | 机构研究 101 | 99 |
| 第 24 章 | 期货交易简介 | 112 |
| 第 25 章 | 远期外汇和期货中的 Alpha | 116 |

第四部分 新工具 WebSim™

| | | |
|--------|---------------------|-----|
| 第 26 章 | WebSim™ 简介 | 120 |
| 第 27 章 | Alpha 与 WebSim™ 的原理 | 126 |
| 第 28 章 | 理解 WebSim™ 的工作流程 | 130 |
| 第 29 章 | 用户帮助手册 | 137 |
| 第 30 章 | 输出结果说明与 Alpha 信号库 | 142 |
| 第 31 章 | Alpha 教程 | 151 |
| 第 32 章 | 常见问题解答 | 161 |

第五部分 结语

| | | |
|--------|-----------|-----|
| 第 33 章 | 卓越宽客的七个习惯 | 174 |
|--------|-----------|-----|

表达式和程序：Alpha 通过包含变量的表达式或者程序来体现。这些表达式或者程序相互对应，这就意味着它们之间总是可以相互转化。

2. 如何基于数据开发Alpha信号

这是一个简单的问题：价格行为总是和社会上的某一个事件对应，事件引起数据变化。如果一个数据永远不变，那就没有 Alpha 了。所以，数据的变化包含了事件的信息。信息的变化又会导致 Alpha 的变化。

变化可以用多种方法表示，如表 1.1 所示。

表 1.1 表达式的变化

| | |
|--------------|-------------------------|
| 简单的区别, $A-B$ | 例如用当天的价格减去前一天的价格 |
| 比值, A/B | 例如用当天的价格除以前一天的价格 |
| 表达式 | 例如当天价格的倒数, 当价格低的时候增大持仓量 |

所有 Alpha 设计都在智能地寻找所有可能的变化。表达式需要反映一种逻辑，如表 1.2 所示。

表 1.2 表达式是理论的体现

| 表 达 式 | 理 论 |
|-----------------------|---------------------|
| $1/\text{价格}$ | 价格低买入（逢低买入） |
| 价格变化（与 3 天前的价格相比） | 价格变化跟 3 天内的价格变化趋势一致 |
| 价格 | 高价格股票还会更高（追高） |
| 当天价格和前一天价格的相关系数 | 股票趋势跟随，跑赢大盘 |
| 当天价格和 3 天前价格的比值，乘以交易量 | 股票趋势加大交易量跑赢大盘 |

3. Alpha信号的质量

一个好的 Alpha 信号需要满足如下条件：

- (1) 理论和表达式够简单；
- (2) 表达式或代码优美；
- (3) 有很好的内测夏普率；
- (4) 对数据和参数的微小调整不敏感；

第 1 章

Alpha 设计简介

作者：Igor Tulchinsky

Alpha 是数学表达式、计算机源代码和配置参数的综合体，Alpha 结合历史数据，可以预测多种金融工具的走势；Alpha 也可以对每一种金融证券进行收益预测；Alpha 是一种基于基本面的观点。上面三种关于 Alpha 的理解其实是相通的，我们开发和交易的 Alpha 也是真实存在的。因为即使市场是接近完全有效的，也肯定存在内在驱动力。交易者执行技术面和基本面的 Alpha 信号，从而驱动价格趋向合理。

1. 怎么体现Alpha信号

Alpha 信号是通过持有证券的种类和持仓量的时间序列矩阵来体现的。矩阵的数值对应特定股票在特定日期的持仓量。股票持仓量每天都在变化，这些变化通过在证券市场中进行交易而发生。Alpha 会产生收益，收益也会对应变化。收益和收益标准差的比值是 Alpha 的信息比率。当 Alpha 股票持仓量和预测的股票收益成正比时，Alpha 的信息比率是最大的。

表达式和程序：Alpha 通过包含变量的表达式或者程序来体现。这些表达式或者程序相互对应，这就意味着它们之间总是可以相互转化。

2. 如何基于数据开发Alpha信号

这是一个简单的问题：价格行为总是和社会上的某一个事件对应，事件引起数据变化。如果一个数据永远不变，那就没有 Alpha 了。所以，数据的变化包含了事件的信息。信息的变化又会导致 Alpha 的变化。

变化可以用多种方法表示，如表 1.1 所示。

表 1.1 表达式的变化

| | |
|--------------|-------------------------|
| 简单的区别, $A-B$ | 例如用当天的价格减去前一天的价格 |
| 比值, A/B | 例如用当天的价格除以前一天的价格 |
| 表达式 | 例如当天价格的倒数, 当价格低的时候增大持仓量 |

所有 Alpha 设计都在智能地寻找所有可能的变化。表达式需要反映一种逻辑，如表 1.2 所示。

表 1.2 表达式是理论的体现

| 表 达 式 | 理 论 |
|-----------------------|---------------------|
| $1/\text{价格}$ | 价格低买入（逢低买入） |
| 价格变化（与 3 天前的价格相比） | 价格变化跟 3 天内的价格变化趋势一致 |
| 价格 | 高价格股票还会更高（追高） |
| 当天价格和前一天价格的相关系数 | 股票趋势跟随，跑赢大盘 |
| 当天价格和 3 天前价格的比值，乘以交易量 | 股票趋势加大交易量跑赢大盘 |

3. Alpha信号的质量

一个好的 Alpha 信号需要满足如下条件：

- (1) 理论和表达式够简单；
- (2) 表达式或代码优美；
- (3) 有很好的内测夏普率；
- (4) 对数据和参数的微小调整不敏感；

(5) 适用于多个不同股票池（例如选取 3000 只流动性最好的股票，还是 SP500 成份股）；

(6) 适用于多个区域；

(7) 利润达到新高。

4. 寻找Alpha的算法

永远重复以下步骤：

(1) 查找数据中的变量；

(2) 理解变化，并进行建模；

(3) 建立数学表达式，从而将变化转为股票持仓量；

(4) 测试表达式；

(5) 如果结果不错，则提交 Alpha。

第 2 章

Alpha 的起源——金融价格 预测量化模型的生命周期

作者：Geoffrey Lauprete

Alpha 是预测金融工具价格的模型。在 20 世纪八九十年代，对市场进行建模和预测价格的想法已经比较成熟，同时计算成本也比较低，使以下情况成为现实：

(1) 在华尔街交易柜台进行计算机建模；

(2) 数据的产生和收集保持指数级增长。伴随着电脑和系统数据收集越来越普遍，使用即时数据建模成为博士们向华尔街迁移的驱动力之一。最后，在技术革命和数据暴增的环境下，量化交易产业产生了。

1. 背景知识

当低成本的计算能力在华尔街实现时，量化交易和 Alpha 研究获得了蓬勃发展。Alpha 是量化交易中的输入预测值。另外一种说法是，量化交易是靠

Alpha 来获利的。请记住，Alpha 是预测模型的形式之一，而并不是简单意义上的无风险套利，“统计套利”这个术语也用来描述利用 Alpha 进行的量化交易。

有些人怀疑 Alpha 的存在，关于“Alpha 是否存在”的争论可以视为“有效市场理论”的一部分。很多金融经济学的相关著作都尝试回答这个问题：验证市场和信息流的本质，基于市场、市场参与者和他们的理性级别等多种假设，以及这些参与者怎样相互影响和处理信息推导出结论。

也就是说，从简单而直观的观点来看，市场是复杂的系统，呈现出一定程度的可预测性。虽然这些预测有成功也有失败，但关于“这些预测是不是套利机会”的讨论，每天都在量化行业发生。

2. 遇到的挑战

即使有人可以通过不同形式的假设，得出“Alpha 存在”的结论，在这个真实世界中，预测的细节还是混乱的。从从业者的角度看，低准确率和适用于价格微小变动的预测并没有太大意义。这个市场是人们意愿的合集，受科技更新、宏观经济现实、监管和财富影响，预测比我们所看到的更富有挑战性。所以，为了更好地进行市场建模，我们需要对影响金融工具价格的外在变量有非常深刻的理解。

3. Alpha 的生命周期

市场的一个基本规则是：潜在的盈利策略会吸引关注和资金。因为市场规模是有限的，所以当很多资金使用一个相同的策略或者使用一个特定的 Alpha 时，对应这个机会而形成的固定“馅饼”会被分成更多份，每一份都很小。最终的结果是，如果 Alpha 是从市场参与者的交流中产生的，那么当 Alpha：（1）足够强大；（2）历时足够久；（3）足以连续被验证，并提供建立盈利交易策略的基础时，它们将开始吸引资金。这种资金流会使 Alpha 对应的盈利机会减少并变得不稳定，一直到资金过量使得 Alpha 失效。然而，在这个过程中将产生其他的机会，从而影响市场，使 Alpha 信号得以生生不息。

4. 数据输入

为了预测金融工具的价格波动，Alpha 需要数据。这里的数据可以是价格

本身，也可以是价格的历史记录。在大部分时候，获得更多信息会比仅仅得到价格有用。比如，一只股票的交易量等信息，都可以丰富历史价格和时间序列。

Alpha 的作用如下所示：

数据（如历史价格）→ Alpha → 价格预测

数据的质量对于 Alpha 的结果有很大的影响。所以在数据被使用之前，对数据的质量进行评估并发现其缺点非常重要。如下原因可能影响数据的质量：技术问题，如硬件问题；一些人为的错误，如一些意外的数据格式变化，冗余的数据等。

5. 预测结果

Alpha 模型的输出其实就是一个结果预测。在很多市场中，预测一个金融工具的相对价格比预测其绝对价格要容易。所以，在股票市场中，很多 Alpha 模型预测多只股票和其他相似股票的相对价格。

一般来讲，Alpha 是通过利用 C++、Python 或其他灵活的现代语言进行编程来实现的。在大一点的机构里，机构内部研发的软件环境可以将 Alpha 开发者从很多 book-keeping 和数据管理问题中解放出来，让开发者专心去做创造性的研究和建模工作。

6. Alpha 的评价

什么是好的 Alpha，什么是坏的 Alpha？没有一个单一的标准能够回答这个问题，答案部分依赖于 Alpha 是怎么被使用的。一些投资策略需要很精准的预测，还有一些策略从简单预测中获利。Alpha 评估的关键点如下：

- （1）好的样本内测效果不能保证好的样本外测效果；
- （2）就像学术统计一样，极值可能破坏一个模型并且导致错误的预测。

为了验证一个想法，需要用很多样本内数据和样本外数据去测试。拥有越多的数据，对一个 Alpha 的信心就越足。相反，考虑的时间周期越长，信号失效的可能性就越大，而且市场发生了根本性的变化，以至于 Alpha 在将来失效。所以，Alpha 的开发信心和有效性之间的关系很紧张，必须达到正确的平衡。

7. 回测

在开发 Alpha 时，有机会去及时回看以及评估预测模型历史表现怎么样。并且，用回测方法评估信号是十分有价值的（为看清市场和获悉 Alpha 被使用的情况提供一个窗口）。为此，有如下两个非常重要的事项：

（1）历史绝对不会完全重复。所以，尽管 Alpha 在理论上看起来很好，但是并不能保证它在将来一定会有效。这是因为计算能力和建模者的创造力有限，可能导致一叶障目。使用计算机资源，我们可以评估大量的想法和这些想法的组合。但是如果没有坚持对想法的监控，也没有考虑到评估模型从人工统计到现实模型转换的可能性，我们就会错把煤块当金子。

（2）现在去看历史，会觉得当时的市场比现在更容易交易。这是因为如下几个原因：① 你当时不会有现在的领悟；② 你现在拥有之前无法得到的数据；③ 你当时没有现在这么强大的计算能力和技术。20 世纪 90 年代的那些想法，现在看起来很简单并且很容易在表格中编程，但当时不是这样的，更不要说当时的研究水平了。每个 10 年都有它自己的市场和独特的市场机会。这取决于 Alpha 开发者是否能使用当时的技术和数据寻找到市场机会。反观历史，诸如“过去预测市场和赚钱更容易”这样的想法是错误的。

8. 统计不等于统计套利

统计套利是另外一个用来描述量化投资的术语。这个词引申出以下几点。第一点，预测并不意味着纯粹的和无风险的套利；第二点，统计模型是用来预测价格的。第一点是绝对正确的，但是对于第二点需要做一些补充。当某些来自科学统计的模型，比如时间序列分析、机器学习、回归、优化等可以作为 Alpha 设计的一部分来使用时，我们也需要认识到大多数从学术界而来的技术，初衷并不是用来解决量化投资公司所关注的问题（比如生产现金流和管理现金流的风险）的。所以，当利用标准差等来判断模型是否优劣的时候，只是数据结果间接地和盈利相关而已。

9. 总结

综上，市场中 Alpha 的存在是市场参与者和竞争个体之间的信息不对称所

致。一个具有长远眼光的产业投资者不是很关注价格的短期变化。相反，一个短线交易者不需理解能影响价格波动的基本面因子，而应该去理解短期的动态供需关系。不同的投资者共存，然后辩证地为对方提供参考。除此之外，市场中每个投资者的行为相互糅合并形成一些模式。Alpha研究者的目的是在不同的制约条件下（比如使用不同的数据和建模工具）发现这些模式，判断哪些和预测未来价格相关。当Alpha的特性随着时间变化时，如何去发现哪些想法和现在相关，同时以高效、优雅和可靠的方法去实现，是Alpha研究者面临的挑战。

第 3 章

止 损

作者：Igor Tulchinsky

人之所以是最成功的物种，在于人可以适应周边的环境，并且能比其他的生物更高效地利用规则。自农耕时代以来，对规则的理解伴随着人类的进化在发展。现在，各个领域都有自己的规则。从财务到锻炼，到社交关系和自我修养，规则都在发挥作用。人们依靠规则生活。

现实世界由无数的规则组成，人类总是下意识地争取和努力，以期发现更多规则。当然，此时此刻，也有唯一的一种规则一直在统治着人们。这个规则就是：没有规则是完美的。我们可以把它叫作“母规则”。

事实上，没有规则可以被证“实”，它们只能被证“伪”，这是一个科学原理。这个原理是伟大的哲学家 Karl Popper 在 1934 年提出的。他说验证普遍真理是不可能的，相反，仅仅一个反例就可以推翻它。Popper 强调，因为纯粹的事实不存在，所以所有的发现和规则都是主观的、基于理论的。

有一个很好的理由可以阐述为什么规则不能被证实。现实是复杂的，人

们和他们的想法总是不完美的。想法通过语言和符号来表达，规则是达到现实的隐喻方式。所以，所有的规则都有局限性，没有会一直适用的规则。没有单独的教条可以形容或者统治这个世界。但是所有的规则都可以描述世界的一部分，有时，所有的规则都有效。

我们就像在画布上画画的艺术师一样，每一次挥洒都可以使图片更靠近现实，但是永远不可能成为现实，而且也永远不可能成为现实的完美解释。然而，虽然永远不会抵达，但我们离它越来越近。例如，几个世纪以来，看上去牛顿定律完美地解释了力学，但结果证明它是有瑕疵的，相对论弥补了其不足。人们对于自我的认识和对于自己在宇宙中位置的理解一直在持续发展，从认为地球是宇宙万物的中心，到意识到我们只不过是那不可能触及的空间的一粒尘埃。同样，在经济学领域，已被定义的各种规则，包括期权模型，也只会证明是错误的，各种假设终会崩塌。

有这样一个悖论：只有一个理论一直适用，就是没有理论一直适用。规则永远是空气中的尘埃，指出朝向现实的方向，但是永远无法到达现实。

这就是为什么当行驶在母规则变化的海洋中时，我们要最大化地减少损失的真正原因。怎么去掌握？能处理数以百万计的、一直在变化的规则的正确方法是什么？基于不同的环境和假设，它们都是不完美的，有时还是相互矛盾的。

交易是现实的微观缩影，百万个规则和思想汇聚成百万个个体，形成一个动态的极其复杂的系统，系统又反过来被个体所影响。交易的难点在于人们很难领悟市场内涵的规则，然后成功利用规则去获取利润，而不是通过摧毁规则去改变市场。

交易中的规则被称为 Alpha，Alpha 是用来预测证券未来收益的算法。从本质上看，管理百万个规则，每个规则都是假设，是很复杂的事情。当在真实交易中应对这些规则的时候，就可以看到许多显现的规律。这又引导我们反观止损行为。最好的也是最通用的处理这种复杂事情（所有的规则都会被破坏）的方法，就是去掌握隐含在止损行为中的长期规则。

在交易的世界中，止损的概念已经存在一段时间了。它发源于也许是已

知的最早的交易：跟随趋势（相信涨的证券还会一直涨）。在这种操作模式下，交易者会在一个新的高点入场，然后当经过历史最高点、损失的利润总额超过一个固定数值的时候退出。

在今天的交易世界中，策略和 Alpha 不再像以前一样单一。不再去跟随一只特定的证券，而是跟随趋势看重策略的整体 P&L。

如果用朴素语言和通用的词汇来描述“止损”，可以这样说：止损即不再采用失效的规则。

止损不仅在交易的微观世界中很容易看到，它的原则适用于生活的各个角落，比如商业、娱乐和社交。

止损需要纪律和自律。通常，所有的思维和决策，包括跟随规则，都可能被心烦意乱的情绪所干扰。脑科学家发现一个脑受损的人很难做简单决定，比如早晨应该穿什么衬衫。在我们使用 Alpha 做交易决定的时候，我们心里面一般都被“自信”这种情绪所驱使：“我理解这个世界是怎么工作的，我相信我的规则，这是我的规则。”因为自尊和骄傲也同时存在于这种自信中，所以一旦产生，即使在截然相反的事实面前，也很难摆脱这种情绪。

也许是因为人性中自尊的原因，止损的原则不能被广泛使用。另外一个原因可能是替代规则的缺乏。切换策略的高成本，对于取消现有的失效规则也是一种阻力。

最好不要让自己迷信于单一的理论或者规则，而要去看全局，不要相信它们的一部分。单一规则总是有时好用，有时不好用。

判断一个规则好坏的标准是看它现在是否运行良好（获利），除此之外的都是投机活动。如果一个规则有效，我们就投资，如果无效，我们就放弃投资。

我们收集所有的策略，然后让时间和结果去验证在什么时候，哪些有效，哪些无效。

我们根据历史提出一个新的想法、规则或者 Alpha 策略。然后经过统计分析（有时也有一点基本的主观思考），进行判断，从而将其放进我们的知识库。

经过广泛的思考，我们对事物的了解越发接近事物本身。然而你必须接

受现实：你永远不可能知道所有的事情。

有种说法：在盲人的世界中，有一只眼睛的人是国王。

反过来回到交易和金融市场，即使有一只眼睛也会很有成就。

如何应用母规则和止损原则？

我们承认，不完美想法的数量是无穷的，现实是未知且不可知的。每个不完美的想法都能成功描述一部分现实。所以我们有越多的 Alpha，就能越好地描述现实的一个方面，就能离“睁一只眼”的获利状态更近。

既然没有规则是完美的，那么一个所有规则的综合体是我们所能做的最贴近现实的方法。

同时应用所有的规则是成功的关键。比如，为了穿过街道，你的脑海中必须有如下规则：

- (1) 向左看，向右看，然后再往左看，接着安全通过；
- (2) 如果听到比较响的声音，则向发出声音的方向看；
- (3) 如果看到朝你开过来的车，则快跑！

也许你确信街道处于第(1)种情况，然后已经开始放心地穿越街道，但是在听到喇叭声的时候，就触发了第(2)种情况。(1)对应的处理方式应该立即停止，因为噪声已经打破了安全的结论。然后就要按照第(3)种情况去处理了。

所以我们受到如下启发：

- (1) 很有必要在刚开始的时候，尽量多地制定好的规则；
- (2) 永远不可以完全相信单一的规则；
- (3) 要有同时使用多种规则的策略。

如何判断一个策略是不是有效？如果它在预期历史收益之外，则会表现为如下情况：

- (1) 回撤超过正常值；

(2) 夏普比率降低;

(3) 呈现出历史数据范围之外的结果, 否定了我们在刚开始的时候观察到的规则。

同时进行几个不同的策略并将一些即时成果及时糅进策略很重要。例如, 我们有个很好的理论可以解释金价什么时候上涨, 该理论在一年内 50% 的时间里有效, 假定一个人有 10 个这样的理论, 将这些理论综合起来肯定比其中任何单独一个都能更好地描述现实。并且发现其中最准确策略的方法, 就是去看现在哪些是有效的。

接下来就有了止损的应用。当一个策略失效的时候, 要下定决心去采取止损的行动。如果没有这种决心, 那么你就是掷骰子碰运气。那么你最好终止这个行为, 然后去做其他能有更多产出的事情。

例如, 你雇用了一些人去重盖你的房子, 他们保证只花 5 万美元, 但是干了不到一半, 他们已经花了 4.5 万美元。此时, 换掉他们, 改用另外一个建筑商会更便宜。

假定我们正从事一个叫 X 的活动, 这个活动将会让你蒙受经济损失。

这个活动可以是任意事情, 比如交易策略或者商业活动。然后你做下面的思考:

(1) 我会在 X 活动中损失金钱吗?

(2) 会损失多少钱?

在我们开始 X 活动之前, 估计最大损失资金是多少。把损失称为 Z, 如果 Z 大于之前预期的最大损失, 并且损失的金额不多, 那就止损。

综上, 在真正做决策之前, 要考虑到每一种可能的情况。要确定:

(1) 目标是什么?

(2) 在正常情况下, 预期的困难是什么?

应提前做好计划, 考虑怎么才能低成本地改变策略, 同时执行多种策略, 停止所有低于预期值的策略。

第二部分

设计和评估

- 第 4 章 Alpha 设计
- 第 5 章 如何开发 Alpha: 逻辑实例
- 第 6 章 如何开发 Alpha: 案例研究
- 第 7 章 基本面分析
- 第 8 章 股票价格与交易量
- 第 9 章 换手率
- 第 10 章 回测——信号或者过度拟合
- 第 11 章 Alpha 与风险因子
- 第 12 章 Alpha 与证券组合风险的关系
- 第 13 章 风险与回撤
- 第 14 章 数据与 Alpha 设计
- 第 15 章 统计套利、过度拟合与 Alpha 多样性
- 第 16 章 提升 Alpha 信号稳定性的技巧
- 第 17 章 自动搜寻 Alpha
- 第 18 章 Alpha 研究中的算法和特殊技巧

第 4 章

Alpha 设计

作者：Scott Bender/Yongfeng He

Alpha 是一种对资产未来价格变化进行预测的方法。例如，Alpha 可能是用来预测某一只特定股票未来收益的计算机程序。

我们在本章中所涉及的 Alpha 是成体系的，而且可以用一系列具体的代码来表达。Alpha 通常以固定频率周期性地做预测，如每天一次。Alpha 的预测是通过一系列预测而持有的资产来体现的。

一个简单的 Alpha 信号，可能每天对前一天涨的所有股票给以-1 的预测，对前一天跌的所有股票给以+1 的预测。这对我们来说是一个有效的 Alpha，因为它系统性地以固定的频率对一系列的资产做了具体的预测。

Alpha 是用来预测的模型，但是如果缺少了执行，人们就没有办法获得预测产生的潜在收益。一般来讲，Alpha 是用来实战的交易策略的组成部分。交易策略在很大程度上被其中的 Alpha 信号所驱动，但是也会考虑其他的实际因素，比如交易成本和在实际执行一个交易时的策略风险。

1. Alpha信号的分类

根据交易金融工具（股票、ETF、期货、期权和债券等）的不同，Alpha 可以被分为三大类。Alpha 可以交易单一工具或多种工具。交易范围可以是在一个国家内，也可以是在多个国家内，甚至可以是在全球范围内。

根据产生 Alpha 的时间及预测产生的频率，我们可以将 Alpha 进行如下的分类。

(1) 日内 Alpha: 在当天交易时段不断进行再平衡。可以对日内 Alpha 进行如下分组。

- ① 每个时间间隔都进行再平衡，例如 1 分钟/ 5 分钟/ 15 分钟等；
- ② 被事件（像 ticks/orders/fills 或者之前设定好的事件）触发再平衡。

(2) 每日 Alpha: 每天进行再平衡。这种类型的 Alpha 可以根据使用信息的时间分为如下 3 组。

- ① 延迟 N ，适用 N 天前的数据；
- ② 延迟 0（快照），适用一个固定时间之前的快照数据；
- ③ MOO（market on open）/MOC（make on close），在市场开市或闭市节点交易。

(3) 周或者月 Alpha: 每周或者每月进行再平衡。

2. 开发Alpha信号

开发 Alpha 的数据来源为公共信息。得到数据的流程越高效，Alpha 能达到的效果就越好。可以通过搜索公共信息或者建立特定模型去处理信息，从而找到 Alpha。在信息库里查找信息或者模型的时候会产生 Alpha。典型的来源如下。

(1) 价格/交易量：可以使用基于价格/交易量的技术分析或预测/回归模型。

(2) 基本面：通过自动分析每个公司的基本面，可以建立基本面 Alpha。这种 Alpha 一般有很低的换手率。

(3) 宏观数据：比如国内生产总值、就业率，这些数据对金融市场有很大的影响。

(4) 大环境：比如美联储（美联储公开市场委员会）会议、公司文件、论文、期刊、新闻或者其他公开社交媒体的信息。将这些大环境信息进行量化（最终体现为买卖数据）非常有必要。环境数据包括现在和将来的事件。

(5) 音频和视频等多媒体也可以用来作为信息源。处理音、视频的技术已经非常成熟。比如，我们可以很方便地使用语音转为文字的技术从音、视频里面抽取文字，然后用文字信息进行建模。

有时，不能直接从信息里面提炼出 Alpha。信息用来提高 Alpha 性能或者产生 Alpha 的效率。举例如下。

(1) 风险模型：利用控制或者降低因子的风险敞口，去改善 Alpha 的效果。

(2) 关系型模型：一般来讲金融标的在固定范围内相互关联。有些金融标的会对其他金融标的产生引导或者跟随行为，这就为套利创造了机会。

(3) 微结构模型：改善了实际交易的执行效率。

目前，伴随着信息的海量增长，从很多冗余信息中淬炼信号变得越来越难。解决方案也都是非凸性、非连续、动态的。好信号往往从人们不注意的地方产生。怎么去提取这些信号？通过圈定寻找范围，采用之前其他人尝试过的方法，比如在之前发现信号附近的地方寻找，保护资源避免挖得太深，同时通过验证过的线索去提高发现新信号的可能性；此外，总是预留一些资源去验证疯狂的想法。

3. Alpha 的价值

Alpha 的价值的最终测试是看在用它进行交易的时候，能给策略增加多少风险调整后的收益。然而，在实际中，这很难精确地计算出来，因为：

(1) 在设计阶段，没有明确的规则指明 Alpha 在哪里使用，以及 Alpha 用在哪些具体策略上。

(2) 即使确定风险厌恶的等级和具体的策略，鉴于在策略组合里面盈利

可能是非线性的，也很难去判断有多少盈利归功于特定的 Alpha。

(3) 基于以上讨论，我们仍然可以：

- ① 确定 Alpha 在策略中是否增加收益的有用预测；
- ② 合理估计 Alpha 在整体策略基础上增加的收益。

4. 实用的Alpha评估方法

由于事先不知道实际使用的交易策略，如果单纯考虑一个 Alpha 信号，那么如何判断信号好坏？或者说，当我们对 Alpha 做任何改动的时候，怎么知道是不是改好了？为了回答这些问题，我们需要一些方法帮助预测 Alpha 信号是否为具体的策略增加了收益。

用来收集交易策略信息的典型方法之一是先模拟（比如回测），然后量化结果（比如信息比率）。对 Alpha 进行模拟测量的方法，就是把 Alpha 的预测和交易策略做配对，然后假定 Alpha 预测的就是在策略中所持有特定资产的仓位。同样，这个策略的交易就是 Alpha 预测的变化。这个方法有个问题，即 Alpha 经常找不到匹配的策略，因为 Alpha 是用来预测收益的，而不是为了交易本身。在这种模拟下，成本会比较低，甚至没有成本。

一旦我们构建了一个如下所述的模拟，就可以有如下的计量。

(1) 信息比率

Alpha 收益和收益标准差的比值。通常：

- 粗略计算怎么持续让 Alpha 做好的预测。
- 通过信息比率结合观察的时间窗口可以判断 Alpha 信号是不是随机的噪声。

(2) 利润

在模拟中，Alpha 创造的利润可以根据交易规模做如下划分：

- 粗略计算 Alpha 对于交易成本是不是敏感。高利润意味着 Alpha 受交易成本影响不大。
- 低利润的 Alpha 不能作为策略使用。低利润的 Alpha 不会创造价值，

除非它们和策略中的其他 Alpha 大不相同。

(3) 独特性

在 Alpha 信号池中，一个信号和其他信号的最大相关性定义为独特性：相关性越低，意味着 Alpha 信号越有价值。

越来越多的复杂测试被开发出来。比如，检测 Alpha 在流动性好（交易量比较大）和流动性差的股票中是否都有很好的信息比率。如果 Alpha 只是在流动性差的股票中适用，那么它在某些交易规模比较大的策略里意义不大。

5. 未来的表现

在预测环节，所有的方法都是用来比较两个 Alpha 信号的，除它们预测的实际结果外，并没有其他信息。但是，有些补充信息（比如 Alpha 是怎么形成的）能为判断 Alpha 信号是否有好的预测结果提供有用信息。最后，也是最重要的一点——Alpha 是否能做出可靠的预测。

假设我们有一个高信息比率的 Alpha 信号，但是它基于小概率经济事件，或基于历史数据的已知规则的参数优化。比如，假设 Alpha 有 12 个参数，每个参数对应单独的一个月 (X_1, X_2, \dots, X_{12})，假定在一月份只是简单地以 X_1 元购买所有股票，在二月份买入 X_2 元的所有股票，依此类推。如果我们基于过去几年，去优化 $X_1 \sim X_{12}$ 的数据，就能取得很好的数据，但是我们不知道怎么去优化，从而在未来达到好的效果。

一般来讲，通过观察历史数据去优化或者改进的 Alpha 信号，都会在基于历史数据的回测中有很好的效果，但是在预测未来的时候，就不会有这样明显的改善了。应该格外注意，Alpha 设计者优化 Alpha 信号的目的是为了预测将来！

当 Alpha 的改善对基于历史数据的回测有很大的改善，但对未来预测的效果只有很小甚至反向的改善时，我们称 Alpha 对于历史数据过度拟合。Alpha 设计者可以去测量过度拟合的影响，通过观察 Alpha 在没有使用过的数据（样本外测数据）时的表现，与改良后的 Alpha 使用样本内测数据对应的结果做对比。样本内测数据和样本外测数据的对比不仅对于 Alpha 本身有意义，而且对于一个信号设计者的一系列信号，甚至对一个设计者的所有信号都有意义。这种对于 Alpha 信号组的比较可以用来计算设计者过度拟合的趋势。

第 5 章

如何开发 Alpha：逻辑实例

作者：Pankaj Bakliwal

Alpha 是用来预测各种金融工具未来价格运动趋势的数学算法。

形成 Alpha 逻辑→信息数据化→提出设想→得到数学表达式→应用运算→得出最终稳健的 Alpha 模型→转换成金融工具头寸→检验历史损益,以及其他性能度量(信息比率、换手率、回撤等)。

我们的目标是利用适当的信息开发数学预测公式,在将风险最小化的同时将利润最大化。为了将数学公式转换为股票头寸,公式给出三种类型的数值意义如下:

正值,表示买入持有,多头;负值,表示卖空,空头;0,表示空仓。

公式给出的每只股票的数值的大小正比于该只股票的持仓金额。

示例如下。

1. 收集信息

考虑科技行业的两只股票：Google 和 Apple。我们可以很容易地获得两只股票的每日历史价格，同时，假设我们可以获得即将发生的事件信息，这些事件可能给这些科技企业带来正面或者负面的影响，但方向性并不明确。

2. 提出设想

一旦我们获得了可以获得的所有信息，下一步则是提出一个合理的设想。假设基于历史价格，我们观察到两只股票的价格在上周都呈现上涨趋势。那么从逻辑上推测，如果没有更多的信息，当股价上涨时，持有股票的投资者会获利了结，卖出股票，造成股价下跌；当股价下跌时，投资者看到一个价格比较低的购买股票的机会，会购买股票从而推动股价上涨。

3. 转换成数学表达式

虽然并不是所有设想都可以直接转换成数学表达式，但是可以直接把上面的例子表述成如下公式：

$$\text{Alpha} = - (1 \text{ 周收益}) \quad (1)$$

负号代表当股价上涨时持空头仓位，当股价下跌时持多头仓位。

持仓金额由公式给出的数值的大小决定。上面的公式表明，价格趋势越强，价格反向移动的概率越大。

我们假设由公式（1）得出的 Google 和 Apple 的 Alpha 值分别是+2.5 和 +7.5。

4. 应用运算来改进原始表达式

原始的 Alpha 公式已经得到，但我们是否可以让它更加稳定可靠？尤其是，假设有一个即将发生的事件会从正面或者从负面影响股票价格，我们可以利用此信息使 Alpha 模型更加盈利吗？如果通过 Alpha 公式得出的结论是买入两只股票，而事实上事件对科技行业产生不利影响，从而会造成股价急剧下跌，那么我们将面临巨额亏损。

避免损失的方式之一是开发行业中性化策略,这种策略使得不管发生正面的还是负面的事件,资产都没有很大的下跌风险。行业中性化策略是指投入一个行业的头寸加总为0。即行业中所有股票的持仓金额加总为0。例如在上面的例子里,这意味着我们持有同等金额的两只股票,但却是相反的持仓方向,也就是说对一只股票持多头,对另一只股票持空头。

5. 得出最终稳健的Alpha

最终的 Alpha 将结合所有信息,将上面 Google 和 Apple 的 Alpha 值由+2.5 和+7.5 改为-5.0 和+5.0。

6. 转换为金融工具的头寸

我们可以简单地应用下面的公式获得最终的仓位:

某只股票的仓位 = (某只股票的 Alpha ÷ 所有股票的 Alpha 之和) × 股票规模

例如,如果我们有 1000 万美元,那么将建立 500 万美元的 Apple 多头仓位和 500 万美元的 Google 空头仓位。

7. 检验稳健性

稳健性指标有如下 8 个:

- (1) 样本内测信息比率高;
- (2) 样本外测信息比率高;
- (3) 适应大多数的交易空间;
- (4) 调整少;
- (5) 设想直观、有趣、简单;
- (6) 在多个区域有效;
- (7) 回撤小;
- (8) 回撤期短。

第 6 章

如何开发 Alpha：案例研究

作者：Hongzhi Chen

Alpha 有多种定义，最广泛使用的定义是：用于预测金融工具未来价格走势的计算机算法。另一个经常使用的定义是 Alpha 值，用于给每个金融工具赋值，表示投入的资金比例。一般来说，我们说的 Alpha 是指第一种定义。当我们说 Alpha 值的时候，指的是后一种定义。

我们先研究一个简单的例子，从而可以更好地理解 Alpha。

假设我们有 100 万美元资金，投入一个由两只股票组成的证券组合：Google (GOOG) 和 Apple (AAPL)。我们需要知道如何将资金分配到两只股票上。如果每天都对证券组合进行仓位调整，其依据肯定是未来几天每只股票的投资收益。我们该如何预测？

有很多因素会影响股票价格，比如交易者的行为、新闻、基本面的变化、内部人士的行为等。我们将预测过程分解成两个步骤：第一步，用单因素预测股价的变化，如仅靠新闻来预测；第二步，汇总所有的预测。下面进行具

体分析。

我们需要一个数据收集器来收集新闻，然后，我们找出将文本新闻转换成向量的算法，这个向量的值即我们要投资到每只股票的资金数值。如果我们认为某只股票的价格即将上涨，就多投一些，反之，则多卖空。用来做这种预测的计算机算法就是我们所说的 Alpha。

在前面的例子里，假设用我们的算法得出如下结果：

$$\text{Alpha}(\text{GOOG}) = 2$$

$$\text{Alpha}(\text{AAPL}) = -1$$

上面的值表示的是 2 : -1，即我们买 2 份 GOOG 股票，卖空 1 份 AAPL 股票，绝对值表示投资金额，负号表示卖空。假设有 100 万美元的资金，那么在当天收盘的时候，我们的仓位是持有 100 万美元的 GOOG，同时卖空 50 万美元的 AAPL。这里假设交易成本为零。

可见，Alpha 模型是将输入数据（如价格、交易量、新闻、基本面等）转换成向量的算法，向量的值就是我们投入到每个金融工具的资金比率。

$$\text{Alpha}(\text{输入数据}) \rightarrow \text{Alpha 向量值}$$

现在我们明白了 Alpha 的含义，那么下面来设计第一个 Alpha 算法。

我们将引入更多的概念。

首先，我们要定义投资空间，即定义一个包含若干金融工具的集合。在美国的股票市场，有很多方式可以用来定义投资空间，例如用标准普尔 500 指数的成份股、选择流通性最好的 3000 只股票等。假设我们用美国流通性最好的 3000 只股票作为研究空间，称其为 TOP 3000。

我们需要一个预测股价的思路。行为金融学发现，短期来看，由于交易人员的过度反应，股价有回调的趋势。我们怎么将这个概念转变成模型？有很多表达方式，为了便于演示，我们选择一个相对简单的公式：

$$\text{Alpha}_1 = (\text{当天收盘价} - 5 \text{天前收盘价}) \div 5 \text{天前收盘价} \quad (1)$$

公式(1)假设股票会回调到5天前的价格。如果今天的价格低于5天前的价格,我们就买入,反之则卖出。我们用5天内的收益作为持有仓位的资金量。也就是说,收益越高,回调越大。

现在得到第一个简单的 Alpha 公式。

为了测试这个设想是否有效,我们需要进行模拟回测,可以通过 WebSim™ 模拟网站进行测试。

通过 WebSim™,公式(1)得出的测试结果如图 6.1 所示。

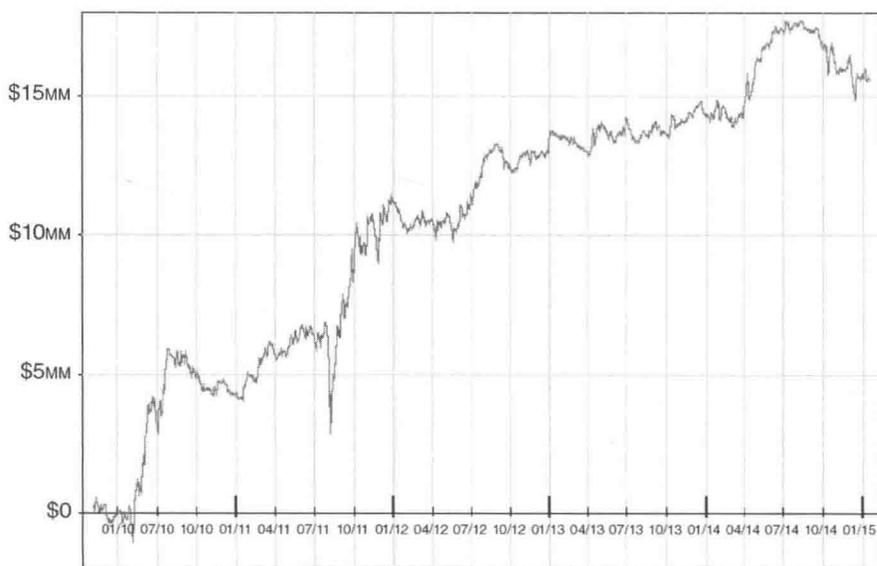


图 6.1 通过 WebSim™ 得出的 Alpha1 的样本模拟结果

表 6.1 中列出了一些在后面会详细阐述的新鲜的概念。我们罗列了一些评估 Alpha 算法时所需要用到的最重要的概念,这些会在本章后面进行讲解。

表 6.1 Alpha1 模拟结果 (仅供说明用途)

| 年份 | 仓位规模 | 损益 | 年化收益率 | 信息比率 | 最大回撤 | 盈利比 | 日换手率 | 每美元交易利润 |
|--------|----------|---------|--------|------|--------|--------|--------|---------|
| 2010 年 | 2000 万美元 | 427 万美元 | 46.44% | 1.32 | 16.63% | 46.52% | 62.90% | 0.15 美分 |
| 2011 年 | 2000 万美元 | 693 万美元 | 68.70% | 1.42 | 39.22% | 50.79% | 64.72% | 0.21 美分 |

续表

| 年份 | 仓位规模 | 损益 | 年化收益率 | 信息比率 | 最大回撤 | 盈利比 | 日换手率 | 每美元交易利润 |
|------------|---------|-----------|---------|-------|--------|--------|--------|---------|
| 2012年 | 2000万美元 | 201万美元 | 20.08% | 0.96 | 14.66% | 51.20% | 63.36% | 0.06美分 |
| 2013年 | 2000万美元 | 104万美元 | 10.34% | 0.6 | 9.22% | 46.83% | 63.26% | 0.03美分 |
| 2014年 | 2000万美元 | 148万美元 | 14.72% | 0.61 | 28.67% | 51.19% | 62.36% | 0.05美分 |
| 2015年 | 2000万美元 | -15.82万美元 | -32.96% | -1.38 | 4.65% | 41.67% | 64.34% | -0.10美分 |
| 2010—2015年 | 2000万美元 | 155.7万美元 | 31.20% | 1.00 | 39.22% | 49.28% | 63.30% | 0.10美分 |

回测时间从2010年到2015年,表6.1中的每一行对应每年的业绩,模拟仓位规模固定不变,为2000万美元,损益为年度损益。

年化收益率公式如下:

$$\text{年化收益率} = \text{年度损益} \div (\text{仓位规模} \div 2) \quad (2)$$

公式(2)计算了Alpha的盈利能力。信息比率是最重要的指标,公式如下:

$$\text{信息比率} = \text{平均每日收益} \div (\text{每日波动率}) \times \sqrt{256} \quad (3)$$

公式(3)计算了Alpha算法的信息比率,信息比率能够大致地反映Alpha盈利的稳定性,越高越好;最大回撤是指从损益最高点到最低点所对应的损失除以1/2仓位规模的比率;盈利比是指盈利天数在一年中的比率;每日换手率是指调整证券组合的频率,公式如下:

$$\text{每日换手率} = (\text{每天的平均交易量}) \div \text{仓位规模} \quad (4)$$

每单位美元的利润衡量的是每一美元的交易创造了多少利润,公式如下:

$$\text{每单位美元的交易利润} = \text{损益} \div \text{总的交易量} \quad (5)$$

在上面的Alpha算法中,总的信息比率是1,收益率大概为30%,最大的回撤为39%,这意味着风险很高,损益不是很稳定。为了减少最大回撤,我们需要降低一些风险,可以通过风险对冲手段实现。行业风险和市场风险是股票市场中最大的风险,我们可以通过在各个行业中同时拥有多头和空头仓位来规避一部分风险。

根据上面的要求，我们调整算法，增加限制条件：

$$\text{Alpha2} = \text{Alpha1}$$

Alpha2 在同一行业的仓位和=0

通过上面的限制，我们得到图 6.2 所示的结果。

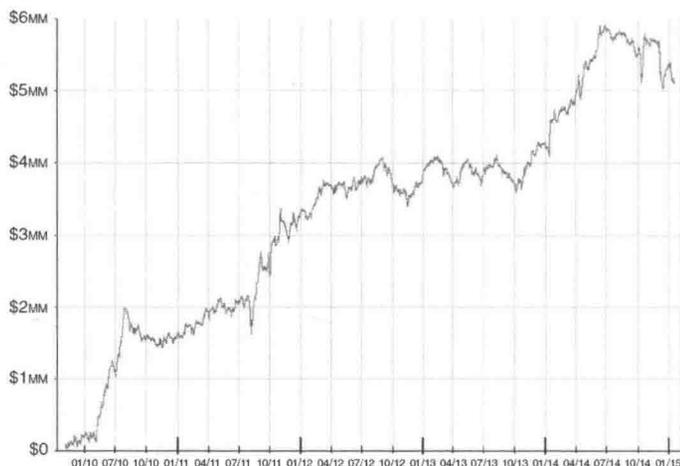


图 6.2 WebSim™ 得出的 Alpha2 的样本模拟结果

从表 6.2 可以看到，信息比率上升到了 1.4，收益率下降到了 10%，但最大回撤下降很明显，仅 9%，有了很大的改进。

表 6.2 Alpha2 模拟结果

| 年份 | 仓位规模 | 损益 | 年化收益率 | 信息比率 | 最大回撤 | 盈利比 | 日换手率 | 每美元交易利润 |
|-------------|----------|------------|---------|-------|-------|--------|--------|----------|
| 2010 年 | 2000 万美元 | 159 万美元 | 17.30% | 2.44 | 5.44% | 51.30% | 63.73% | 0.05 美分 |
| 2011 年 | 2000 万美元 | 166 万美元 | 16.50% | 1.81 | 5.27% | 49.21% | 63.85% | 0.05 美分 |
| 2012 年 | 2000 万美元 | 51.824 万美元 | 5.18% | 0.90 | 6.66% | 55.20% | 63.12% | 0.02 美分 |
| 2013 年 | 2000 万美元 | 45.088 万美元 | 4.47% | 0.80 | 4.97% | 51.59% | 62.99% | 0.01 美分 |
| 2014 年 | 2000 万美元 | 111 万美元 | 11.02% | 1.24 | 8.73% | 53.17% | 62.86% | 0.04 美分 |
| 2015 年 | 2000 万美元 | -23.14 万美元 | -48.21% | -5.96 | 2.88% | 33.33% | 62.30% | -0.15 美分 |
| 2010—2015 年 | 2000 万美元 | 510 万美元 | 10.22% | 1.37 | 8.73% | 51.92% | 63.29% | 0.03 美分 |

为了进一步改善 Alpha 算法，我们之前用的是 5 天的收益，这个指标并不

能准确地进行预测。现在用相对大小作为预测指标。我们引入了等级的概念。等级为相对的 Alpha 值。

$$\text{Alpha}_3 = \text{等级}(\text{Alpha}_1)$$

Alpha₃ 的值在同一行业的和=0

结果见图 6.3。

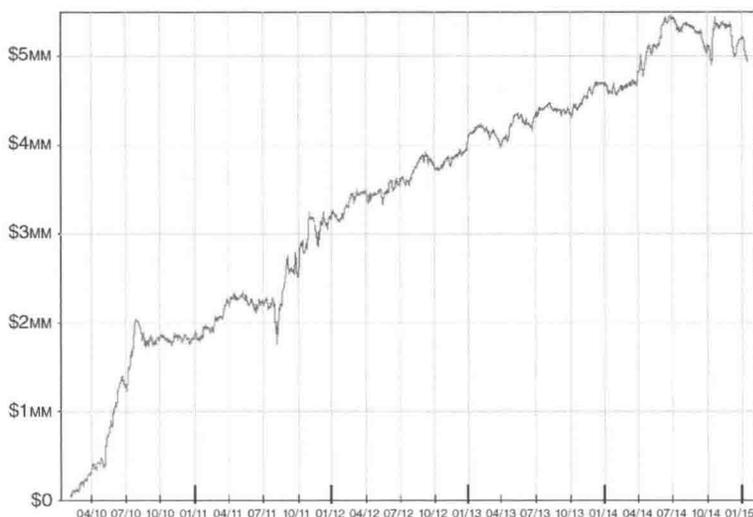


图 6.3 WebSim™ 得出的 Alpha₃ 的样本模拟结果

从表 6.3 可以看出，算法得到了极大的优化。虽然换手率还是有一些高，但是效果已经好多了。我们可以通过衰减运算来降低换手率，即对 Alpha 取某一时间窗口的平均值。

表 6.3 Alpha₃ 模拟结果

| 年份 | 仓位规模 | 损益 | 年化收益率 | 信息比率 | 最大回撤 | 盈利比 | 日换手率 | 每美元交易利润 |
|-------|----------|------------|--------|------|-------|--------|--------|---------|
| 2010年 | 2000 万美元 | 183 万美元 | 19.94% | 3.43 | 3.11% | 56.52% | 59.43% | 0.07 美分 |
| 2011年 | 2000 万美元 | 134 万美元 | 13.30% | 1.70 | 5.82% | 53.17% | 59.49% | 0.04 美分 |
| 2012年 | 2000 万美元 | 80.174 万美元 | 8.02% | 1.89 | 1.93% | 55.20% | 58.94% | 0.03 美分 |
| 2013年 | 2000 万美元 | 69.273 万美元 | 6.87% | 1.94 | 2.49% | 53.57% | 58.69% | 0.02 美分 |
| 2014年 | 2000 万美元 | 51.806 万美元 | 5.14% | 0.93 | 5.43% | 52.38% | 59.20% | 0.02 美分 |

续表

| 年份 | 仓位规模 | 损益 | 年化收益率 | 信息比率 | 最大回撤 | 盈利比 | 日换手率 | 每美元交易利润 |
|-------------|----------|------------|---------|--------|-------|--------|--------|----------|
| 2015 年 | 2000 万美元 | -25.14 万美元 | -52.37% | -10.45 | 2.78% | 33.33% | 59.59% | -0.18 美分 |
| 2010—2015 年 | 2000 万美元 | 496 万美元 | 9.89% | 1.76 | 5.82% | 53.93% | 59.15% | 0.03 美分 |

新 Alpha = Alpha + 前 Alpha 加权平均值

引入三天的衰减，经过 WebSim™ 模拟，我们得到图 6.4。

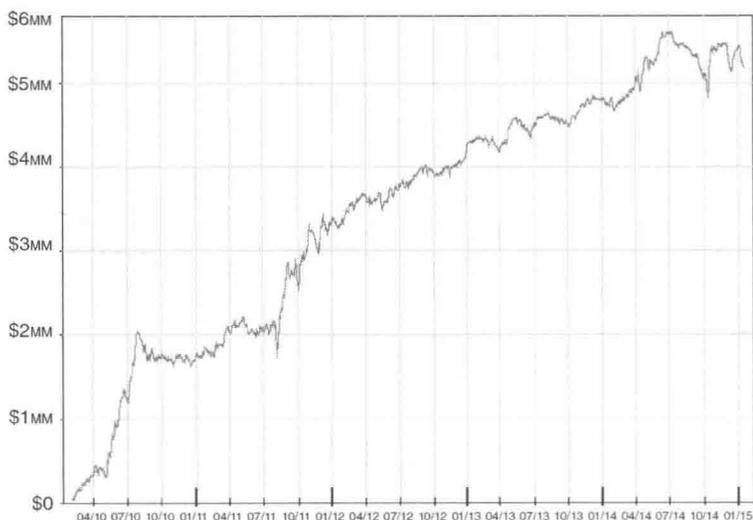


图 6.4 WebSim™ 得出的新 Alpha 的样本模拟结果

从表 6.4 可以看出，结果很好。不仅每日换手率下降了，而且信息比率、收益、最大回撤都得到了改善。

表 6.4 新 Alpha 模拟结果

| 年份 | 仓位规模 | 损益 | 年化收益率 | 信息比率 | 最大回撤 | 盈利天数占比 | 日换手率 | 每美元交易利润 |
|--------|----------|------------|--------|------|-------|--------|--------|---------|
| 2010 年 | 2000 万美元 | 172 万美元 | 18.66% | 3.09 | 4.11% | 53.91% | 42.48% | 0.09 美分 |
| 2011 年 | 2000 万美元 | 161 万美元 | 15.94% | 2.01 | 4.87% | 51.19% | 42.28% | 0.08 美分 |
| 2012 年 | 2000 万美元 | 81.403 万美元 | 8.14% | 1.90 | 2.05% | 57.20% | 42.09% | 0.04 美分 |
| 2013 年 | 2000 万美元 | 64.329 万美元 | 6.38% | 1.88 | 2.48% | 54.76% | 41.87% | 0.03 美分 |

续表

| 年份 | 仓位规模 | 损益 | 年化收益率 | 信息比率 | 最大回撤 | 盈利天数占比 | 日换手率 | 每美元交易利润 |
|------------|---------|------------|---------|-------|-------|--------|--------|---------|
| 2014年 | 2000万美元 | 59.921万美元 | 5.94% | 1.03 | 7.74% | 51.59% | 42.09% | 0.03美分 |
| 2015年 | 2000万美元 | -19.434万美元 | -40.49% | -7.20 | 2.58% | 33.33% | 41.82% | -0.19美分 |
| 2010—2015年 | 2000万美元 | 519万美元 | 10.39% | 1.82 | 7.74% | 53.53% | 42.15% | 0.05美分 |

现在我们知道了如何将一个 Alpha 的策略转换成算法,以及优化算法的方法。你可以通过更多的方式进一步优化算法。

这就是 Alpha 研究的基本方法^①。

接下来要做的就是找出一些独特的策略和数据,独特的策略能够让你捷足先登,获得更大的盈利。

^① 样本 Alpha 与收益仅出于释义的目的,不意味着世坤投资或附属机构优化了任何策略。

第 7 章

基本面分析

作者：Xinye Tang/Kailin Qi

在 Alpha 的研究中，研究员试图找到能够在统计上准确预测资产收益的策略。这种策略，我们称为“信号”，通常基于资产价格的均值回归、“超前/滞后”效应、动量效应，以及对与资产相关的分析师评级、新闻情绪等信息的分析。基本面分析是证券投资分析方法的基石，显然是 Alpha 策略研究中一个很重要的方向。

股票的基本面分析，是通过挖掘可能会影响公司实际业务和未来前景的潜在因素从而判断证券内在价值的技术手段。我们试图通过基本面分析来回答以下问题：公司的盈利是否稳步增长，公司的债务偿还能力如何，公司盈利和收益质量是否良好，以及相对于债务，公司是否有足够的流动资产，等等。

从广义上来说，基本面分析是指分析一个金融实体的经济状况，而不是特意去分析其价格走势。可以从行业（产业）或者经济整体情况分析，而不

是单一地分析一只股票。相对于基本面分析，另外一种主要的证券分析方法，同时也是 Alpha 策略中的一个主要方向——技术面分析，主要考虑证券价格和交易量的变化，而不考虑背后的基本面信息。

基本面分析既可以是定性分析，也可以是定量分析。在 Alpha 的研究中，我们主要考虑定量因子，可以通过计算得到结果或者通过数字表达式来表示。财务报表是公司披露其业绩的标准媒介，我们通常从财务报表中获取定量信息以设计 Alpha 信号。很多实证研究都试图通过发现财务报表数据与证券价格之间的关系，来提高基本面分析的准确性。

主要的财务报表有资产负债表、利润表、现金流量表。

资产负债表是一个静态报表，描述的是一家公司在特定时间点的资产情况。而利润表、现金流量表是动态报表，描述的是一段时期内的资产变化情况。资产负债表记录了公司的各项资产和负债情况，以及公司的净资产，股东权益，即公司的净资产。负债水平可以通过从资产负债表里面得出的财务比率（如资产负债比、速动比率、流动比率）来衡量。这些比率可以帮助投资者来理解一家公司的债务偿付情况、信用评级和相对行业的平均业绩水平。其他的危险信号如留存收益的大幅减少、存货的大量增加、应收账款的大量增加也可以从资产负债表中看出。

利润表描述的是一段时期内的盈利水平，从中可以看到所有的收入、利润、成本、费用。息前收入去除了债务融资的利息负担，考查的是企业的经营盈利能力。出于谨慎的原则，费用项目可能包含当下与商品出售不直接相关的成本，而一些研发费用并未列示其中。

通过公司一年中的现金增减情况，现金流量表揭示了公司的健康状况。如果现金流量表显示该公司无法满足其股息需求，却依靠运营现金流保持其股本生产力，或者其经营现金流低于投资现金流，都是出现债务问题的危险信号。

进一步地深入分析财务报表，可以知道一家企业现在和将来的业绩表现。例如，如果从不同方面研究一家公司的应计收益，你会发现报表可能造假，从而得出不真实的公司状况。在投资者看来，公司现在有高利润，意味着将

来也会有高利润。而公司的利润由两部分组成：公司运营产生的实际现金流和应计收益，应计收益部分由会计师来计量和确认，这就有造假的可能。为了分析应计收益的准确性，我们应该去学习 Sloan 等人在 2011 年的一个案例。

首先，Sloan 用总资产对收益、应计科目和现金流进行标准化，来比较不同规模的公司。

其次，Sloan 分析了利润和应计收益的关系。提取了样本公司的应计收益、利润和现金流数据，根据利润分级。Sloan 根据分级将所有公司、年份进行均匀分层，计算每一级的平均利润。然后他跟踪该年份上一年和接下来 5 年的公司利润。

最后，Sloan 将跟踪结果按照应计收益进行分级，然后做比较。

结果表明，当用利润进行预测时，如果一家公司当年的利润很高，那么在未来的几年内，该公司很可能维持很高的利润。但如果根据应计收益分级，利用利润做预测的效果就不是很好。可以看到，和现金流相关的项目比应计收益的效果要好得多。也就是说，在分析未来利润的时候，由现金流产生的利润比应计收益可靠。

为了理解公司价值和财务业绩，我们分析主要通过财务报表计算得出的估值比率。比较常见的有市盈率、市净率。实证研究表明，一些反映多个估值比率的同时变化的因子，可以产生显著的超额收益。

除财务报表中的常见科目外，一些标准格式外的相关信息也会在附注中说明。附注会披露一些关键信息，帮助投资者更好地了解公司，做出切实的投资决定，比如，会计政策的具体内容、会计准则变更、长期采购合同、未决或正在进行的诉讼、管理层的股票期权等。很少会有公司将其经营失误和困难在标题或主表中披露，所以对于审慎的投资者来说，附注的字里行间是获取信息优势的来源。有的时候，公司会在财务信息披露里隐藏会计准则变更的事实和相关影响，从而避免股价下跌。通常来说，如果在财务报告里有很多新的附注信息，那么在这大段文字里面往往掩藏着危险的警示。如果你没有注意到附注里的这些信息，那么很可能是公司故意模糊事实。要想从普通投资者中脱颖而出，成为一个优秀的投资者，从附注信息里看出早期危险

信号的能力必不可少。

季度电话会议也是我们了解公司披露信息的手段。财务报表提供的是公司过去某个时刻的业绩表现，而投资者通过管理团队的季度电话会议，能够同时了解公司的现状和未来业绩的预期。一些评论员认为，从公司 CEO（首席执行官）和 CFO（首席财务官）在电话会议中说话的方式和语调中，尤其是在问答环节中解释公司与预期不符的情况时的表现，可以得到长期模型和技术指标需要的关键信息。实证研究表明，电话会议中的情绪因素能够影响未来 60 天的股票价格变化。一些特定的宏观经济因素也可能会反映相关行业的股票价格，例如油价的变化和交通运输企业股价的关系。

由投资银行或者证券公司等卖方提供的分析报告也是一种信息来源。这些分析报告会对某只股票和其相关产业的基本面和历史信息进行收集整理，然后分析、预测并建立估值模型，这些分析预测、股评、评级及推荐都会对股票和产业产生影响。而大型的机构投资者通过其持有的大量资产影响市场，他们往往会参考分析师的观点，这些信息都不应该忽略。

由于基本面数据更新频率较低，基本面相关的 Alpha 信号，相对于其他类似量价的信号，其换手率和股票覆盖率都很低。另一方面，基本面信息反映的是长期的股价。基于基本面的 Alpha 信号的累计收益，通常集中发生在财务公告的前后，然后在信息披露一年后逐渐平稳。也就是说，大部分超额收益是和公司前一年利润变化相关的。

基本面分析也给我们提供了一些股票分类思路。例如，我们可以根据公司的基本面状况把股票分为价值型和成长型股票。价值型股票通常指股价低，市盈率、市净率、市售率低，高股票分红的企业。成长型股票相反，通常指股价高，市盈率、市净率、市售率高，低股票分红的企业。同样，我们可以通过其他的基本面因素对股票进行分类。这些基于基本面的分类方法，可以帮助我们更好地理解不同类型股票的市场表现，从而设计出更好的 Alpha 信号。

第 8 章

股票价格与交易量

作者：Cong Li

在经济学中，有效市场假说（EMH）认为经济市场是“信息有效”的。也就是说，在投资行为发生时所有信息都可知的情况下，经过风险调整，投资者无法持续获得超过市场平均收益的收益。

有效市场假说有三种形态：弱有效市场、半强有效市场和强有效市场。弱有效市场认为在弱有效的情况下，金融资产（包括股票、债券和资产）的市场价格已充分反映出所有过去的公开信息；半强有效市场假说认为价格已充分反映出所有过去和现在的公开信息；强有效市场假说认为价格能充分反映所有公开信息，而且能反映所有内部未公开信息。所以在有效市场当中，仅通过现在和历史的量价数据，我们无法获得超额利润，也就没有量价 Alpha 策略。

真是这样吗？当然不是！实际上，随着信息技术的发展以及计算机信息处理、自动化交易等技术的出现，市场越来越接近于有效市场，但仍不是强有效市场。量化交易员通过发现市场无效性获得利润。

我们如何从这种无效性中套利呢？

信息比率（IR）（每日损益的平均值÷每日损益的标准差，用来衡量证券或者证券组合经风险调整后的收益）是信息系数（预测与实际收益的相关系数，IC）乘以交易次数（交易频率）的平方根： $IR=IC \times \sqrt{\text{breadth}}$ ，我们认为能够获得较好的风险调整收益。

与长期投资者一个季度才调整一次仓位不同，主动型的投资经理会交易得更频繁。他们每天都多次调整证券组合。交易的频率越高，越容易呈现统计特征。我们引入高斯分布或者正态分布概念，这个分布是一种连续的独立正态分布，通过落入区间的概率来表示事件发生的概率，当我们把 n 个独立分布的正态分布相加时，相对于原来的概率分布，正态分布的均值一样，但是标准差只有 $1/\sqrt{n}$ 。相同的 IC，4 倍的交易频率，IR 是原来的两倍。

从另一个角度看，我们把交易的所有资产当作一个整体（证券组合），而不是很多单独的股票。我们在看单一金融资产的时候，可以看技术指标，例如平滑移动平均线（即 MACD 指标，它是运用快速（短期）和慢速（长期）移动平均线及其汇合相交与分离的征兆，加以双重平滑运算得出的，一般用的是 14 天的数据）。当看一个股票组合的时候，我们关注的指标不同于单一金融资产，比如需要关注股票组合内所有股票间的关系、作为一个组合的运动一致性，以及不同股票对于相同事件的不同反应。在把证券组合视为整体的时候，才能做全局优化，使风险中性化。

最重要的是，我们要用别人没有的模型进行交易。交易员都很珍惜他们的模型而且不会公开，模型只有在适度使用的时候才有价值，一旦公开了，模型的预测力就会减弱，而且会立刻失效。市场也在不断进化，新的模型出现，老的模型消失。不断地寻找新的模型是一些公司得以生存的关键。模型和思路的来源很广，可能是经过学术论文启发而来的，也可能是由自己对市场的理解而来的，甚至可能是洗澡时的突发灵感。股票的微观架构很有趣，我们会问：基于什么情况，股票的价格会上涨或者下跌？什么因素会影响股票的未來走势、交易量或者反转走势？有很多工具可以用来寻找答案，如数学和统计方法在股票交易市场用得很多，模式识别和信号处理程序近年来也用得越来越多，很多简单的股票量价数据也可以用在量化金融里。

第 9 章

换 手 率

作者：Pratik Patel

我们一般用信息比率（Information Ratio, IR）和信息系数（Information Coefficient, IC）等指标来衡量 Alpha 策略的预测准确度和效果。IR 指的是基于参照物的超额收益和这些收益所对应的风险。Alpha 策略的高 IR 意味着在一段时间内低风险、高收益。IC 是预测与实际收益的相关系数，IC=1 意味着预测与实际一致。

高 IR 和 IC 意味着策略很好，但是我们要记住，这些比率忽略了现实交易中的一些限制条件，比如这些比率假定市场有无限的流动性，没有交易费用，而且除了我们之外没有其他交易者。但现实中的交易策略必须遵循一定的约束，关于市场情况做合理假设的 Alpha 会做出准确的预测，也更容易被使用。

1. 什么是换手率

新的信息会造成预测的变化。股票的瞬息万变，分析师改变他的推荐意见，公司公布经营业绩，都会造成交易活动的变化。我们通过换手率来衡量

交易，即总的交易量÷持有的总仓位。一个公司的股票价格比公司的每股收益变化得更为频繁，因此基于股价变化的 Alpha 策略会比基于基本面策略换手率高。交易越频繁，越能抓住盈利机会。我们发现基于股价的 Alpha 策略的 IR 和 IC 比基于基本面策略的高。

2. 低换手率一定意味着低收益吗

换手率和收益必须找到一个平衡。降低换手率并不意味着降低预测的准确性，相反，某些触发频率过低的信号通过平滑曲线的方法还可以提高表现。去极值（限制极端的数值）和衰减因子能够降低换手率，可以用来防止对信息变化太过敏感，从而造成预测值不必要的改变。最终结果取决于 Alpha 策略如何设计。但不管结果如何，理解在不同换手率下的 Alpha 策略的表现，能够帮助我们获得交易灵感。我们宁愿选择加杠杆而不太影响收益的 Alpha 策略，也不会选择为稍微降低换手率而损失所有收益的策略。

3. 流动性因素有什么影响

每笔交易都有费用，我们称为交易费（例如付给经纪人或者交易所的费用）和价差成本。交易费用是进行金融产品交换时产生的。在买入股票时，我们不但要支付经纪佣金，也要支付价差。买方给出的最高价一般都会低于卖方给出的最低价，这就是买卖价差。为了确认收益，你必须在给定的时间内卖出你持有的股票。而我们买入此股票的价格要高于卖价。

我们认为，买卖价差是市场的流动性造成的。美国股票市场最具流动性的 500 只股票的平均价差只有 5 个基点（每基点为 0.01%）。而小的股票市场如东南亚股市，其平均价差为 25 至 30 个基点。流动性不高的市场的交易成本就比流动性好的市场成本高，换手率就更加重要了。

了解市场流动性对我们设计 Alpha 策略非常有用。而当检验一个 Alpha 策略在各类金融工具中的可行性时，可交易性是一个很重要的因素。在大多数流动性市场适用的换手率，如果延伸到低流动性市场，就不一定可行了，或者换句话说，其可能在一个发达国家的市场有效，在其他的发展中国家的市场失效。例如，一个 Alpha 策略以 $x\%$ 的换手率交易美国最具流动性的 500 只股票能够盈利，但如果用于较低流动性的金融工具，则不可行（如用于流

动性排前 3000 名的股票)，用于发展中国的市场也不可行。所以在考虑策略的业绩的时候，要考虑交易成本，维持低换手率。

4. Alpha策略一定起作用吗

假设我们用价格和交易量，设计两个 Alpha 策略来预测未来股价。假设两个策略都操作相同的金融工具，而且具有相同的收益和 IR。一个策略根据金融工具最近的价格波动投资，另外一个根据金融工具的交易量投资。

$$\alpha_1 = \text{std}(\text{收益})$$

$$\alpha_2 = \log(\text{交易量})$$

我们发现第一个策略投资波动大的股票，而高波动的股票一般交易量不大，所以当资金规模大幅提升时，策略很难实现。第二个投资策略，会投资到高流动性、交易量较大的金融工具，而且很容易加入杠杆。如果假设交易量的波动性相对于价格波动更稳定，第二个策略的换手率也不会很高，进一步增加了其吸引力。

5. 什么样的换手率比较适合Alpha策略

这是一个平衡的问题。换手率边际衡量的是相对于换手率 Alpha 策略的实际收益。定义为利润除以总的交易量（交易资金的总量）。好的换手率要能最大化“利润÷IR”和换手率的比值。更重要的是，Alpha 策略在流动性不同的资产和不同的换手率下的表现，能够让你了解该策略的稳健性和可交易性。最后，所有的这些理论都是相对而言的。

第 10 章

回测——信号或者过度拟合

作者：Peng Yan

传统意义上，Alpha 是指经风险调整后的主动投资收益。在本章中，Alpha 指用来预测未来投资收益的量化模型。

1. 回测

Alpha 模型设计可能来源于一个假设、一篇论文、一个故事、一个灵感，甚至一个随机的想法。

(1) 想法需要被测试

和学术研究一样，很多假设是错误的，很多实验是徒劳的，仅有少数想法是成功的。市场的参与者都是人，每个人都不同，想法也不同，只有少数想法能够在现实环境中持续盈利。一开始你可能对模型很有信心，但是只有在测试之后，你才能判断模型是否真正有效。

资产价格被很多因素直接或间接影响。有些想法可能会影响某类资产而

对其他种类的资产无效。

(2) 模拟和回测

我们称测试想法的过程为模拟。模拟的方法有很多，举例如下。

① 蒙特卡罗模拟：通过大量模拟能够影响金融工具价格的不确定因素，得出结果的范围和区间。

② 定价模型：计算资产价值（Black-Scholes 模型就是一个计算期权价格的模型）。

③ 解释模型：建立模型用来解释历史事件。

我们把模拟测试理解为回测。也就是说，当你产生一个想法时，可以通过历史数据来测试这个模型的表现。回测理论基于如下的假设前提：如果这个模型对历史数据有效，那么这个模型对未来的情况可能也有效。反之亦然，如果一个模型历史数据表现得不好，那么我们就不考虑采用它。

回测的结果用于模型预先筛选，比较不同模型，判断一个 Alpha 模型的价值。回测的评价指标有很多，例如夏普比率、换手率、收益、相关系数等。

回测仅仅是我们对于模型的一个附加测试，好的回测结果并不能保证策略一定能获得收益。有很多因素会影响投资结果。一般来说，投资不能仅仅依靠回测结果，主要基于以下原因。

① 当前市场和历史市场有区别：市场规则会变化。例如，会有新的投资者、新的理论、新的技术。

② 模拟环境与真实环境的区别：比如，在买卖资产时，买卖的动作可能会对市场造成影响，而且买卖会有交易费用。是否能合理考虑到这些影响，对于测试结果至关重要。

③ 先验误差：如果你看到有人跟随市场趋势获得收益，然后设计一个跟随趋势的模型并进行测试，可能基于历史数据你会获得一个很好的测试结果，但是，如果不能真正理解其内在原因，设计出来的模型在未来可能并不能获得收益。

④ 过度拟合：有些时候，看起来很好的回测结果，其实只是随机产生的错误信号或者是干扰信号，没有任何预测的效果。

过度拟合是本章的主题，是统计机器学习的范畴，也是回测模型中一个比较重要的概念。金融市场总是充斥着各种干扰信息，即使是非常好的模型，其预测能力可能也很小。在有效市场假说理论中，是假定没有套利机会的，所以当你看到很好的模拟结果时，要当心模型过度拟合的风险。

2. 过度拟合

很多方法被用来降低过度拟合的风险。例如，十折交叉验证、正则化、先验概率等。十折交叉验证指将一个数据分为 10 组，轮流将其中 9 组作为训练数据，另外 1 组作为测试数据，重复 10 次，每次都能得到对应的正确率，并将 10 次结果中正确率的平均值作为对算法精度的估计。正则化，与统计学和机器学习中的方式类似，在模型筛选时，对一些模型中的极值进行“惩罚”，从而避免过度拟合。先验概率，一个不确定变量 P 的概率分布，可假设某些事件发生，以此为前提，对 P 的不确定性进行猜测。近年来在量化投资领域有很多关于过度拟合的论文，例如 Bailey (2014a)，Bailey (2014b)，Beaudan (2013)，Burns (2006)，Harvey 等人 (2014)，Lopez de Prado (2013)，Schorfheide 和 Wolpin (2012) 的著作。

(1) 过度拟合很容易发生

在对 7 组参数进行回测以后，研究员希望能找到一个至少在两年时间内，样本内测数据年化夏普比率超过 1，而样本外测数据夏普比率为 0 的模型。当这个研究员测试了大量的参数之后，总能找到一个在指定时间内符合这个要求的模型，所以在测试数据固定的情况下，需要有一个最小回测时间长度的要求，否则很容易发生过度拟合。

(2) 相关系数可能会欺骗你

假设我们有大量的随机时间序列，这些序列与一组新的随机序列的最大相关性很可能会超过 0.2，而实际上因为两者都是随机干扰信息，所以相关系数应该是 0。

（3）金融市场有记忆效应

过度拟合会降低样本外测数据测试的意义。在交易规模比较大的时候，交易成本和市场影响等带来的类似市场冲击的交易干扰信息，会在实际交易中造成损失。

（4）学术论文会有偏差

学术论文只说成功经验，而并不说他们验证了多少次，也不说他们失败了多少次，很多模型的结果甚至无法去重现。

（5）更高的接受标准

近年来，越来越多的模型被发掘出来，而且这个数量还将持续增长。因此对 Alpha 模型的要求会更高，尤其是对于横向预测的模型要求会更高。

3. 如何避免过度拟合

有很多减少过度拟合风险的方法，其中一些借鉴了统计学和机器学习的相关知识。

（1）样本外测试

为了测试一个 Alpha 模型，样本外测必须真正基于样本外。也就是说，我们建立一个模型，每天在真实环境中进行测试，观察测试结果。以下几种做法是不对的：① 我们用近 N 年的数据作为样本，用 N 年以前的数据作为样本外数据；② 用样本空间里一部分数据作为样本，而将另一部分作为样本外数据。在①中，近 N 年的数据包含过去的历史信息，所以今年奏效的模型当然很可能对历史数据有效；在②中，样本空间的数据是相关的，模型在一部分样本中有效，当然也会在另一部分样本中有效。

请注意，随着样本外 Alpha 模型的增加，样本外的测试也会有偏差。一个 Alpha 可能只是碰巧表现不错，基于单 Alpha 的样本外测试是不充分的。

（2）增加样本内测要求

增加夏普比率，延长用来测试的历史时间，扩大样本空间，这些测试方式都能降低过度拟合的风险。但在现实环境中，总有各种限制：并没有一个

系统化的方式来增加夏普比率；也没有足够多的历史数据，或者市场变化了致使历史数据无效；受到实际工具数量的限制，没有足够大的样本空间。

(3) 优化模型

符合这些条件的 Alpha 模型更佳：① 简单易懂；② 有好的理论和逻辑支持，而不仅仅源于历史经验；③ 能被解释，而且你能说明后面的原理。例如“Alpha=收益”就可能是一个很好的模型，而“Alpha=收益+交易量变化”就不是一个很好的模型，变量的单位不同造成模型可能无效，因为收益是美元单位，而交易量是数量，如份额等。

(4) 参数和操作

与机器学习模型类似，参数越少，模型对参数变化就越不敏感，这样能够降低过度拟合风险，用于调整参数的时间就会少很多。

此外，还有如下可以用的方法：

(1) 可视化

图表会比统计数值包含更多的信息。

(2) 记录实验次数

用同样的方法，记录实验次数对于评估过度拟合风险很有帮助。

(3) 使用人工数据

用一些人工合成数据来测试模型也是有用的。

(4) 动态模型

动态学习模型优于单一时间静态学习模型。

第 11 章

Alpha 与风险因子

作者：Peng Wan

本章将回顾提取 Alpha 的发展历史。我们将分析几个已经被充分研究过的市场反常现象，最后可以得出结论：Alpha 可以演变为对冲基金 Beta 或者风险因子。

1952 年，Markowitz 根据“期望收益与收益方差”建立了证券投资的证券组合。在此基础上，Treyner (1962)、Sharpe (1964)、Lintner (1965) 以及 Mossin (1966) 在 20 世纪 60 年代创建了资本资产定价模型 (Capital Asset Pricing Model, CAPM)，即股票的预期收益是对其所承受的市场风险的回报，公式如下：

预期回报率 = 无风险回报率 + 市场系统性风险 Beta × 市场风险溢价

为了对 Alpha 进行评估，我们需要将其中的市场 Beta 部分过滤出去，之后专注于剩余部分。在实际操作过程中，当我们将 Alpha 表达为股票仓位的向量时，通常认为货币和 Beta 均呈中性。

自产生之日起，资本资产定价模型因为其假设的限制性太大，并且与经验数据相差也较大，因此受到多方质疑。但另一方面，该模型将股票收益分解为一般风险因子和非系统性风险，这种方法在学术研究和实践过程中被广泛使用。

1992 年，Fama 与 French 为了解释股票收益，在他们的论文中增加了两项风险溢价因子：市值与价值。市值效应是指小市值股票的表现往往优于大市值股票。价值效应是指账面价值和市值比值高的股票在其收益中具有正向的风险溢价。价值效应的测量指标有收益率和股息率等。

1986 年，Amihud 与 Mendelson 提出了流动性溢价理论：流动性较低的股票具有较高的预期收益。流动性可以由买卖价差、股票换手率及交易量来衡量。2003 年，Pastor 与 Stambaugh 关注整个市场流动性，并将流动性 Beta 作为对股票总体流动性敏感程度的衡量指标。

1993 年，Jegadeesh 与 Titman 提出股票的动量效应，即新近获利者通常会比新近亏损者表现得更好。

这些因子，当然还有其他已经被深入研究过的风险因子，在金融理论与实践 中扮演了重要角色，并且在未来它们仍将是驱动股票获利的重要因素。我们发现，任何 Alpha 多少都会无意中引入一些风险因子。例如，尽管在人们未经处理的 Alpha 的构建过程中，即获得初始惯性的过程中没有包含任何价格信息，但因为新闻撰写者很有可能由于股票价格的飙升而兴奋，所以从新闻情绪中获得的 Alpha 很有可能产生惯性效应。

这些因子可能在将来继续保持正向风险溢价。但是在研究过程中，我们需要注意以下问题：

- 考虑到这些风险因子的高公开性，它们的夏普比率（风险调整后的回报）不可能很高，否则掌握内情的资金会不断涌入，直到股票的收益不再吸引任何资金。
- 一些风险因子，比如市值和流动性，要求在风险因子多空双方有较大的不均衡流动性，但在现实交易中人们不希望发生这样的情形。

这同样是风险管理的关注点，因为很难对多空双方均衡地进行平仓，尤其是在发生市场危机的情形下。

- 这些风险因子通常会在单位货币情况下产生更大的波动性。受宏观趋势的影响，这些因子可能会经历长期的回撤。此外，还会存在一种产业趋势，即一些风险因子被包装为“另类 Beta”产品（见 Institutional Investor，2014 年 9 月），这样可能会增加它们未来的波动性。图 11.1 中就有这样一个案例，2009 年一个简单的惯性因子在市场回撤的作用下进入了长回撤期。

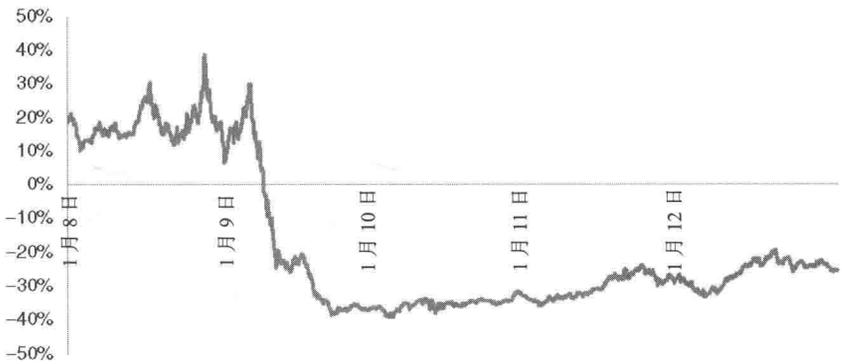


图 11.1 简单动量因子的累计收益率曲线

- 充分研究过的风险因子已经在整个量化投资领域中被广泛使用。如果某些投资人突然减持他们的股份，那么价格产生的影响可以高到促使其他人跟风减持，从而使局面更加恶化。这种风险在 2007 年 8 月份的“量化危机”中得到了充分表现。根据 Khandani 和 Lo 的说法，2007 年 8 月 7~9 日，由于一些参与者清仓，常用的量化风险因子蒙受了很大损失。因此，在这种情形下，最好采用有别于他人的策略，并且控制对一般风险因子的敞口。图 11.2 是一个假设的量化因子在 2007 年的“量化危机”中暴跌的情形。

因此，尽管这些被充分研究过的市场效应在长期的运行过程中可能会继续产生正向收益（可能是因为对风险和收益的理性考虑，也可能是因为非理性投资），我们在研究过程中更倾向于称其为风险因子，而不是 Alpha。在一

个新的 Alpha 策略里我们并不希望引入太多的风险因子。

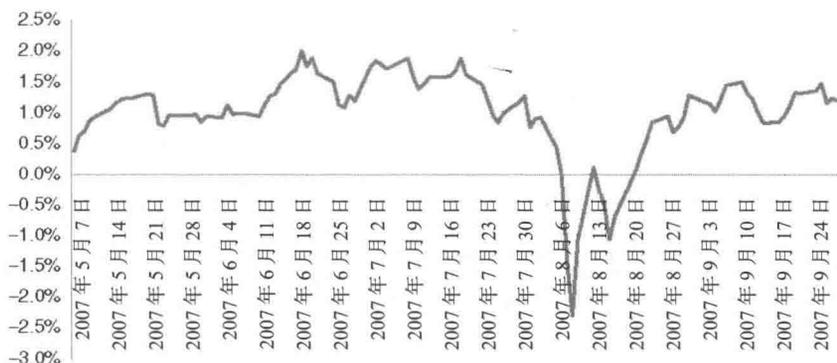


图 11.2 2007 年量化危机中假设因子的累计收益率曲线

我们需要在抵消掉风险因子后检查 Alpha 的余量。在实践过程中，一个好的 Alpha 在风险因子呈中性后应该具有高的夏普比率（风险调整后的回报），尽管单位货币收益在某些情形下会有所降低。

寻找 Alpha 通常是一个不断改善的过程。有些 Alpha 可能会随着时间慢慢失去价值。要想使 Alpha 的研究持续成功、有效，需要不断引入新的数据，不断进行创新。

第 12 章

Alpha 与证券组合风险的关系

作者：Ionut Aron

对于 Alpha 有很多种理解方式。最直接的方式是把它理解为一种技术手段，引申开来，就是利用这种技术手段来获取收益。大多数理性投资者会想尽办法在给定的风险水平上获取最大可能的收益。要想确定最优策略从而持续获取尽可能多的收益，需要一定的技巧。Alpha 可以反映出基金经理在不提高风险的前提下扩大证券组合收益的能力^①。

1. 证券组合风险

在 Alpha 的定义中提及“资产组合”并非随意之举，这是 Alpha 评估过程

① 在不增加风险的前提下增加收益并不意味着 Alpha 对于收益来说是一种没有风险的源。每种收益的源都有风险，但关键是要明白 Alpha 的收益来自证券组合中没有表示出来的风险源。换句话说，就是说 Alpha 从与证券组合里的风险正交（理想情况）的源获得收益，因此它对风险证券组合具有多元化效应。真正的 Alpha 并不能对证券组合里已有的风险产生杠杆作用，它只能增加不同的风险源。

的基础：我们需要根据初期的风险水平和收益来确定一个 Alpha 是否能够增加价值^①。这就意味着 Alpha 对背景信息有很高的依赖度，也就是说，一个 Alpha 可能可以增加某种证券组合的价值，却对另外一种证券组合无效。

为了帮助理解为什么会这样，可以将证券组合看作风险的综合体。每种收益源都衍生于其对应的风险源，因此，任何证券组合都可以看作是风险的综合体，根据证券经理的技术、偏好与/或约束条件，不同风险具有不同权重。

两种证券组合的不同效果取决于风控经理挖掘风险源和收益源的数量，以及他们在每种风险源上赋予的相对权重。极端情况下，一部分风控经理坚持倾向于关注少数几种风险，尝试通过增加赌注去增加价值，通过频繁交易弥补多元化投资的不足。另一部分风控经理则尽量避免市场择时或者因子择时（也不会在少量因子上下大额赌注），他们通过在证券组合中引入更多风险源来进行风险管理。介于两者之间的证券经理，要么更偏好于将两种操作方式结合起来，要么被指定必须要满足特定要求，在这些要求中，有些会限制他们可以涉及的风险种类，有些会控制他们在每种风险源上设置的权重，等等。例如，证券经理可能被要求他的证券组合里不可以承担任何利率风险，这样的话他就必须确保利率风险因子的敞口为零。这可以保证不论这个因子怎么变化，他的证券组合都能不受影响，但是这也迫使证券经理放弃与之相关的收益。为了弥补收益上的损失，证券经理可能会在其证券组合中的其他因子上冒更大的风险，或者创造 Alpha。

2. Alpha

很多风险源已经被广泛认识并理解^②。在极端的情况下，证券经理可能会被要求在其证券组合中去除大多数（甚至是全部）的已知风险（例如，根据某种特定的风险模型）。为了依然可以产生收益，证券经理就不得不对其他市

① 这并不是一个限制，因为参考证券组合里可以包含任何形式，如股票指数、债券指数或仅仅只是现金。

② 最被大众所了解的风险源很容易成为商业风险模型里的常态因子。Alpha 通常来源于正交空间，就是未知（或未被广泛认知）的风险。然而，随着一种新的风险源逐渐被认知，这种风险源慢慢地就会在风险模型中被考虑，建立证券组合时为了避免碰撞或拥挤，人们就会逐渐地减少对其的敞口。

场参与者尚不了解的、或尚未被广泛利用的风险源进行研究。

这就是创造 Alpha 工作的意义：寻找别人不了解的或与其他证券经理不同的风险源和收益（理想情况下为正交的）。正是因为了解的人不多甚至没有人了解，这种 Alpha 收益的衰减速度很慢，可以为发现它的人带来长期利润。又由于其特殊性，与之收益相关的风险不会对证券组合中已有风险产生杠杆作用，只是会使风险更加多元化。在效果上，每个 Alpha 都可以提供一种新的收益流，稀释证券组合中的风险压力（假设它与证券组合中的风险正交），这样单个风险就不可能会对证券组合产生重大影响。

有一点非常重要：要尽量降低对已被其他市场参与者广泛知晓的风险源的敞口。因为这些已被广泛知晓的风险很容易受到拥挤效应与从众效应的影响，对证券经理的证券组合造成重大威胁。只要其他市场参与者忽视它，这种因子的风险就会很小。然而一旦它的使用变得过度拥挤，其风险—收益曲线就会发生剧烈变化。随着追逐某一因子的资本的增加，证券经理从该种因子收获的收益就会减少，而采用该种风险的证券组合面临的风险值就会增大。这是因为在某些时刻，资本的撤出更可能大于资本的注入，而且通常的情况是，这种撤出会像暴雨一样突然而至（除非受到中央银行的调控），而资本的注入通常是个缓慢的、长期的过程。这就是为什么很难选择退出某种因子的恰当时机的原因，同时也是为什么关注少数几个因子的证券组合比多元化证券组合波动性更大的原因。恐惧对情绪的驱动力比贪婪大得多，因此抛售发生得很突然，几乎不给证券经理任何预警或反应的时间。

另一方面，未知因子（至少为他人所不知）不应该被等同论之，其风险敞口不应该被严格限制，因为它们发生拥挤的概率并不高。因此，对这些因子增大风险敞口有益于优化证券组合，因为收益并没有在其他市场参与者之间进行分配，又因为参与者少，所以引发突然抛售的情况也很少。然而在实践中，很难确定究竟有谁知道并在使用同样的 Alpha，以及在它们上面究竟有多少资本在流动。因此，证券经理通过使用尽量多的不同的 Alpha 来稀释这些风险是非常明智的。

3. 减少Beta，增加Alpha

对于有风险意识的证券经理来说，在理想的证券组合中应该尽量将那些

广为人知的风险因子的风险压力 (Beta) 控制为零^①, 同时应该尽量多地发现新的未知因子 (Alpha)。Alpha 应当与所有已知因子正交 (这是为了真正使其不受其他市场参与者市场行为的影响), 并且不同的 Alpha 之间也要相互正交 (从而确保证券组合中具有最大的风险多元性)。这样的证券组合的收益将全部来自于 Alpha 成分, 在很大程度上其余市场参与者并不能从中获益。

在现实中, 很少有对已知风险敞口为零的证券组合。证券组合中的因子, 无论是否众所周知, 彼此之间都几乎不可能完全正交。因此, 证券经理的 Alpha 通常是以下两种策略的结合: ① 因子择时 (根据时机调整不同风险因子的敞口); ② 掌握其他市场参与者未知的风险源。

在第一种情形下, 其他市场参与者虽然不太可能知道证券经理的 Alpha 具体是什么, 但是他的 Alpha 只是一些已知因子的组合。尽管这种 Alpha 并不会使证券组合中已有的风险源更加多元化, 但是它可以提高证券经理的因子择时效果。证券经理对风险压力可以进行更好的组合, 使得每种风险因子可以获得尽量多的正向影响, 同时尽量减少负面影响。这种类型的 Alpha 可以通过改变操作时机和频率等方式重新建立平衡的行为, 对证券组合中已有的风险敞口进行调整。尽管这样做可以提高收益, 但是同时也意味着波动表现的降低, 所以应当注意, 使用这种类型的 Alpha 需要付出一定代价。原因很简单, 如果一种 Alpha 提高交易频率 (或规模), 那么这种证券组合的可测量性及证券经理对其的控制能力就会受到严重影响。一般情况下, 证券经理进行因子择时的次数越多, 就越不可能扩大其证券组合的规模, 因为频繁 (尤其是大额) 地转手可能对收益有负面影响, 原因在于转手需要一定的手续费, 并且价格影响较大。

与发现全新的 Alpha 相比, 最有价值的是发现不会对已有风险源产生杠杆作用的 Alpha, 这样可以做到风险多元化。正如人们所料, 这些 Alpha 很难被

① 如果证券经理有风险因子时间管理技巧的话, 赋予人尽皆知的风险因子一定的权重也是可行的, 因为在这种风险因子代价相对较低的时候将其引入、在这种风险因子代价相对较高时候将其卖出可以扩大回报。有风险因子时间管理技巧的证券经理可以不必寻找其他不为人知的风险源, 因为他可以只通过建立关于已知风险因子的证券组合并根据时间模型调整敞口来产生巨大的收益。

发现，尤其是在已知风险因子的数量已经相当大的情况下（并且在不断增长，每年在核算商业风险模型里风险因子数量的时候可以看出这种趋势）。为了发现这种 Alpha，就必须对已知风险因子的正交空间有意识地进行深入研究。这个过程代价很大，产出可能很小。然而即使这样，也是值得的。这种 Alpha 除了会产生收益之外，还会使证券组合实现真正的多元化。这种 Alpha 的绝对回报虽然很诱人，但其真正价值在于发现了一种完全不同的风险源，可以为证券组合带来不同于其他风险源收益。

5. 需要注意的问题

考虑到 Alpha 与证券组合风险之间的动态（进化）关系，对于寻找 Alpha 或运用 Alpha 构建证券组合的从业者来说，除要评估 Alpha 的绩效之外，还要对 Alpha 在以下关键性问题上的表现进行评估：

- 它真的与众不同吗？为了促进风险多元化，Alpha 必须与众所周知的因子及其他证券组合中的 Alpha 正交。否则，这个 Alpha 就会对风险产生放大作用。在这种情况下，Alpha 与已知因子及其他 Alpha 的重叠度越低越好。证券经理至少应当认识到这个重叠度的存在，并且清晰了解在每个因子上承担了多大的风险。
- 谁不知道它？由风险带来的收益必须不被其他市场参与者知道。如果所有人都知道的话，它就不能被称作 Alpha 了。就算有人知道，证券经理需要做的也是尽可能地减小敞口。Alpha 收益（并非无限）要在所有的使用者之间进行分配。同时，在证券组合中持有这种 Alpha 的风险，会随着追逐其收益的资本和市场参与者数目的增加而增大，这就意味着负面效果会比正面收益更大。
- 它合理吗？如果一个 Alpha 从不符合市场行为的活动中获得获益，那么这个 Alpha 基本上是没有用的。如果一个 Alpha 交易的全是非流动性的资产，那么它怎么会产生收益呢？
- 它能规模化吗？对因子择时依赖度很高的 Alpha 非常难操作，原因在于这种 Alpha 要想成功的话，必须要有更多其他市场参与者的配合。你需要一个交易伙伴执行你的 Alpha。更重要的是，你对这样的人是

否存在以及他将要采取什么样的行动，必须判断正确。这种情况发生得越频繁，你需要进行的交易量就越大，你从 Alpha 中获益的机会就越小。因为，你希望交易的条件可能不会一直存在，或者其可能会与你的假设不符。

- 它会持久吗？好的 Alpha 必须不能被初阶市场参与者直觉性地感受到（最好对他们没有吸引力），但是对于进阶型市场参与者来说，它是非直觉性的、有逻辑的。这样就可以通过时间差，减少其他市场参与者对收益的瓜分。即使是一个不容易被观察到的因子，也可能迅速地被其他人发现，从高收益低风险源变为低收益高风险源。

第 13 章

风险与回撤

作者：Hammad Khan

风险有多种定义方式。Alpha 的风险可由其收益波动水平和预期风险价值 (VaR) (这种方法可以衡量在统计置信水平下 Alpha 可以引发的最大可能损失) 来衡量。风险发生后, 可以被精确测量。但在 Alpha 风险发生前, 也可以使用一些简单的方法进行预测。

- 头寸集中在特定证券或特定证券组合里。例如, 在股市中, 如果一个 Alpha 大部分头寸都在特定股票或特定行业或板块里, 那么我们可以预期, 一旦这些股票或者股票组合面临系统性事件, Alpha 会承担很高的风险。同样, 在期货中, 如果 Alpha 在商品、债券或股票指数中头寸比较集中, 那么它也会面临承担系统性风险的可能。如果我们知道 Alpha 在这些头寸处的敞口, 就可以对风险进行预判。
- 世界各地许多量化研究者和证券经理已经对一些 Alpha 及其相关技术有所了解。因为在世界范围内追逐这些 Alpha (我们不妨称其为因子) 的资本太多了, 它们已经成为一般风险源。如果这些因子的绩

效衰退，资本就会从使用这些因子的证券组合中外流。而资本外流可能会加剧投资组合的亏损。通过衡量头寸与 Alpha 绩效之间的相关性，可以确定这种因子的敞口，因此它是可以预测的。值得指出的是，在样本内测期间，尽管承担了已知风险，但 Alpha 的绩效也可能是非常好的。然而只要承担着这些风险，Alpha 未来的绩效还是有可能急剧恶化的，其收益（PnL）会剧烈波动。在金融领域，风险这个课题覆盖范围很广，每年都有许多书籍、论文及调查报告对它进行论述。在这里我们只简要地讨论在本书主题范围内设计 Alpha 时的风险。

投资多元化程度可以降低 Alpha 或证券组合的风险。在头寸集中度可控的前提下，使用的金融工具种类越多，多元化效果越好。例如，只基于英国富时 100 指数（FTSE100）构建的 Alpha 比基于整个英国和欧洲股票市场构建的 Alpha 的多元化程度低很多。然而需要着重指出的一点是，我们在扩大标的选择区域的同时，也会引入其他风险。这些风险包括在相同的 Alpha 内对不同国家和不同货币体系的风险敞口。应当认真考虑、谨慎对待这些风险，有时还要通过中性化方法来降低这些风险。同样，使用丰富的资产种类，如期货、股票、债券等，也会提高多元化程度。

关于风险，需要指出的一点是，一个好的 Alpha 尝试着区分股票中的胜者和败者，从而建立多空组合，但是这并不是百分之百会发生的。我们假设有一个 Alpha 可以区分胜者和败者，而且这些股票的胜败趋势长期稳定。如果这种情况真的发生了，那么这个 Alpha 对这些股票的价格趋势就有了风险敞口。有时，交易某只股票赚取到一笔确定的收益后，Alpha 会减弱这只股票的持仓信号，而不是持续跟进这只股票的趋势，直到产生回撤。同样的事情会发生在利用 Alpha 预先设置止损的情况下。如果样本空间不够广的话，以上这些技术就尤为重要了。

1. 回撤

回撤是 Alpha 损失相对其峰值的百分比。例如，如果开始时，某个 Alpha 可以赚得 20% 的收益，之后的几天（或几周）收益率下滑到 18%，那么回撤就是 2%。因为每个 Alpha 都会有或好或坏的时候，所以每个 Alpha 都会发生回撤。我们在回测时通常关心的是回撤的两个重要指标：

- Alpha 的历史最大回撤（以及每年最大回撤）；
- 最大回撤历时。

在进行回撤时，对回撤的测量应当结合 Alpha 的其他相关指标，如它的年化收益率和投资率。有时好的 Alpha 也会发生突然回撤，之后 Alpha 又恢复为之前的模式。有时，回撤发生得非常缓慢、平滑，这种下跌会持续很多天，直到 Alpha 重新开始起作用。当然，在实际情况下或者在进行样本外测时，下跌发生时我们可能没办法知道是 Alpha 彻底失效了，还是只是发生了回撤、以后还能恢复。因此，在样本内测阶段，统计历史最大回撤和最大回撤历时是非常重要的。这可以为我们评估 Alpha 现状绩效提供标准和依据。

如果在设计 Alpha 时考虑了其他评估方法（如避免过度拟合），则应当注意避免发生大规模回撤，在上升阶段不太需要关注。回撤与风险密切相关。以下一些简单的最小化风险的方法可以帮助你降低回撤量最小化：

- 最小化所有单个证券的最大敞口；
- 最小化所有相似证券组合中的最大敞口；
- 最小化对一般 Alpha 因子的敞口；
- 最小化对整个市场（局部市场或全球市场）的敞口。

在设计 Alpha 时，人们往往容易被其收益所引诱，而忽略其回撤。但是使用以上（或者其他未指出）的技巧可以降低未来发生回撤的风险。还要注意在减小风险敞口的同时也会系统性地降低 Alpha 的绩效（因为某些样本内测绩效正是由这些风险驱动的）。

2. 风险和预期回撤的计算

需要注意的另外一点是绩效集中。尽管对敞口进行了一定的控制，但许多 Alpha 可能在相似的证券组合中获利，这将放大 Alpha 在某些方面的潜在风险。参照如下案例：在一个受制于板块敞口的股票 Alpha 中，样本内测的大多绩效来自能源和信息技术板块，其他板块并不很好。

图 13.1 为这种 Alpha 的一个案例。

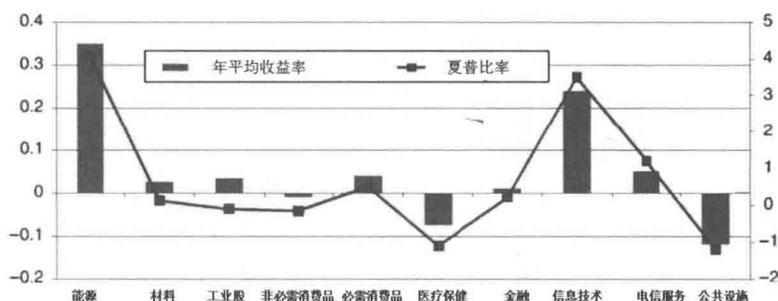


图 13.1 主要受能源和信息技术板块驱动的股票 Alpha

尽管整体样本内测绩效可能看起来非常合理，但是如果以上两个板块中有一个失效了，Alpha 就会退化并发生回撤。好的 Alpha 应当在多个证券（或多个分组，比如板块）中尽可能地平等分配其绩效。图 13.2 是一个很好的案例。

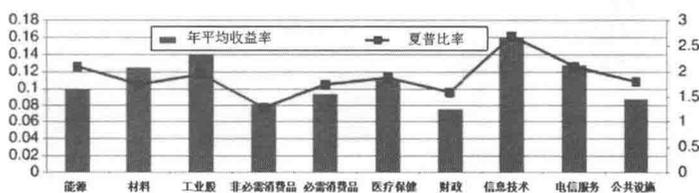


图 13.2 各板块绩效差异性较小的证券 Alpha

然而，为特定证券组合设计 Alpha 时不应该将这种情形视为硬性规定（如某些 Alpha 的目标样本很小，只有少数板块或行业）。

类似的一个方法是，研究员应当经常检验 Alpha 的五分位绩效。最好的检验方法是将目标样本分为 5 份并分别对应 Alpha，然后计算每份的收益和平均值（及标准差）。在最好的一份中 Alpha 是具有最优未来收益的 Alpha（大多是具有最高正值的 Alpha），在最不好的一份中 Alpha 则是具有最差未来收益的 Alpha（大多是具有最低负值的 Alpha）。其他的则按照未来收益值从高到低依次排列。

图 13.3 为一个五分位 Alpha 分布的案例，这个案例中展示了较为理想的绩效分布。

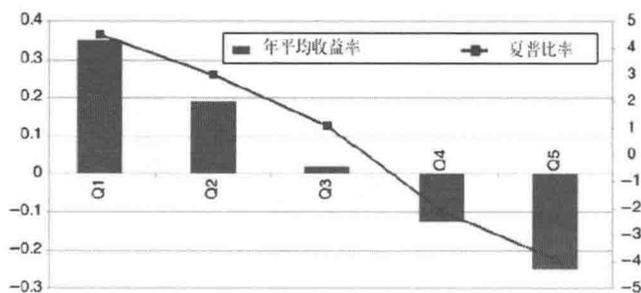


图 13.3 人们期望发生的 Alpha 分布

我们在许多 Alpha 中发现，Alpha 的绩效基本取决于头 1/5 或末 1/5，中间的 2/5 ~ 4/5 一般是噪声。图 13.4 中为这种类型 Alpha 的一个案例。

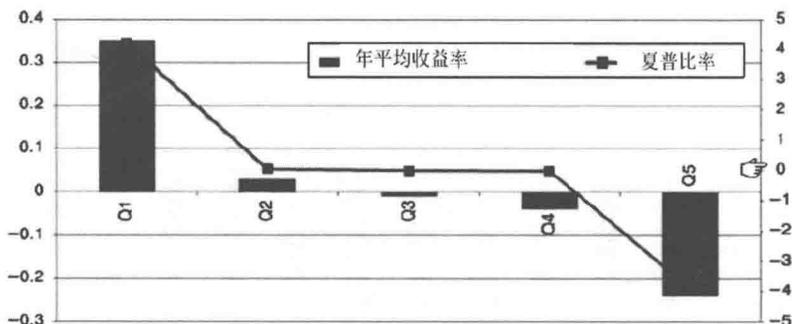


图 13.4 Alpha 只在末区域具有预测效力的五分位分布

由于这种类型的 Alpha 其预测能力只来自于尾端分布，如果尾端区域未来萎缩，则其适用范围会降低，发生回撤的可能性会增大。由于在这种 Alpha 的中间区域没有信息，因此选择中间五分位的信号是不可行的。Alpha 只能应用在末区域的案例有很多。有时，我们也能发现中间 2/5 ~ 4/5 的效能系统性地超过头 1/5 或者末 1/5 的情况。

图 13.5 和图 13.6 中的两个案例，它们的头 1/5 或末 1/5 较其他五分位效能更低，预测能力也较低，因此会导致未来绩效变坏。

综上所述，在设计 Alpha 时考虑风险和回撤非常重要。样本内测试绩效图和总结数据只能反映 Alpha 的部分情况。通过研究与已知 Alpha 的相关性、绩效集中性及五分位绩效分布，研究人员可以清楚自己正在面临什么样的风险源。

这些工具可以帮助研究人员更好地理解并预测 Alpha 面临的风险。通过构建与本章中提到的特性相近的 Alpha，可以降低未来发生回撤的可能性，设计出更好的 Alpha。

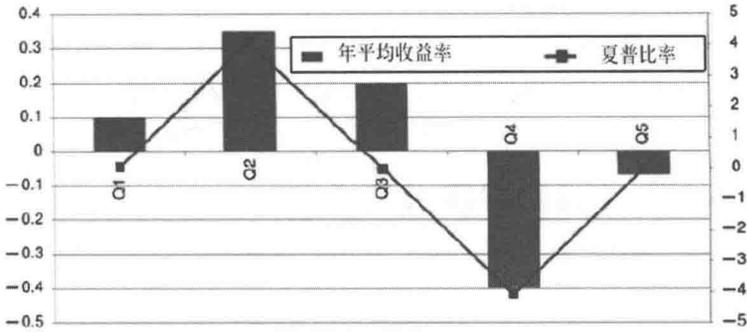


图 13.5 尾部预测能力较低的五分位分布

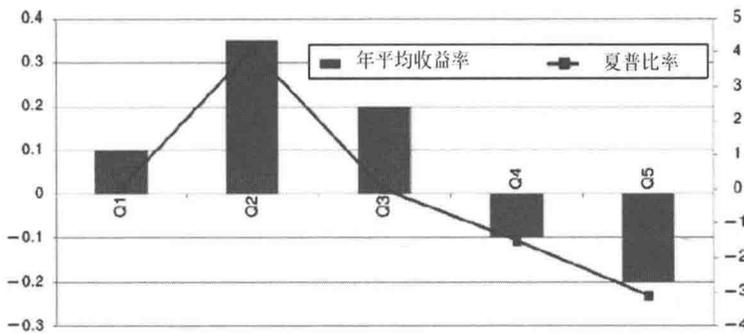


图 13.6 第二五分位与第三五分位效能超过头五分位的分布

第 14 章

数据与 Alpha 设计

作者：Weijia Li

数据在设计的过程中扮演了核心角色。首先，我们需要利用基础数据进行模拟。基础数据是指股票价格和交易量。不论你想要对什么类型的 Alpha 做回测，你都需要利用这些基础数据计算收益率、夏普比率及换手率等指标。如果没有这些指标，我们永远也不会知道这个 Alpha 是好还是坏。其次，数据本身对 Alpha 具有启发作用。例如，你可以画出某些股票的价格/交易量图表，检验历史上是否有类似现象发生，你也可以使用价格/交易量数据进行技术分析。如果你有方法获得公司盈利的数据，那么你就可以顺理成章地基于公司盈利数据进行股票交易。

1. 如何寻找 Alpha 所需数据

对于 Alpha 研究人员来说，寻找新的数据是非常重要的技能。人们通常更喜欢绩效好、相关性低的 Alpha。新的数据集可以达成上述目标。有时我们可以从一个数据集中获得信号，但是不管我们如何努力去改善它们，这些信号都还是不够强。如果我们能够得到另一组数据，就可以从不同的角度审视目

标公司，增强原始信号，获得更优质的信号。我们总是想要创造出相关度较低的 Alpha，提高 Alpha 信号池的多元性。然而，尽管构建 Alpha 的方法不同，但有时，基于同一个数据的 Alpha 信号也会表现出高度的相关性。这是由于使用了相同的数据，Alpha 信号呈现出天然的相关性。如果我们有一个新的数据集，新的数据集会激发出构建 Alpha 的新想法，这样根据新数据集构建 Alpha 的方法也是新的。使用新数据的 Alpha 信号和使用之前数据集的 Alpha 信号相关性会很低。通过使用新数据，可以改善绩效，提升多元化程度。

（1）文献中的数据

在构建新 Alpha 时，如果有新数据，将会事半功倍。但是，我们该如何获得新数据呢？事实上，如果你能找到合适且相关的数据，那么你就成功了一半。数据的来源有很多，最常用的来源是学术文献。正如我们前面所提到的，数据通常与 Alpha 的构建相关。如果我们在互联网上搜索“股票收益”，则会查到成千上万篇关于“超额收益”的论文。在这些论文中，我们可以查到作者们在研究中所使用的各种各样的数据，如价格/交易量数据、基本面数据、收入数据等。在获得数据之后，我们就可以使用论文中所用的方法构建 Alpha。你也可以利用互联网上的公共数据，但是需要指出的一点是，越鲜为人知的数据，可利用的价值越高。如果这些数据广为人知并且公共可用，那么可能会产生相似的 Alpha 模型，这将使得 Alpha 套利空间（如超额收益）逐渐减少。但是如果这些数据从来没有用于价格预测，那么即使很多人都知道，它们仍然具有被使用的价值。

（2）供应商处的数据

数据是非常有价值的信息，因此售卖数据也是一门生意。专门从事数据采集、解析、处理和销售的数据供应商有很多。如果数据比较简单，数据供应商可以提供他们收集的原始数据（比如价格/交易量数据）。对于比较复杂的数据，供应商可以对其进行解析和处理，然后交付给客户，如基本面数据。对于非结构化而且非常复杂的数据，如新闻、推特等，供应商会采用自然语言处理技术对原始数据进行分析，最后将机器可读取形式的数据交付给客户，而非仅是人类可读的原始数据。供应商甚至可以直接售卖 Alpha 模型，这就意味着一些 Alpha 模型的输出结果就是数据。顾客只需要导入数据即可据其进行交易。

2. 数据检验

作为 Alpha 研究人员，不论什么时候，只要拿到新数据，首先要做的事情不是利用数据检验 Alpha 信号，而是检验数据的可用性。在 Alpha 模拟中，数据的发布时间是个非常重要的因素。对于任何一种数据，只要它没有时间戳，就是没用的数据。这是因为如果不知道数据创建的时间点，我们就不可能有效地了解它的进程，只能盲目地使用它。有了正确的时间戳，我们就可以正确地对其进行模拟实验。如果我们尝试使用当前还不可用的数据（未来数据），则我们会引入前瞻性偏差，将使得 Alpha 绩效惊人。如果我们不能及时地使用数据，Alpha 效能就会低于预期，我们会损失一部分收益。例如，如果苹果公司（AAPL）的收入超出了预期，那么它的股票价格可能会上涨，假设其他条件不变，因此在发布通告时买入是最好的时机。然而，如果我们延迟一周买入 AAPL，将不会获得这些收入，因为好消息已经通过股价体现出来了。因此，对于每一个我们想用于 Alpha 研究的数据集来说，都需要掌握公布它的时间，确保我们在数据可用的时间点及时地引用数据。这样，我们就可以在确保不引入前瞻性偏差的条件下，获得有意义的模拟结果了。此外，我们需要确定数据可以支撑 Alpha 产品化，这意味着数据要按照固定的时间表持续被生产。有时，数据生产商会停止生产数据。这种情况下的数据对 Alpha 是不适用的，因为在 Alpha 模型中没有被及时地持续输入数据。

数据可能产生的另一个问题是幸存者偏差。例如，数据商提供了一种 Alpha 模型，测试的时候绩效很好，但是这并不意味着这个 Alpha 模型在未来也会表现很好。因为我们不知道数据商开发了多少模型，这些模型是基于什么建立的。如果数据商尝试了 1000 个模型，只有一个成功了，我们就可能会面对幸存者偏差。这个偏差是由数据商引入的，是超出我们的控制范围的。在这种情况下，一些样本外测的数据就更有用了。因此，样本外测是非常有用的，如果 Alpha 在样本外测绩效表现依然很好的话，则我们可以认为该 Alpha 的稳定性很好，因为测试是在不受控制的样本库进行的。

我们在进行 Alpha 设计的时候，应当对数据的完整性进行检查。在进行历史模拟时，哪怕一个坏数据点也能破坏整个 Alpha 信号。在现实生产中，如果你认为数据总是正确的，那么你将面临很大的风险，因为错误的数据可能会彻底扭曲 Alpha 信号，导致巨大的损失。因此，我们需要进行一些基本的检测，

如在 Alpha 编码中去除极值。有了这些基本保证之后，Alpha 的稳定性会更好。

3. 在使用数据前先了解数据

严谨的 Alpha 研究需要对数据进行深入的理解。我们一定要理解数据，这样才能更好地利用它。对于某些简单数据来说，只是对数字进行简单的处理就好。但是，对于复杂的数据，我们需要在 Alpha 研究中进行深入探索。有时，为了理解某些数据，我们需要了解一些其他事情，比如，为了理解成百上千的基本因子，需要了解一些公司财务。只有当我们完全理解数据时，才能构建更理性的 Alpha。基于对数据理解深刻的 Alpha 的稳定性更好，也更有可能成功。

4. 迎接大数据时代

如今，数据呈现爆炸性增长。数据增长表现在 3 个方面：种类多样、数量惊人、增速极快。过去，用于预测股票价格的数据可能只有价格/交易量和和其他基本面数据。现在，我们有更多选择。我们可以用的数据类型非常多，有时甚至可以尝试非常疯狂的想法。Kamstra 等人在 2002 年提出了 SAD 效应（季节性情绪失调）：股票市场收益随着白天时间的长短呈现季节性变化。Hirshleifer 和 Shumway 在 2003 年发现，通过某个国家早上的日照强度可以预测那一天股票收益的市场指数。Preis 等人于 2013 年利用 Google Trend 的数据获得了 326% 的收益率，成功地超过了市场平均水平（16% 的收益率）。

每天都有大量的数据被创造出来。以美国的股票市场来说，一级 tick 级别数据每天大概是 50GB，二级 tick 级别数据每天超过 100GB。社交媒体也创造了大量数据，比如推特用户平均每天发送 58 000 000 条推特信息。

数据创造的速度也是很快的。现在高频交易公司的延迟期为微秒级，尽管如此，人们还是希望购买更快的处理器，选择更快的接入方式，尽快完成交易。数据商也正在尽力地将速度提升至极限，以获得更多的客户。

只要在控制范围内，数据当然越多越好。管理快速增长的数据是一项非常具有挑战性的任务。但是成本也是需要考虑的重要内容，比如存储设备、计算设备、定制数据库等。另外，如果我们能正确处理这些数据，那么就可以获得更多可用的数据，创造更多、更快、更好的 Alpha。所以，在大数据方面的大手笔投入意味着巨额收益！

第 15 章

统计套利、过度拟合与 Alpha 多样性

作者: Zhuangxi Fang

Alpha 可以直接预测股票市场吗？或者更确切地说，我们可以预测苹果公司 (AAPL) 在 2030 年 6 月 29 日的股价吗？不幸的是，我们很可能预测不到。统计套利的本质就是我们只能在统计意义上进行预测。

统计套利的关键假设认为金融商品价格是有规律的，通过研究历史数据可以发现这些规律，并运用规律预测未来。证券的价格由多种规律驱动，对于特定时刻的特定金融商品，这些规律并非必然适用。但是，真正驱动价格的规律，或者基于价格驱动规律的好的 Alpha，应该在我们研究的交易日中表现出具有统计意义的预测性。一个简单的规律就是人们所熟知的均值回归，它认为股票价格围绕其均值随时间波动。但即使在适用均值回归规律的股票市场中，也不难找到股价持续上涨或下跌从而背离其均值的股票。但即便存

在这样的股票，如果检测 M 个交易日内的 N 只股票的集合，你会发现在 $N \times M$ 个样本点中，遵循均值回归规律的比例大于 50%，这样我们就有机会基于这种驱动规律构造 Alpha，以获取利润。

在量化世界中，利用统计套利寻找 Alpha 的过程具有过度拟合或发现伪价格驱动规律的风险。如果一个规律是伪规律，即使 Alpha 在历史数据上表现出统计显著性（因为它是根据历史数据拟合的），在未来它的这种显著性也会消失。“代码中含有 C 的股票通常在星期三上涨”这样的 Alpha 很有可能不是值得投资的 Alpha，即使它在过去可能表现出了盈利性。

这种现象在日常生活中也常出现。例如在 2006 年世界杯足球赛之前被发现的“3964”公式：阿根廷在 1978 年和 1986 年夺冠，将两个年份加起来是 3964，德国在 1974 年和 1990 年夺冠，将这两个年份加起来也是 3964，巴西在 1970 年和 1994 年、1962 年和 2002 年夺冠，这两组年份之和都是 3964。一切看起来都是那样完美，但是如果根据这个规律预测 2006 年的赛事，则应该是在 1958 年夺冠的巴西获胜，然而事实上最终的冠军是意大利。这个规律毫无意外地在 2010 年继续失效，西班牙成为冠军俱乐部的新成员。虽然这个 Alpha 规律不怎么准确，但其也不是没有一点可取之处。之前获得过冠军的球队具有胜过其他球队的实力，因此再次赢得比赛的概率确实更高。

这一点对于构造统计套利 Alpha 很有启发作用。在过去表现出统计显著性的所有规律或模式中，那些可以在某种程度上反映价格驱动力的规律对未来的预测能力更强。虽然单纯地对数字进行挖掘可以发现一些显著的结果，但更为重要的是深入理解在 Alpha 的构造过程中潜在的价格驱动规律。采用合适的实现方式，将驱动规律转化为 Alpha。

金融商品的价格由多种因素驱动，如交易微结构、基本面估值、投资心理等。因此，能够用于构建 Alpha 的价格驱动的规律有很多。此外，基于同样的价格驱动规律，利用不同的实现方式也可以构建不同的 Alpha。例如前面曾提到的均值回归规律，其中关于“均值”的算法有多种，关于“回归”的定义也不止一个，这样就可以构造出不同的 Alpha，这些 Alpha 都可以表现出盈利性。

这里，笔者用一个简单的例子对本章内容做总结：车轮。车轮通常都是圆的，但为什么必须是圆的呢？对于车轮形状的真正限制是它必须具备凸的、宽度固定的截面。因此，只要遵循这个规律，制造出其他形状的轮子也是可行的，正如勒洛三角形理论（the Theory of the Reuleaux Triangle）中所论述的一样。

第 16 章

提升 Alpha 信号稳定性的技巧

作者：Michael Kozlov^①

做 Alpha 研究的主要目的是预测并超越市场表现。然而，大多数投资人期望达到的不仅仅是回报，还包括低风险敞口、交易结果的可预测性等。所有这些因素都引导我们去思考稳定的 Alpha 策略的定义。

稳定的 Alpha 策略应该具有如下特性。

(1) 不随交易环境的变化而变化：Alpha 策略应该体现它统计的属性，它独立于整体交易市场的选择。或者说，我们不希望我们的 Alpha 策略依赖于某个金融工具或者一组工具。这一约束频繁地被市场所强加，其原因包括政策调整、客户偏好、流动性降低、做空禁令等。

(2) 在极端市场下的稳定性：Alpha 信号所有指标都不应该出现急剧下跌的情形。最常用的用于反映 Alpha 信号性能的量化指标为：信息比率、最

^① 感谢 Andrey Perfilyev, Itay BenDan 和 Boris Dvinsky 的宝贵意见。

大回撤、收益等。

1. 改善稳定性的简单方法

下面我们介绍改善 Alpha 信号稳定性的几种有效方法。

(1) 排序法

① 划分等级：划分等级指按照某种排序方法，将一个向量的元素替换为它们的序号（从 0 到“向量长度减 1”）。通常将所有向量元素除以“向量长度减 1”，以进行归一化。如果两个值相同，那么它们应该具有相同的序号，该序号等于它们对应位置的均值。

划分等级可以用在 Spearman 于 1987 年提出的等级相关模型中，该模型在很多情况下比 Pearson 在 19 世纪 80 年代提出的经典皮尔逊相关系数模型要稳定得多。皮尔逊相关系数模型对于非平稳或非线性输入的结果也被公认是不稳定的。

② 分位点近似：分位点就是对一个随机变量的累积分布函数以固定间隔分位采样得到的点集。将有序数据分别归类在有相同数量的数据子集被称为 Q 分位法；分位点就是标记连续数据子集之间边界的数据值。

举例来说，我们熟知的最小平方回归法（也叫最小二乘法）对于非平稳、非线性输入可能是不稳定的。然而，它可以被最小分位数平方（LQS）所取代，其目标是使平方残差的某些分位数最小化。最常用的 LQS 方法是使中位数平方最小化。

(2) 向正态分布逼近

① 费希尔变换公式：费希尔变换定义为 $F(r)=\text{arctanh}(r)$ 。

如果 $F(r)$ 是 r 的费希尔变换，并且 n 是采样大小，那么 $F(r)$ 可近似为一个标准差为 $1/\sqrt{n-3}$ 的正态分布。

② Z 得分：数据的 Z 分数（也叫标准分数）均值为 0，并有单位标准差的正态分布。 $F(r)=(r-\text{mean}(r))/\text{stdev}(r)$ 。

(3) 限制性方法

① 截取法：限制每只股票投资资金在总投资资金上所占的最大比例。

② 温莎化：温莎化是统计数据的变形，通过限制统计数据的极值来减少假的极端值带来的影响，卢梭和勒罗伊于 1987 年曾对此做过讨论。

2. 改善稳定性的高级方法

当我们要从一个有限数据集中预估数值时，重采样方法会非常有效，比如自引导法、交叉验证法等。

3. 结论

Bertsimas 等人在 2004 年曾提到，提高稳定性可以在极端市场环境下显著改善 Alpha 的测评指标。

第 17 章

自动搜寻 Alpha

作者：Yu Huang

随着超级计算机和复杂统计模型的发展，自动化搜寻 Alpha 信号成为可能。自动化的 Alpha 搜寻就是利用计算机算法从海量云数据中找到 Alpha 信号。只要使用得当，它可以显著提高 Alpha 搜寻的效率，一日之内即可得出数以千计的 Alpha。这是有代价的：找到的所有信号并非都是真的 Alpha。很多看起来很棒的 Alpha 信号都是噪声，而不具备任何预测能力。因此，必须在输入准备、搜寻算法、信号测试算法等方面做出特别努力，以改善这些 Alpha 的稳定性。

总体来说，一个自动化搜寻有 3 个组成部分：输入数据、搜寻算法，以及信号测试（如图 17.1 所示）。

首先，我们选择一组有意义的财务数据（如价格、利润、新闻等）作为输入预测变量。接下来，对输入预测变量进行预处理以去除极端值，再输入搜寻算法。其次，我们选择一个目标函数 Y ，用来表征未来的股票回报或其

变体。再次，适配算法会为—组预先选择的函数族（最简单的例子是线性函数）找到—组参数，以此来模拟目标函数。最后，对适配好的函数进行稳定性测试。

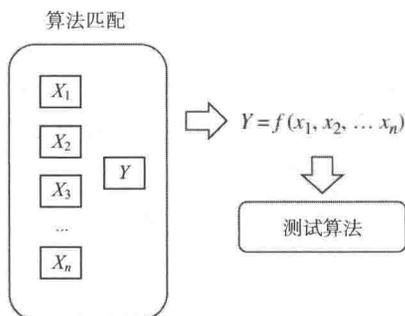


图 17.1 算法匹配

搜寻算法的效率和可靠性各有不同，而且也没有哪个算法比其他算法更可靠。然而，有时在数据准备和信号测试中有一些技巧，会使 Alpha 更加可靠。

1. 将输入数据比率化是一个好办法

好的股票收益预测指标应该对所有股票具备同一性和可比性。因此，原始财务数据（如价格或利润）通常因为在不同股票上不具备可比性，所以不是一个好的预测指标。例如，利润数据并不是跨板块可比的，但是利润与营收的比值却是可比的。有两个常见的方法可以使一个财务变量比率化。第一种方法是用该变量的当前值比其历史值。这一操作描述了该变量的近期变化，往往包含预测信息。第二个方法是将该变量比上同类的一个相似变量，例如净收入和营收的比值，净收入和营收对于所有股票都是有可比性的。

2. 输入数据不应取自太多的类别

众所周知，给函数添加太多输入变量会使其在样本内测中有很好的表现，但是通常会因为过度拟合而导致其损失预测能力。除此之外，一个常常被忽略的问题是数据类别的数量。配置太多变量很可能会导致过度拟合，但是适配来自很多不同类别的变量则更糟糕。这里，类别指代数据的类型和来源。例如，一些通用的类别包括（但不限于）价格交易量、分析师评级、基本面数据、新闻、内幕交易等。每个类别的数据有其自身特征频率，例如，基本

面数据通常呈现出清晰的季度周期性，价格交易量分布稳定，而内幕交易数据通常是随机的。如果模型包含来自很多不同类别的数据，那么其复杂度将大大提高，而该模型也将对数据中的噪声更加敏感。

3. 测试周期并非越长越好

增加测试周期会增加数据点的有效数量，并增加计算结果的统计重要性。然而，这里有一个假定，数据之下所隐藏的潜在驱动逻辑是不变的。对于金融市场这并非总是正确的。市场玩家和他们的行为千变万化，从而改变了市场动态。因此，在选择输入数据的长度时需要折中。如果测试周期太短，我们得到的数据太少，就会对计算结果信心不足。如果测试周期太长，驱动逻辑可能已经发生了变化，就会让计算结果变得不可靠。

4. 敏感度测试和重要性测试很重要

一个好的 Alpha 信号应该对噪声不敏感。最常用的稳定性测试技术包括在随机的数据集上，或者在每个独立的股票板块上测试来自不同周期、不同时段的数据。另外，每个输入数据都应该对结果有显著作用。最简单的重要性测试方法就是去除一个输入变量，然后检查结果是否发生显著变化。如果每个输入变量都作用显著，那么我们对该信号会更有信心。

第 18 章

Alpha 研究中的算法和特殊技巧

作者：Sunny Mahajan

Alpha 研究涉及对海量原始数据的处理。有几个来自于工程学（如机器学习和数字信号处理）的算法和技巧可以用来提取有用的模式、关系和特征，从而为生成 Alpha 服务。

在 Alpha 研究中，我们频繁地处理弱的分类器和预测器。根据机器学习中的提升（boosting）概念，可以在很多场合通过学习一个合适的权重函数，从若干弱学习者中建立一个强健的学习者。

Alpha 研究还需要与时间序列数据打交道。来自数字信号处理的滤波概念，可以有效地对时序数据降噪，并将时序数据分解为趋势和周期分量。

Alpha 研究中的另一个有益的技巧是特征提取。诸如主成分分析（PCA）之类的算法可以帮助降低特征空间的维度。

1. 提升算法

提升算法的基本思想是通过将很多相对弱的和不太准确的规则组合起来，创建一个高度准确的预测规则。被研究和应用得最广泛的助推算法之一就是 Freund 和 Schapire 于 1999 年提出的 AdaBoost。

AdaBoost 使用一个权重函数将若干弱分类器组合起来，通过迭代学习，形成一个强分类器。在每次迭代中，检查之前被分类器错误分类的实例，对能进行正确分类器进行奖励。只要每个分类器的表现比随机猜测稍好，最终的模型经验证就会收敛为一个强分类器。

2. 数字滤波

数字滤波是对原始时序数据进行数学运算，以稀释或放大信号的某个特性。

数字滤波最初和最重要的应用是平滑滤波（即对时序数据中的高频分量进行抑制），最简单和最广为使用的数字滤波是简单移动平均滤波。然而，平滑滤波引入了时滞，而时滞在交易过程中则是极度不利于盈利的。因此，可以使用一些替代性滤波技术，以最小时滞来达到要求的平滑等级。

数字滤波另一个常见的应用是将时序点分解为趋势和周期分量，由 Mohr 在 2005 年详细论述过。趋势抽取涉及对高频分量的抑制（低通滤波），而周期提取涉及对低频分量的抑制（高通滤波）。

巴特沃斯于 1930 年提出的巴特沃斯滤波是一种常用和有效的低通滤波。高通滤波则可以通过从原始时序数据中减去低通滤波后的数据得到。同时，低通滤波和高通滤波可以串联起来以选择（带通）或排斥（带阻）原始时序数据中一个特定的频率波段。

3. 特征提取

特征提取将数据从高维度空间变换到较低维度空间。在量化金融学中一种常见的特征提取算法是 PCA。

PCA 是一种正交线性变换算法，将数据基于新的矢量基进行表达——并且新的矢量基可以解释原始数据的绝大多数变化。

在 Alpha 研究中应用 PCA 包括在高维度数据中寻找模式、对观测数据进行聚类（采用在新矢量基下的投影之间的距离作为相似性的量度）、主成分回归，以及收益率曲线分析。

第三部分

拓展研究

第 19 章 新闻与社会媒体对股票收益的影响

第 20 章 由股票期权市场所反映的股票收益

第 21 章 动量 Alpha 简介

第 22 章 财务报表分析

第 23 章 机构研究 101

第 24 章 期货交易简介

第 25 章 远期外汇和期货中的 Alpha

第 19 章

新闻与社交媒体对股票 收益的影响

作者：Wancheng Zhang

股票价格受新闻影响。近年来，社交媒体对股票价格的影响程度越来越大，然而利用新闻来确定 Alpha 却是一个非常具有挑战性的任务。非结构性数据通常包括文字与多媒体信息，因此计算机无法直接理解这些信息。本章将简略描述如何利用新闻和社交媒体发现 Alpha。

1. 新闻

对于机器来说，很难将新闻中的信息进行精确解析和转译，得到直接而有用的信息。金融领域的统计套利与其他领域的相似，算法通常在计算速度和覆盖范围方面具有很大优势，然而精确度却会随着二者的提高而降低。如今，许多交易公司可以在 1 毫秒内完成对新闻的解析并立刻做出交易决策。重大新闻往往会立刻引发大幅度的价格波动，但是由于超量效应的存在，此类价格波动持续一段时间后又会发生反转。

自 2007 年起,对新闻与社会媒体复杂语言分析方法的应用已经从研究阶段进入成熟产品解决方案阶段。专业的数据供应商可以运用复杂的算法解析新闻并及时传送解析结果。新闻分析与新闻情绪同时被买方机构与卖方机构运用在 Alpha 构造、交易和风险管理等方面。

2. 情绪

简单来说,情绪可以衡量新闻的质量。可以将新闻情绪分为基本的几类:好、坏和中性。高级的情绪分析方法或者工具可以分析出新闻中更为复杂、精细的情绪,比如,愤怒、惊异或者超出预期。

新闻情绪的构建通常包括自然语言处理过程与统计/机器学习算法(如朴素贝叶斯法(Naïve Bayes)、支持向量机等算法)。

情绪程度通常需要进行标准化处理,通过打分(如 0~100 分)的方法可以对它们进行横向对比。分数越高,意味着新闻越好、越正面,分数越低,意味着新闻越不好、越负面,若分数接近 50 分,则意味着新闻处于中立状态。根据情绪分数确定交易决策的典型方法如下:

- 如果 A 股情绪>70 分,则做多 A 股;
- 如果 B 股情绪<30 分,则做空 B 股;
- 对于无新闻报道的股票,按照中性处理。

情绪对于风险管控也是有用的。比如,一个基金经理可能会因为预期外的消息而直接缩减头寸,或者因为此新闻情绪的分数或者新闻的频率而重新估算投资协方差矩阵。

个体的新闻情绪也会影响到市场集体情绪、季节性或者其他的时间因子(比如在收益之前或者之后)。在中性策略中,最好参照中性情绪。

3. 新鲜度

新鲜度可以衡量新闻是最新消息还是某些旧新闻的后续报道;数据提供商可能会将一个长报道分解为多个部分。另外,在第一次报道之后新闻可能被多次进行后续更新。新鲜度低的新闻对市场的影响较小,因为新闻刚出来时的影响已经在市场中有所反应了。如果我们将新闻看作时间序列上的事件,那么其新鲜度通常与事件发生的时间间隔成反比。

4. 相关度

新闻可以对多只股票产生影响，相关度可以衡量新闻对特定股票的影响程度。一些新闻关注的是公司主体，如公司利润或公司行为，此类新闻和企业的相关度通常较高。一些关于工业与宏观经济的新闻与单只股票的相关度则较低。银行业的一般性新闻则会影响许多银行股票；然而与此相反的是，像苹果公司新产品的新闻则只会影响苹果公司与其竞争者（如三星公司），此类新闻表现出来的相关度是：与苹果公司的相关度较高，与三星公司的相关度较低。换言之，相关度将新闻与具体的股票联系起来。

5. 新闻分类

除了可以将新闻分为“好新闻”和“坏新闻”之外，还可以进一步对新闻进行分解归类，以对其进行分析和利用。比如，盈利这一分类可以包括所有与盈利相关的报道，如盈余公告、盈利预测、盈利预测调整、盈利指导、盈利电话会议、盈利日历等。当然，分类也可以更加精细化，如公司法务问题。在对新闻进行分类时，有几个方面需要特别注意。首先，不同类别的新闻对市场影响的持续时间不同。某些类别的新闻可能对企业的估值有长期影响，而另一些类别的新闻则只会在短期内引发股票价格波动。其次，市场在不同的时间点上对新闻有着不同的偏好。分类轮换策略就是利用了市场对新闻风格偏好的变化。最后，将新闻分类更益于将不同类型的新闻与其他信息结合使用以产生 Alpha。例如，可以将盈利类的新闻和分析师盈利预测调整结合起来。

6. “预期内”新闻与“预期外”新闻

一个看似正面的新闻，如果符合市场预期并已经反映在价格中，那么将不会引起正向的价格波动。例如，一则消息称：“公司盈利相对于 2013 年有 150% 的大幅增长”，分析此条新闻通常会得出正面情绪。然而，如果此前市场一致预期此家公司盈利将增长 200%，新闻中的 150% 的增幅实际上是低于市场预期的，那么会引起价格下跌。因此，有必要将新闻与市场共识与市场预期结合在一起进行分析。文字分析与时间分析同样有助于辨别新闻是日常更新还是一些非比寻常的事件。突发和预期外的新闻和事件通常会引起比较大的价格波动。

7. 新闻标题与正文

新闻标题中通常包含最有用的信息，语法格式也是制式的，因此，新闻标题很容易被解析与理解。另外，正文包含更详细的信息，然而处理起来相对困难。研究人员曾经从一份学术研究报告中得出一个非常有趣的结论：一段文字中最有用的信息通常会被总结在第一句或最后一句话中。同样的道理，我们也可以更加关注文章的首段或者末段章节，以及章节中的首句或者末句。段落和句子结构同样是有关注的点。

8. 没有消息就是好消息吗

有句老话“没有消息就是好消息”。从某种角度讲，这个说法是正确的，因为新闻意味着变故、重大事件，或者不同于以往的现象。同样，新闻通常与更剧烈的波动、更大的交易量和分析师预测调整相关，而且更多新闻往往会随之而来，这会给公司带来潜在风险。越来越多的机构投资者在风险管理中使用新闻，那些有过多新闻的公司有可能会被机构投资者缩减头寸或者直接空仓，低回报便由此产生。

9. 新闻惯性

如果新闻中的信息没有立即在市场中引起充分反应，那么股票价格随后会有漂移或有惯性。对于关注度小的公司的新闻和意料之外的新闻，此类效应会更强烈。而对于大公司的新闻和预期中的新闻来说，初始超量后的价格反转很快便会被人们注意到。

10. 学术论文

从 2000 年起，新闻对于股票回报的影响就已经成为热门课题。对这方面的研究课题包括：情绪的累积与释放，运用新闻计算 Beta，领先对新闻做出反应的股票，文字分析中权重分配的方法，根据盈余公告日的股票回报率证实新闻的真实性，没有新闻就是好新闻，对新闻较敏感股票表现较好的探讨，根据交易量确定新闻的真实性，股票市场中新闻覆盖程度的差异性，新闻之后的惯性、超量、反转效应，新闻与分析师预测修正的关系等。

11. 社会媒体

2013 年 4 月，美联社在推特（Twitter）网站上发文称白宫发生爆炸事件，

总统奥巴马受伤。此次攻击事件为虚假言论，然而却立刻在市场中引发了巨大波动。当天从 13:08 到 13:10，道琼斯工业平均指数（Dow Jones Industrial Average Indices, DJIA Indices）下降了 100 多点。当然就像下跌速度一样，回弹也是在几分钟内完成的，十分快。由于推特的此次虚假消息，市场在 2 分钟内损失了 1360 亿美元。

此次事件清晰地勾勒出社交媒体在交易算法中应用的广泛程度。许多数据提供者也在此次事件中抓住了机会。比如 Gnip 公司（后来被 Twitter 公司收购），每天提供数以百万的社会信息，它们同样会为售卖情绪分析产品的第三方提供数据，此类情绪分析产品已被许多对冲基金使用。

12. 什么样的社交媒体平台最有影响力

对于社交媒体平台，大家普遍的选择是 Twitter，因为 Twitter 上的信息最容易与股票对应起来（通过 @ticker 的方法即可实现查找功能）。当然也有研究基于在线论坛、专业投资者/交易者的博客、Facebook 和 Wikipedia 的投资者。

13. 学术研究

首篇相关的学术论文是“Twitter 情绪可以预测股票市场”。论文中称：利用 Twitter 情绪预测每日 DJIA 收盘价的上涨或下跌，准确率可以达到 87.6%。自此之后，在以下方面陆续有研究跟进：多种社会媒体的预测力量，社交媒体在单只股票上的应用，对社交媒体噪声的讨论，从社会名人的 Tweet 和相互回复中挖掘重要信息，社交媒体情绪与长期公司价值的关系。大多数论文可以在 SSRN 上找到。

14. 运用社交媒体进行预测所面临的挑战

社交媒体是量化基金研究中的热点领域。运用社交媒体信息中的情绪分析将面临若干挑战。首先，社交媒体记录的信息数量巨大，更新速度极快。其次，社交媒体内容通常语法形式随意、散乱，一条 Twitter 评论中可能包含很多缩写，格式混乱。这就加大了语言处理的难度。最后，如何找到原始的重要记录？相当多的社交媒体内容是对某条新闻的回复或评论，与初始的社交媒体内容相比，这些评论的引导性和效应性就差了很多。社交媒体中存在许多虚假信号，因此，运用社交媒体进行预测面临着诸多困难。

第 20 章

由股票期权市场所反映的 股票收益

作者：Swastik Tiwari

在金融学中，期权是用来反映卖方（期权签署者）和买方（期权持有者）双方合同的金融衍生物，它赋予买家在行权日或者行权日之前，以特定成交价买入或者卖出标的资产或者金融工具的权利。如果期权持有者行权的话，卖方就相应地有义务去履行此交易（买或者卖），买方也要向卖方支付行使这项权利对应的费用。看涨期权给予行权人以固定价格买入的权利，因此买方希望标的资产价格上涨；看跌期权给予行权人以固定价格卖出的权利，因此买方希望标的资产价格下跌。投机者利用期权杠杆对标的物进行投机，而套期保值者利用期权来降低持有资产的风险。

目前在美国有 10 多个期权市场，由纳斯达克 OMX 集团、BATS 全球市场、德意志交易所的 ISE 分部、加拿大多伦多交易所的波士顿期权交易所、

迈阿密国际控股公司、纽约洲际交易所和芝加哥期权交易所控股公司所经营。股票期权越来越受到散户投资者和机构投资者的欢迎。

期权清算组织（Options Clearing Corporation, OCC）在其 2013 年年度报告中公布了如图 20.1 ~ 图 20.3 所示数据。

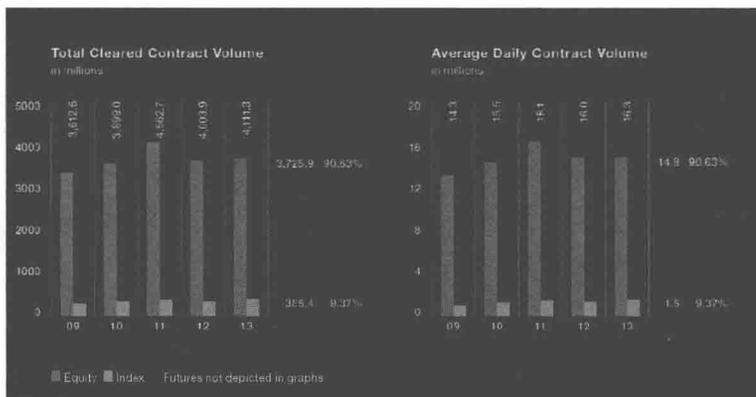


图 20.1 2009—2013 年期权票据和日合约交易量

来源：OCC2013 年年度报告，OCC 芝加哥，伊利诺伊州，第十四页。获得 OCC 授权。版权所有。OCC 及其相关单位不为本书的内容和其他非 OCC 提供的素材负责。OCC 不保证信息的准确性、科学性、完整性和可用性，并对错误、遗漏和基于此信息的所有结果不承担任何责任。

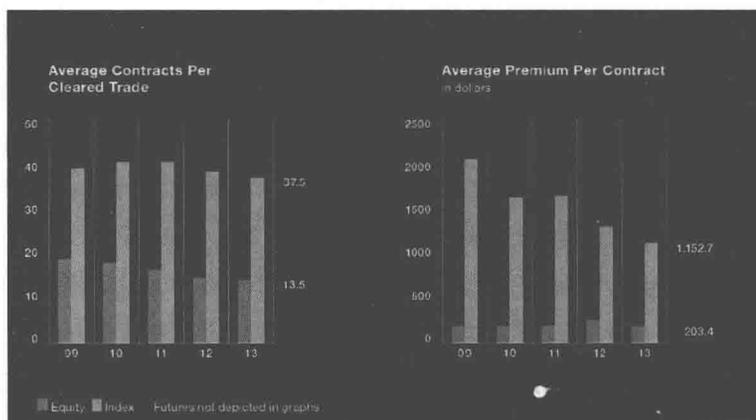


图 20.2 2009—2013 年期权买卖日交易量

来源：OCC2013 年年度报告，OCC 芝加哥，伊利诺伊州，第十四页。获得 OCC 授权。版权所有。OCC 及其相关单位不为本书的内容和其他非 OCC 提供的素材负责。OCC 不保证信息的准确性、科学性、完整性和可用性，并对错误、遗漏和基于此信息的所有结果不承担任何责任。

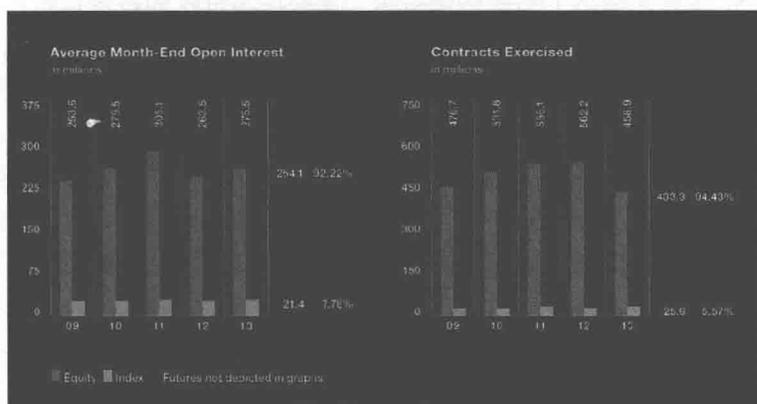


图 20.3 2009—2013 年期权持仓量和合约执行

来源：OCC 2013 年年度报告，OCC 芝加哥，伊利诺伊州，第十五页。获得 OCC 授权。版权所有。OCC 及其相关单位不为本书的内容和其他非 OCC 提供的素材负责。OCC 不保证信息的准确性、科学性、完整性和可用性，并对错误、遗漏和基于此信息的所有结果不承担任何责任。

OCC 期权统计数据如表 20.1 所示。

表 20.1 OCC 期权统计数据

| OCC 期权合约类型 | 数 量 | 占 比 |
|------------|---------------|--------|
| 股票 | 3,725,864,134 | 90.63% |
| 指数/其他 | 385,411,525 | 9.37% |
| 合计 | 4,111,275,659 | 100% |

纳斯达克官网显示美国期权市场 2015 年 2 月份数据，如图 20.4 所示。

股票期权市场为预测股票收益提供了很多有用的信息。股票期权有助于价格发现，因为它们可以让交易者根据信息对价格的正反影响和影响程度调整交易策略。伴随着这种调整，股票期权的杠杆作用产生了额外的激励措施，从而产生私人信息。通过这种方式，股票期权交易可能会比资产本身的交易更精练、更准确地反映标的资产的价值。了解为什么股票期权对价格发现会有影响以及会有怎样的影响，才能理解信息对于资产定价的影响。

1. 波动率偏离

其中一种有用的信息来自股票期权中隐含波动率。期权合约中的隐含波

动率是指金融资产理论上的波动率数值。该数值通过将期权市场数据代入期权定价模型（例如 Black-Scholes 模型）反推得到。对于股票期权，隐含波动率和协议价格是倾斜的曲面。波动性偏离是价外期权、平值期权和价内期权之间隐含波动性的差异。波动性偏移受情绪和供需关系的影响，能解释基金经理选择看涨期权或看跌期权的原因。在股票期权市场中，偏离经常发生，因为操盘者更多地倾向于看涨期权，而不是看跌期权，如图 20.5 所示。

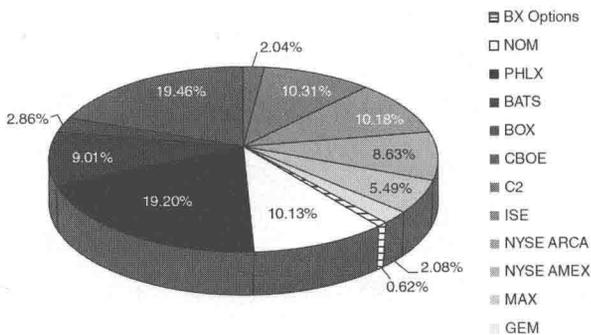


图 20.4 美国 2015 年 2 月期权市场比例

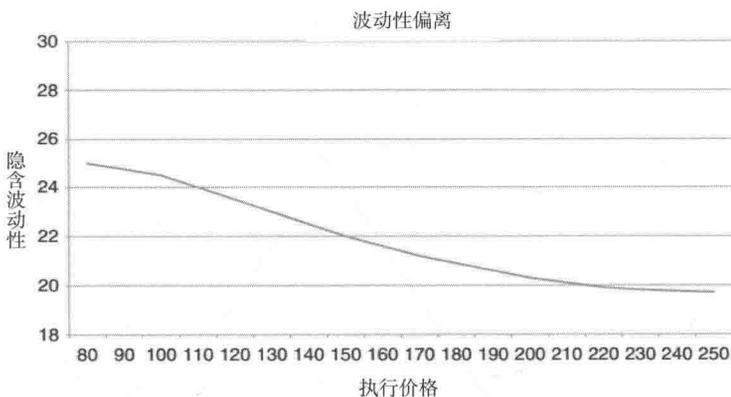


图 20.5 股票期权市场样本波动性偏离

在 Jin 等人（2012）的论文“shape of observed volatility skew and its predictiveability to the buying pressure due to the information possessed by option traders”中，提到了对这个方向的很多研究，比如 Bollen and Whaley（2004）、Garleanu 等人（2009）、Van Buskirk（2011）、Xing 等人（2010）、Bradshaw 等

人 (2010), Bollen、Whaley (2004) 和 Garleanu 等人 (2009) 发布了“观察中的波动性偏离的形状和对于购买压力的预测能力, 是来源于期权交易者所拥有的信息”。Bollen 和 Whaley (2004) 发现“日间隐含波动率的同期变化是被净购买压力的变化所驱动的”。获得利好消息的期权交易员会买进看涨期权或者卖出看跌期权, 以使得价格和看涨期权的隐含波动率相对于看跌期权上涨。同样, 获得利空消息的期权交易员会买进看跌期权或者卖出看涨期权, 以使得价格和看跌期权的隐含波动率相对于看涨期权上涨。因此, 当期权交易者获得关于一系列利空消息可能发生的信息时, 对价外看跌期权的需求相对于看涨期权的需求增加, 从而增大波动性偏离。研究者已经证明, 波动性偏离和股票市场中公司收益存在负相关关系, 这个发现和波动率反映利空消息是一致的。Xing 等人 (2010) 指出, “在股票的成交期权中波动率越大, 不佳绩效越明显”。他们用高偏离值来定义期权标的股票的不佳绩效, 针对那些期权的标的股票, 每年风险调整基准最小都在 10.9% 之上。最近, 也有一些研究聚焦在波动性对于极端利空事件的预测能力上。Van Buskirk (2011) 发现, 高波动性偏离预测价格短时间内的跳跃伴随着收益公告, 在稍长的时间窗口中并没有收益公告, 而在经营收益预测或者股息声明附近时也不会预测跳跃。Bradshaw 等人 (2010) 公布, “波动性偏离的预测能力和现金流风险的计算复杂性是相互作为补充的”。根据作者的观点, 利空信息比利多信息具有更多的信息优势, 体现出对波动性偏离的预测, 期权市场的预测能力实际在不同潜在影响力的消息下都是充分适用的。

2. 波动价差

看涨/看跌期权平价公式表明, 在完全市场中, 对于不分红股票的欧式期权适用下列等式:

$$C - P = S - D.K$$

其中 C 和 P 分别为当前的看涨期权和看跌期权价格, D 为贴现因子, K 为协议价格, S 为现货价格。对于美式期权, 允许提前行权, 等式方程演化为非等式的形式: $S \geq D.K + C - P$ 。从这些关系可以看出, 具有相同协议价格的欧式看涨和看跌期权应该有相同的隐含波动率, 而美国的看涨和看跌期权因为提前行权费用而会有波动价差。(见 Hull, 2008)。

然而，Ofek 等人（2004）证明“波动价差”不能完全由提前行权费用解释。Ofek 等人（2004）、Cremers 和 Weinbaum（2010）发现，这种波动性价差预测了未来股票的收益。比如，Cremers 和 Weinbaum（2010）发现有高波动性价差的股票每周都比有低波动性价差的股票优胜大概 50 个基点。Bollen、Whaley（2004）和 Garleanu 等人（2009）将波动性价差的预测能力归因于基于需求的期权模型。更高的波动价差表明大部分超额需求是基于看涨期权而不是基于看跌期权的，表明期权交易商获得有关正面消息的信息。因此，波动价差可以被视为通过在期权市场捕获整个市场的净购买压力，表明期权交易者掌握的新闻的本质（正面的或者负面的）和潜在影响。

3. 期权交易量

股票期权的交易量也可以提供关于未来股票收益的有用信息。在“The Option to Stock Volume Ratio and Future Returns”一文中，作者约翰逊和 So（2011）就聚焦在期权和标的股票的交易量对标的价格的影响上。作者提供了理论和实践证据表明 O/S，即整个期权市场的交易量（包括看涨和看跌）和整个证券市场交易量的比例，代表着知情交易者掌握的秘密信息。O/S 的方法最早被 Roll 等人（2009）发现，当时他们的研究表明“O/S 比率在截面数据和时间序列上的变化，可能是由掌握着秘密信息的知情交易者所驱动的”。作为这个发现的延伸，作者约翰逊和 So（2011）在“The Option to Stock Volume Ratio and Future Returns”一文中判断了 O/S 和未来收益的关系，并发现了低 O/S 的公司收益优于高 O/S 的公司。他们的方法包括每月月底根据 O/S 对公司进行排序，并且计算出一个策略的平均收益，包括对于高 O/S 的空头和低 O/S 股票的多头。然后他们持有这个组合一个月，该组合提供的平均风险调整后每月对冲回报率为 1.47%。作者认为，在证券市场中，O/S 和未来证券收益的负相关性导致了卖空的成本，因为资本限制和证券卖空成本限制，知情者在知道利空消息的情况下，会比知道利多消息的情况下更加频繁地交易期权。作者表示，“O/S 预测意外收益、标准化无须解释的收益，以及下一个季度的财报公布的异常收益。”相同的基于 O/S 方法而构建的投资组合，也包含了关于未来收益公告的信息，而这些都将在后续持有的月份发生。他们声称这与反映秘密信息的 O/S 保持一致，而当新闻被公开时，私人信息被纳入股票价格。此外，他们宣称“模型还可以预测在卖空成本高或者期权杠杆低的情

况下，O/S 是一个强烈的信号”，所有的这一切都经过了数据验证。

4. 期权持仓

持仓量反映了期权的未平仓合约数目，Fodor 等人（2010）在“Do Option Open-Interest Changes Foreshadow Future Equity Returns?”一文中，分析了持仓量变化和未来收益。他们表明，当标的资产是短期看涨行情时，期权交易者购买相对较多（较少）看涨期权（看跌期权）。同样，当标的资产是短期看跌行情时，期权交易者购买相对较多（较少）看跌期权（看涨期权），这种行为导致那些能影响未来收益的信息的证券发生总持仓量的变化。他们声称，知情投资者通过增加多头买入（卖出）头寸体现他们看涨（看跌）的观点。

在他们的实证调查中，论证了最近总持仓量变化的等级和未来标的资产收益的强负相关性。最近持仓量增加的公司会劣于最近持仓量减少的公司。他们发现，如果看涨期权持仓量变化，那么相反的而且弱很多的关系也会存在。随后他们发现，最近看涨期权持仓量变化和看跌期权持仓量变化的比值对于未来资产收益是最有效的预测，从大比值变化会紧跟着未来资产大收益来看，这种关系是正向的。

Fodor 等人（2010）论证了由于相对小的初始化费用需求，知情交易者[像之前被 Black（1975）所说的]偏好于通过期权反映他们的观点（看涨的通过长期看涨头寸，看跌的通过长期看跌头寸）。Fodor 等人（2010）进一步论证，期权和证券市场在现实中不同的信息化导致了价格体现信息的不同速度。

第 21 章

动量 Alpha 简介

作者: Zhiyu Ma

动量 Alpha 是 Alpha 的一种,而且是占比很大的重要部分。此处我们将动量 Alpha 定义为其平均的 Alpha 值(股票的预期中性收益)与该股票或某些其他股票的往期收益具有共同信号。

1993 年, Jegadeesh 和 Titman 发表了第一篇关于股票价格动量的论文 *Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency*, 描述了在此前 3 ~ 12 个月中的盈利者或亏损者将在未来继续盈利或亏损。在此之后, 其他的研究人员就同样的现象进行了研究, 其研究结果确认了上述理论在国际上大多数股票市场及其他金融市场也是适用的。近年来, 此种策略带来的收益锐减, 尤其是在 2008 年金融危机时遭受了大幅回撤。之后, 许多论文尝试在不变动该策略的本质的前提下修正这个原则, 以达到增加潜在收益, 减少回撤的目的。当前, 动量 Alpha 仍是学术界内活跃的研究领域。

研究人员尝试通过提供合理的解释来解释动量 Alpha 的适用性。其中一个

被广为接受的理论是投资者对新信息反应不足。在非完全有效市场中，新信息需要一定的时间被消化和评估，伴随着股票价格的逐渐调整，当我们研究某些事件对市场产生的影响时，上述理论发挥着作用。当公共消息被披露时（如发布年报时）股票价格会产生波动；披露的信息越重要，其产生的动量影响越剧烈。有一个有趣的现象：股票动量往往在新消息尚未披露前（换句话说，当公共消息还属于内部消息时）就已开始。

不仅投资者对新信息反应不充分，股票分析师也是如此。在行业压力下，分析师缺乏做出突出（但是可能错误）的预测的能动性，他们更倾向于逐步调整其对于未来的收入及目标价格的预测。因此，当市场中的投资者依据分析师的推荐做出自己的投资决策时，其整个决策行为未能有对信息的充分反应，这在某种程度上解释了价格的动量效应。

因子动量是另一种动量。在套利定价理论中，数以千计的股票回报来源于相对而言少得多的因子组合。股票因子的敞口经常变化。但是，与单一股票相比，单个因子的回报却稳定得多（至少在一段时间内或某种市场条件下如此）。基于所有因子回归测试的 Alpha 都假设因子的收益受动量影响，因子动量的另一种应用是，人们在市场上交易那些因共同基金的基金经理投资而热门的因子。

群体动量是动量的另一种类型，群体动量经常与“共同运动”现象相关联，该现象表现为：相关的股票价格（商业关联度高或回报受同一因子影响大）倾向于同时变化。这里所说的“同时变化”不是指在同一时间同时上涨或下跌同一幅度。在这个群体中通常有几只股票价格率先变动（有可能受到新信息的影响），然后该群体中的其他股票价格会跟随这几只股票变动。我们将这种现象称为“超前滞后”效应。延迟的股票享有动量带来的盈利。除了相关股票之外，其他的某些相关群体（如供应链上的其他行业）也会将某一群体的回报传递给其他群体。

前面提到，第一篇讨论动量的论文并没有提到“动量”这个词，但提出了“市场有效性”。事实上，动量效应与非有效性相关。在流动性市场（市场中拥有更多有效性高的股票）中研究动量 Alpha 是颇具挑战性的工作，这要求我们要更加深入地去研究。

第 22 章

财务报表分析

作者：Paul A. Griffin

一家公司会根据季度和年度时间表公布财务报表。由 Graham 和 Dodd (1940) 提出并推广的证券分析是指，对每家公司的财务报表进行深入研究，并以此为依据推断该公司是否具有赚取超额收益的投资潜力。这一分析被以 Warren Buffet 和 Berkshire Hathaway 为代表的“基本面价值法”投资者和对冲基金所采用。基本面价值法与 Lefevre (1923) 所讨论的股票价格走势变化不同，也有别于其他技术分析方法（如以历史价格动量预测未来股价的方法）[参见 Chan 等人 (1996) 的参考文献]。

财务报表分析试图系统地衡量从收益报表中获取的因子，并确定其预测未来回报的能力；随后，它对公司进行排序、分类、筛选，以创建具有更好投资收益的证券组合（请注意，在这种情况下，初步季度公告和随后提交的声明可能不同；经审计的年度财务报表通常被认为是最权威的）。

通过财务报表分析可以得到，Alpha 最初是存在争议的，原因是在有效的市场中，所有可获得的信息都已反映在市场价格中了。然而，通过对基于多元

且合理选择的收益信息构造的多因子模型的后续研究发现,从长期看,有很多违背有效市场假说的情况出现。如 Piotroski (2000)、Chan 等 (2001)、Mohanram (2004)、Abarbanell 和 Bashee (1977)、Beneish 和 Nichols (2009) 等。

传统财务报表分析的典型产出是一个只有多头仓位且低换手率的证券组合,组合中包括通过基本面分析法挑选出的数量有限的个股。

现代财务报表分析使用财务报表来确定证券组合中的股票头寸,其可能受到诸如净值、 β 和行业中性度等其他因素的限制,但不会特地缩小待选股票池。用更现代的话说,很多人可能会认为这些因子是多因子回归分析或机器学习算法中用于预测的数据。

本章不是为了将基本面分析整体呈现给读者,因为市场表现相关的额外数据、分析师建议、盈利意外等信息可以和具体的财务报表相结合来得到额外的 Alpha。关于基本面分析的更全面的概述已在本书第 7 章中提供了。本章旨在介绍财务数据使用的基本原理,并希望为未来的研究方向提供一些启示。

1. 资产负债表

会计的基本方程:

$$\text{总资产} = \text{负债} + \text{权益}$$

由表 22.1 中的资产负债表所示。

请注意,资产负债表展示的是某一时点上公司的情况,就像给公司的运营情况拍了一张快照。

通过对比不同时点的快照,读者可以发现引起一家公司流通股价变动的的原因。

表 22.1 资产负债表

| 资产负债表 YYYYMMDD | |
|----------------|----------|
| 资产 | 负债+所有者权益 |
| 流动资产 | 流动负债 |
| 其他资产 | 长期负债 |
| 无形资产(商誉等) | |
| 总资产 | 所有者权益 |

总资产通常被用作归一化因子来比较不同公司的常规因子，还可以被用来比较同一家公司在不同时间点的表现。对于美国公司，包括商誉这样的无形资产，其定义为公司为另一家公司支付的超过账面价值的费用。虽然商誉包含品牌等项目，但是在统一各因子的过程中，应该测试是否给商誉的价值打折。

如 Piotroski (2000) 指出的，1976 年至 1996 年间，从资产负债表构建的一些与未来收益正相关的知名因子有：

- 流动性增加（流动资产覆盖流动负债）；
- 提高资产周转率；
- 不发行股票；
- 较少长期负债。

2. 损益表

损益表将资产负债表从一个时期更新到下一个时期，如表 22.2 所示。大多数公司使用权责发生制记账，所以损益表并不反映现金的流动，而是反映了支付的义务。例如，如果公司签订多年的产品供应合同，则在每项合同义务被履行时，都会确认收入，而不会在对方完成转账时确认。

表 22.2 损益表

| 损益表 YYYYMMDD | |
|--------------|-------------|
| 净销售额 | A |
| 利息收入 | B |
| 主营业务成本 | C |
| 运营费用 | D |
| 所得税 | E |
| 毛利润 | $A-C$ |
| 运营收入 | $A-C-D$ |
| 总收入 | $A+B$ |
| 净利润 | $A+B-C-D-E$ |

根据 Piotroski (2000) 的研究，基于损益表的以下因子与 1976 年至 1996 年间的美国市场预期收益率呈正相关关系：

- 净利润 >0 ;
- 总资产收益率提高;
- 毛利率提高。

3. 现金流量表

现金流量表, 如表 22.3 所示, 描述了两个时期内, 公司现金余额变动的情况。

与 1976 年至 1996 年间美国未来收益呈正相关的因子有:

- 经营现金流 >0 ;
- 经营现金 $>$ 净利润。

表 22.3 现金流量表

| 现金流量表 | |
|--------|---------------|
| 期初现金余额 | A |
| 经营性现金流 | B |
| 借款 | C |
| 抛售股票 | D |
| 购买支出 | E |
| 税费 | F |
| 现金流 | $B+C+D-E-F$ |
| 期末现金余额 | $A+B+C+D-E-F$ |

4. 成长性

上述因子侧重于发掘公司绩效质量。不过, 投资者也可能对增长前景感兴趣。

Chan 等人 (2001) 和 Mohanram (2004) 对成长型股票做了类似的回归分析。他们发现了 1979 年至 1999 年间专门与成长型股票 (具有较低账面价格/市值比率) 的未来回报相关的信号:

- 净利润/总资产 $>$ 行业中位数;
- 现金流/总资产 $>$ 行业中位数;

- 净利润方差<行业中位数；
- 总收入方差<行业中位数；
- 研发费用/总资产>行业中位数；
- 资本性支出/总资产>行业中位数；
- 广告费用/总资产>行业中位数。

研发、融资和广告费用是表 22.4 所示的运营费用项目内的单独项目。增长中的公司将会基于对未来销售收入增长的预期加大对这些项目的投入。

表 22.4 损益表—运营费用

| | |
|--------|-----------------|
| 研发费用 | $D1$ |
| 资本性支出 | $D2$ |
| 广告费用 | $D3$ |
| 其他运营费用 | $D4$ |
| 总运营费用 | $D=D1+D2+D3+D4$ |

5. 公司管理

管理层使用某些指标来监督和提高公司业绩，市场参与者将会倾向于通过观察这些指标的变化来判断股票的收益。根据 Abarbanell 和 Bushee (1997) 的说法，与未来收益呈正相关的指标是：

- 每单位销售对应的库存减少；
- 每单位销售额对应的应收账款减少；
- 改善销售毛利率；
- 每单位销售额对应的管理费用减少；
- 改善的税率；
- 盈利质量，用先进先出法取代先进后出法；
- 审计意见，变更为有保留的审计意见；
- 每位员工的销售额。

6. 非美国市场的因子分析

最近的研究证实，因子分析提供了非美国市场的超额回报预期。

(1) 负面因子

有一些文献特别适用于构造做空证券组合。Beneish 和 Nichols (2009) 指出的基本思想是寻找改善会计盈利或现金流而不实际改善核心业务的行为，以及盈利操纵、并购历史、股权发行和其他形式的管理工作。

- 净销售额高于自由现金流；
- 高 PB 率；
- 高销售增长；
- 每股经营性现金流与股价比率低；
- 最近五年进行过收购；
- 股权发行 > 两年行业平均水平。

高财务杠杆，定义为净融资债务超过普通股股权。

Nissim 和 Penman (2003) 的研究显示，公司高财务杠杆与公司盈利能力负相关。高额且长期的金融债务会引发资金流动性问题甚至导致公司破产。长期资本管理投资公司的历史就是一个著名的金融杠杆误用的案例。此外，安然公司的会计丑闻也是源于其用特殊方式将长期债务从资产负债表中扣除，账目被纠错后直接导致其破产的。一般来说，额外的杠杆可能增强盈利能力，也可能提高破产风险。因此，读者迫切需要将财务杠杆与其他因子结合在一起，对企业进行统计分析，才能确定与未来收益正相关或负相关的证券组合。

(2) 特别注意事项

大宗商品行业的销售受相关商品价格影响，因此销售增长不一定是公司业绩提高的一种反映。银行有不同的报告要求，应给予特殊对待。经济周期所处阶段可能会影响债务与价格增长的相关性，并且对与增长率密切相关的因子有重大影响。

(3) 将因子转化为策略

在关于投资的文献中，关于因子应用方法，大家讨论最多的是建立样本集合，特别是在通过测试时为公司+1 分的机制，然后结合所有因子的分数，买入评分高的公司。根据 Kahneman (2011)，在没有其他数据分析结果的情况下，这通常是一种合理的融合方法。

(4) 将因子转换为 Alpha

请注意，大多数分数是以变动率为基础的，通常会减去前一年度的最新报表数据，因为季节性将直接影响季度之间的比较。报表数据对于 Alpha 分析来说可能会有迟滞性，因此应注意数据使用中的时间延迟问题。即时的财务数据提供了更现实的（更糟糕的回测）结果，因为这些数据中没有包括声明财报内容的正向偏差。

用一些回归模型执行对因子的统计分析可能是有益的，测量因子相关性也可能是有用的。现在还可以采用诸如遗传算法的机器学习技术来找到好的因子组合。如果对所采用的因子的含义有深入透彻的理解，那么证券组合很有可能获得成功。市场将 Alpha 奖励给有新意义的因子组。因此，研究人员应该创造性地思考，并为新的交易灵感参考大量文献。

第 23 章

机构研究 101

作者 Benjamin Ee

本章重点介绍了 Alpha 研究人员为了创造新的交易灵感,是如何将机构研究资源转化为自己的开发工具的。

本章第一部分,针对过去几十年来蓬勃发展的各种金融市场学术研究提供了一个“参观指南”。第二部分是研究人员经常在金融媒体中遇到的分析师研报和股票推荐的一般概述。我们将讨论如何获取分析师的建议,以及如何使用这些建议去激发市场策略灵感。

1. 金融市场的学术研究是否像大海捞针

如何才能获取新的市场策略灵感?灵感可以有许多来源,包括与同学、同事或朋友的讨论,以及金融新闻、研讨会、书籍等。本章将介绍可以转化为研究人员开发工具的另一个灵感来源:可免费获得的金融市场相关学术论文。

乍一看，过去几十年，由教授、研究生、各种市场评论家等公开发布的研究特别多，而对于行业新手来说，在现有文献中找到适合自己的文章是一项艰巨的任务。2014年12月，在 Google Scholar 上搜索“corporate finance”（企业金融）和“asset pricing”（资产定价），分别产生 150 万和 130 万个搜索结果；如果把搜索词设置得更具体，如“cross-section of stock returns”（跨板块股票收益），就会产生近 10,000 个结果，其中包括 2014 年诺贝尔奖获得者的著作。

因而消化所有现有的文章或许不是获取灵感的最快办法。

本章尝试为新手提供一个“简单”而“实用”的指导，以便他们可以快速地把学术研究变成个人开发工具的启发之源。论文可以通过几种手段访问，其中一些只需花费互联网上网的费用。费用不同主要是因为来源不同：“经过审查的”（如期刊）来源和“开放的”来源。

1) 期刊

正式的学术期刊是教授和专业研究人员发表研究成果的地方。正式期刊上的论文均经过了同行的严格评审，评审中其他研究人员会对其进行点评。因此，在这个过程中“幸存”下来的研究论文通常没有严重的研究方法错误。你不仅可以从中获得交易想法，还可以获得新的研究方法的启发。

在这些出版物中讨论的内容通常是专家学者们认为在发表的时点上非常重要的主题。这可能与市场策略开发者优先考虑的问题仅有部分重叠。学术界的主要焦点不是寻求市场策略，而是要了解主导经济相互作用的根本力量。虽然两者之间息息相关，但并不是每一篇文章都会同时涉及前者。为便于理解两者差异，我们做一个物理学中的“类比”：比如一方面是研究电磁学的基本定律，另一方面是开发一个核磁共振机器。

具有更“普适”结论的论文通常发表在高影响因子的期刊上，而且这些论文也会受到学术界的更多关注。对于希望利用这些论文结论的市场策略开发者来说，这个特性有其优缺点。一方面，这表示你将会有更多的工作要做，从对一般经济理论（如“道德风险”）的讨论到具体的战略。期刊上的讨论可

能与市场脱节，你需要将它与市场联系起来。另一方面，在这些讨论中，可能會有重要的经济基础理论，能够改进所构建的许多策略。同时，自己通过基础理论的启发设计新策略，而不是直接复制现有的结论，会使新策略有别于其他市场参与者。

与“综合”杂志相对的，还有聚焦在投资领域的杂志，其中的内容与投资策略和策略研究方法联系更紧密。在某些情况下，你甚至可能直接对所讨论的策略进行编辑和复制。这种情况的缺点我们也提到了，这些策略被描述得非常详细，其他人也可以执行这些策略。不过，一旦你的核心策略代码通过测试，就可以考虑将其用作进一步创新的基础，这比从大致想法开始简单多了。

如果学者有可行的市场策略，为什么他们不对其采取保密措施并自己交易呢？可能包括以下几方面的原因：

(1) 学者的目的是能解释经济内部运作的原理，他们的理论与市场反应一致只是证明的一部分。

(2) **执行事宜**：开发便于有效执行优秀市场策略的基础设施，并不是一件容易的事情，时间和金钱成本都很高。

(3) **职业和声誉考虑**：出版的预期收益，例如对其专业领域的学术影响、媒体报道、潜在的咨询合同等，都可能导致创作者倾向于分享信息。

(4) **时滞**：一篇新的学术论文可能需要几年的时间才能从一个概念发展为期刊发表的合格文章。如果该文章描述了一个高度可操作的策略，那么它可能已经被其他市场参与者所采纳和实施了。

(5) **成本**：学术期刊价格昂贵；你可以通过你所在职的大学或本地图书馆阅读这些论文，或者在某些情况下，直接向作者发送电子邮件索要副本。大多数学术作者很乐意发送他们的论文副本。

2) 开源资料

开源资料通常是可以自由访问的文档存储库，其中少部分允许用户上传文档。“自主上传”意味着论文可能没有经过传统的同行评议过程。在

许多情况下，它们可能是正式学术论文的早期版本，以最终通过审查为目标。开源资料库有 SSRN 或 arXiv 的论文库。一旦你确认了自己喜欢的学术作者名单，还可以在他们的个人页面上查看正在撰写的论文。

开源系统的特点如下。

(1) 滞后性小：根据定义，最新的研究通常是一份正在撰写的论文，而不是已发表的文章。容量有限的市场策略最好尽快实施。

(2) 一般是免费的：正在撰写的论文通常可以由作者免费分发，大多数作者都希望最大限度地传播自己的论文。

(3) 品种更多样：经过正式审核的期刊反映了学术界的倾向性。在所有领域中，一些细分领域会在某一时段得到更多的关注和讨论。所以，不能保证正式期刊彼时关注的焦点会和市场策略开发者匹配。然而，在 SSRN 或 arXiv 上可以找到更宽泛的领域，那里的作者具有较大的自由度；缺点是读者需要在挑选文章上更有辨别能力。

(4) (暂时) 没有同行评议：没有同行评议一般意味着没有质量保障，论文可能会有严重的方法错误，不过可以用来锻炼自己对主题做出判断的能力。有几种简单的标准可供参考：

① 作者是否发表过同行评议过的论文？

② 本论文是否将进行同行评议（他们可能会在他们的网站上或以脚注的形式声明“已提交给‘XYZ 杂志’”）。

③ 论文结论是基于常用数据库（如 Compustat、CRSP）还是基于罕见或自主收集的数据。如果作者的发现基于罕见数据，有没有办法使用常用数据的理论复现他们理论的本质？为了创造一个可行的市场策略，你通常本就应该这么做。

④ 其方法论是否常用且直观？

最后一类值得研究的论文来自自由市场数据供应商提供或发布的论文。这些论文中展示的策略都是基于这些数据供应商（例如 CRSP、WRDS 等）提供的数据，因此至少一部分是用于营销的目的。尽管如此，这些论文可以帮

助你快速上手一个新的数据集，并且策略可能具备可移植性。

3) 跟进讨论

公开分享研究成果的一个好处是它会引发讨论。其他研究人员将撰写后续的论文，对已发表的作品进行评论和改进。如果你了解如何改进现有框架，或想知道大多数人对其优点和缺点的看法，那么这个讨论可能会非常有用。

大学院校中的大部分学术研究人员通常都会了解学术界对最新文章的看法。同样的情况可能不适用于坐在办公室或在家里阅读最新一期的 *Journey of Finance* 的你。那么如何才能了解其他专业研究人员对你刚阅读的文章的看法呢？

方法之一是关注一篇论文的引文。引文是会涉及你刚阅读的论文的其他学术论文，通常会进行讨论改进、调整或批评。当代信息技术使得我们找到这些引文相当容易。以任意一篇文章为例，比如“*The Cross-Section of Expected Stock Returns*”（对于新手策略开发者来说，这算个不错的起点）。将文章名输入 Google Scholar（这与普通的 Google 不同，请使用 <http://scholar.google.com>），具体可以参阅图 23.1。

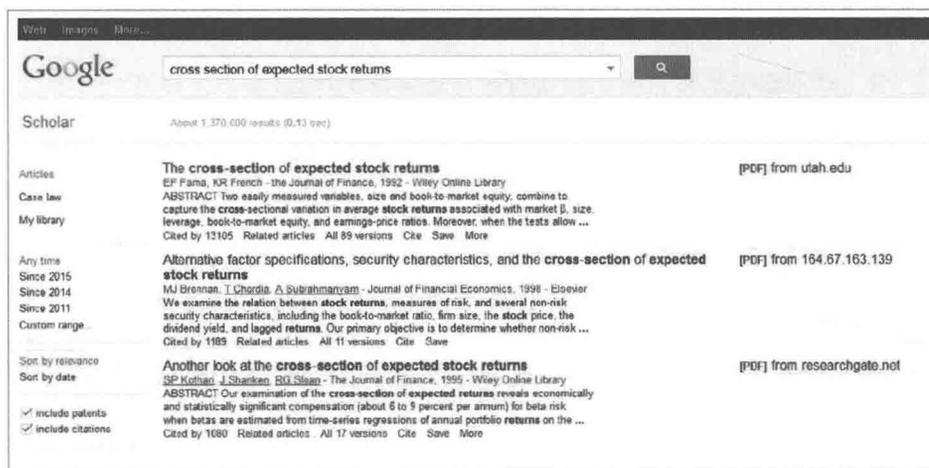


图 23.1 在 Google Scholar 上搜索 “The Cross-Section of Expected Stock Returns” 一文的搜索结果

资料来源：Google 和 Google 徽标是 Google Inc. 的注册商标，经许可使用。

我们可以立即看到，截至 2014 年 12 月，该论文已被 10,000 多个其他论文引用。单击“XXX 引用”的链接，Google Scholar 帮助返回参考上述论文的论文，如图 23.2 所示。

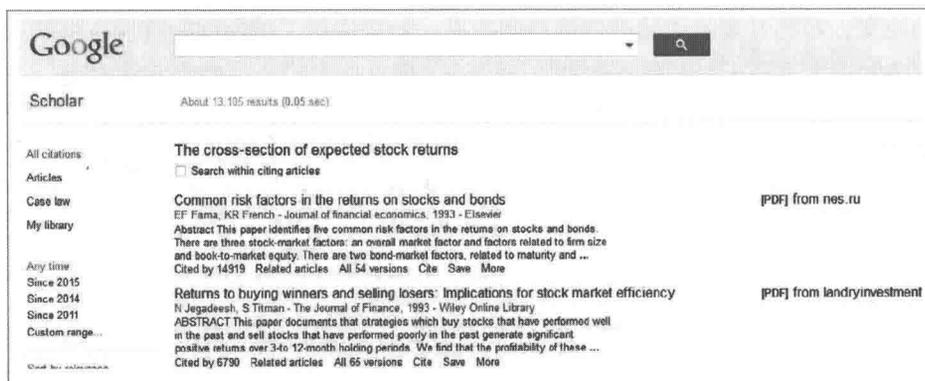


图 23.2 在 Google Scholar 上“*The Cross-Section of Expected Stock Returns*”被引用的历史细节
资料来源：Google 和 Google 徽标是 Google Inc.的注册商标，经许可使用。

若要找出所有关于某个主题内容，以及关于这个话题所有已讨论过的改进和评论，跟踪引用是一个很好的方法。正如人们常说的：“没有必要重新发明轮子！”

2. 学术研究的作者真的相信有效市场策略的存在吗

恭喜！你已经成功地阅读了本章关于学术研究讨论的上述内容，这是个不小的成就。

最难的问题留到最后，你可能会问：

- 学术研究人员是否真的认为自己可以写出可行的市场策略？
- 如果他们认为那是不可能的，那么我们为什么还要读这些人写的东西？

两个都是好问题。如果你对市场真正感兴趣，你可能与你的大学教授在这一点或者某些方面针对怎么战胜市场进行过辩论。“有效的市场假说”（EMH）和“不可能”这几个词甚至会进入脑海。辩论可能是令人沮丧的，至少不是因为有些天的市场回报就像州际公路 10 英里限速一样。

笔者认为学术研究人员非常重视 EMH，不是因为这是一个可以和重力与运动定律一样坚不可摧的定律。相反，这是因为市场异常确实存在，但它需要一而再、再而三的仔细核实。一方面，由于各种历史、制度、法律、政治或经济原因造成的市场异常^①提供了获利的机会；另一方面，历史上有很多投资者，将市场异常与风险因素混淆，并因此付出了惨痛的代价。

总结

(1) 为了寻找新的想法，你可能需要为你的策略开发工具包添加新来源——学术研究。

(2) 过去几十年来，在金融研究领域产生了无数论文。

(3) 筛选这些论文的策略至关重要。成熟的知名期刊可以作为一个不错的起点。通过引文将论文分组为不同的“会话主题”是另一个补充策略，你只需关注你感兴趣的主题。

(4) EMH 不是一个不可破灭的定律，但是它值得被严肃对待。

3. 分析师研究

(1) “高盛提高 Netflix 评级”——TheStreet.com, 2014 年 7 月；

(2) “摩根士丹利降低了对科技行业评级”——雅虎财经, 2014 年 12 月；

(3) “花旗集团保持对美铝公司的评级”——MarketWatch.com, 2014 年 12 月。

在金融报纸、会议、博客和数据库中，卖方分析师对企业和整个行业的研究都占有重要地位。在探究股价的大起大落时，分析师的荐股、调高评级、降低评级或修改目标价位等被认为是主要原因的情况屡见不鲜。行业协会和学者界的大量研究发现，分析师研究中确实包含了大量重要信息[见 Francis 等

① 严格来说，市场异常存在（即使是极端的例子，如真正的套利机会）也不能反驳 EMH。市场回报的定义是什么？如果我们将所有可能的市场异常利润纳入“市场回报”的定义，那么 EMH 仍然是无法挑战的。无论如何，如果市场异常让我们在控制风险的同时赚更多的钱，那么这一点就是学术性的。

(2002) 和 Frankel 等 (2006)]。

尽管如此，“股票分析师”给大家的形象，都是在高盛 (Goldman Sachs) 或摩根大通 (JP Morgan) 里的资深研究员。他们与 500 强企业 CEO 强势对话，了解其收益情况并把搜集来的信息展示给数十亿美元的机构基金。对应这个形象，你作为一个新的 Alpha 研究员 (管理的可投资金稍少于几十亿美元)，如何才能接触到这个价值连城的分析？同样重要的是，一个对构建系统性市场策略有兴趣的研究员，为何要对特定公司级别的分析同样感兴趣呢？

1) 获取研报 (当然了，免费)

关于股票研究的一个有趣而意外的事实是，其中一些研究是非常容易获取的，金融媒体就是一个有价值的中介渠道。在这种情况下，金融媒体不仅包括《华尔街日报》或“Bloomberg”等传统信息来源，还包括雅虎财经和 Google 财经等综合网站。后者在查询分析师分析、预测或公司盈利电话会议的问题时尤其有用。与学术研究一样，大多数这类网站至少有一些免费/开放的内容，你可以根据自己的研究需求进行查询。

2) 用户将无法通过公共资源获取所有的银行分析师公司研报 (甚至大部分都不行)

卖方分析师进行的是成本昂贵、过程复杂并且耗时的分析，他们自然希望把分析先推荐给最有价值的客户。尽管如此，进入公众视野的一部分分析师研报也可以为新的 Alpha 研究人员提供一个有价值的学习工具。

3) 在金融门户网站上查看分析师的讨论和预测

鉴于金融媒体上关于分析师研报的报道激增，找出提及这些分析师研报的新闻通常和把公司股票代码输入你最喜欢的金融门户一样容易。例如，进入雅虎财经门户，在 2014 年 12 月^①输入苹果公司股票代码“AAPL”就会出现如图 23.3 所示的头条。

^① 这里以一个门户网站 (Yahoo Finance) 为例。你可以在其他主流融资门户网站 (如 Google 财经, Bloomberg.com 等) 上找到相似的内容。

| Headlines | HPQ | 72.52B |
|---|---|--------|
| <ul style="list-style-type: none"> Pa. school district uses \$291K Apple gift on iPads AP (Wed 8:52AM EST) Why Even Apple Is Surprised by the Enormous Demand for the iPhone 6 at TheStreet (Wed 8:45AM EST) I'm not loving these big phones Yahoo Finance Blogs (Wed 8:42AM EST) Will Apple (AAPL) Stock React Today to This Analyst Action? at TheStreet (Wed 8:35AM EST) Meet 5 Big Caps Whose Earnings Are Expected To Surge at Investor's Business Daily (Wed 8:02AM EST) TiVo is a good acquisition target for Apple, Alibaba among others: analyst at MarketWatch (Wed 7:49AM EST) iPhone 6S Rumors Point To A Curious Early 2015 Release at Forbes (Wed 5:06AM EST) Don't Confuse Morgan Stanley Decreasing Apple's Weighting By 1% With Being Negative On The Shares at Forbes (Wed 4:97AM EST) Apple iPod Antitrust Trial Gets Under Way at The Wall Street Journal (Wed 1:05AM EST) Jobs Said Apple Took Pains to Protect Record Contracts at Bloomberg (Wed 12:23AM EST) Mozilla wants to put Firefox on iOS at Engadget (Tue, Dec 2) Apple vs. Samsung in '15 at Barrons.com (Tue, Dec 2) Google's Chromecast Overtakes Apple TV in Video Streaming, Survey Says at The Wall Street Journal (Tue, Dec 2) Your first trade for Wednesday at CNBC (Tue, Dec 2) In emails, Jobs determined to keep iPod Apple-only AP (Tue, Dec 2) | <p>» More Competitors</p> <p>Key Statistics</p> <p>Forward P/E (1 yr): 13.42</p> <p>P/S (ttm): 3.68</p> <p>Ex-Dividend Date: 06-Nov-14</p> <p>» More Key Statistics</p> <p>Analysts</p> <p>Annual EPS Est (Sep-15): 7.74</p> <p>Quarterly EPS Est (Dec-14): 2.53</p> <p>Mean Recommendation*: 1.9</p> <p>PEG Ratio (5 yr expected): 1.29</p> <p>* (Strong Buy) 1.0 - 5.0 (Sell)</p> <p>Analyst Opinion Estimates</p> <p>Business Summary</p> <p>Apple Inc. designs, manufactures, and markets mobile communication and media devices, personal computers, and portable digital music players worldwide. View More</p> <p>Company Profile Industry</p> | |

图 23.3 在 Yahoo Finance 上搜索“AAPL”的结果截图

资料来源：Yahoo Finance©2014 版权所有。经许可转载。

在截图的左侧，很快，我们不难找出分析师研报的标题和对应链接。在页面右侧，雅虎财经还为分析师对 AAPL 每股收益的预测和平均分析师操作推荐（例如强力买入、卖出等）进行了非常有用的总结。单击上面高亮的链接通常会连接到一个分析师的综述页面，包括对 AAPL 的看法、他的思考过程、基于的数据、注意事项以及价格目标和操作推荐。

同时，你可能希望在其他门户网站（如 Google 财经、Bloomberg.com 等）上尝试此过程，然后再选择最适合自己的网站。当所有其他的尝试都失败时，试试直接用搜索引擎（例如在 Google 通过输入“appl analyst reports”）来查看。

与分析师研究一样，通过财务门户网站可以获得公司盈利电话会议的内容，同时也能获取股票分析师的提问和通话期间公司对这些问题的回应。在下一个例子中，我们将使用名为 Morning Star 的另一个财务门户网站，网址为 <http://www.morningstar.com>。这是互联网上可以找到这类信息的众多途径之一。其他的途径还有通过 Seeking Alpha 网站，<http://seekingalpha.com>，证券交易所网站（如 <http://www.nasdaq.com>）以及公司网站的投资者关系板块。其中一些还包含非银行市场评论员对特定行业和股票的大量讨论。这些网站的讨

论可能类似于股票分析师的研究，涉及诸如企业特定基本面、宏观经济学、区域政治和市场条件等方面的观点。

看看你是否可以在 MorningStar 网站上找到 AAPL 的 2014 年第二季度附有分析师问答的盈利电话会议手稿。可点击的重要链接在黑方框中高亮显示，如图 23.4 所示。



图 23.4 苹果公司在 MorningStar.com 上的高亮截图

资料来源：MorningStar.Inc 保留所有权利，经许可转载。

4) 这些是不错，但你为什么还要关心这些呢

好问题。大多数分析师报告（或市场评论）专注于单一股票或行业，而你作为 Alpha 研究人员，正在寻找每天交易成千上万只股票的系统性市场策略。那么，当 XYZ 银行的一些分析师特别偏爱某家公司时，你该怎么办？我们如何从这个结论出发，在全球 20 个不同的证券交易所交易成千上万只股票？

以下解释了为什么我们认为寻求系统性市场策略的研究人员可以从阅读股票分析师报告中学习的原因：

- 比任何具体的“买入”或“卖出”推荐更重要的是分析师的思考过程：他决定提高 AAPL 的评级是不是因为行业原因？（例如，智能手机市场正以三位数增长）还是因为企业特定的原因（净利润在过

去四个季度有所增加)或更普遍的原因,例如公司的市盈率低于行业内的其他企业?无论是何原因,一个有趣的问题是:我可以把它应用到其他公司吗?例如,如果分析师说他喜欢 AAPL,是因为这家公司的 CEO 也购买了它的股票,那么这个逻辑只能适用于 AAPL 还是对一般上市公司都可行?这一推理方式已经产生了几种新的策略思想。

- 分析师通常会在盈利电话会议期间提出非常好的问题。鉴于他们拿到的高额回报,他们也应该这么做。对于一位试图理解现代公司密集财务数据的新手研究员来说,这些问题可以帮上大忙。信息越多并不代表越好,特别是当每个公司都有 20 多页财报的时候,试图把信号和噪声分开是个技术活。你应该关注哪个会计科目?一条线索是关注分析师聚焦的数据和趋势,以及他们关注这些问题的目的。他们是不是为库存每一季度的非季节性巨大变化感到困惑?为什么这很重要?和往常一样,我们关注的是,其中是否具有普适性的逻辑。
- 分析师具有丰富的行业知识。Larrabee (2014) 指出,行业知识是股票分析师最重要的特征(和竞争优势)之一。最好的卖方分析师甚至能够通过股票评级和预测来改变股票的价格,并且这种影响在具有行业经验的分析师中更明显。这些特征都与研究人员相关,因此我们可以从股票分析师的工作中学到很多方法论。例如:
 - 评估方法因行业而异。对于构建现金流贴现模型。制造企业与周期性公司或非周期性公司的区别很大。分析师可能会关注不同的估值指标,例如一个行业用市盈率,另一个行业用市净率。对于 Alpha 研究人员来说,应了解这些差异的根本原因,以便将其推广应用到交易领域中。
 - 每个行业可能都有自己独特的驱动力或衡量运营水平的指标,通常在分析报告中占据突出地位。在 20 世纪 90 年代后期,互联网企业曾经通过“眼球”指数来衡量运营水平可能现在在一定程度上还在用;航空公司则考虑“乘客里程数”;生物技术公司可能会专注于生产线上的新药或药物试验。了解每个行业运营绩效的关

键驱动因素可能有助于 Alpha 研究人员尝试了解其策略表现中的行业间差异。

- 分析师研报可以提供有效的交易信号。分析师研报偶尔还会影响股价。你可能已经看过，新闻把个股股价的暴涨和暴跌归因于分析师对这只股票评级的升降或目标价位的调整。有广泛的学术研究表明分析师研报和股价之间的联系，你可以在 Google Scholar (2014) 的“stock analyst research” 的搜索中找到。更好地了解分析师的建议可能有助于你在构建策略时更好地利用这些信息。

5) 阅读分析师研报中的注意事项

无论你是通过阅读分析师研报来寻找新的市场策略灵感，还是想直接在策略设计中使用其建议和目标，把分析师研究的一些优缺点和特质都牢记在心是非常有用的。

- **正向偏见**：虽然不同的银行可能有不同的方法，但学术研究人员认为，股票分析师作为一个群体可能会表现出正向偏见。其中一种情况是，分析师买卖的建议分配可能会出现倾斜。例如，如果我们统计“买入”和“卖出”的建议，可能就会发现“买入”的建议比“卖出”的建议更多。学术研究人员进一步研究了这种情况发生的原因，比如 Michaely 和 Womack (1999)、Lin 和 McNichols (1998) 进行的研究。
- **羊群效应**：羊群效应指分析师不希望其给出的公开建议和目标与市场上其他分析师给出的有太大不同。正如理论所说，有一部分原因可能和行为学相关。公开股票价格预测（或分析师说的“目标价位”）对分析师的职业发展有一定风险。在其他条件相同的情况下，跟随大多数人做出相同决定可能会更安全。对此的一个推论是，更自信或声誉更好的分析师通常愿意做出与主流意见不同的决策建议。

除行为学的原因外，还有其他合理的理由可以解释为什么分析师容易得出类似的结论。比如，大多数人可能都有相同的信息来源。因此，了解著名分析师为什么会得出与其他人不同的结论，他们的分析方法或数据来源有什

么特殊性，如果可以，将其进行“系统化”，可能是一件很有趣的事。

6) 股票分析师为什么要和金融媒体接触

如果你准备为了创造出精彩的交易思路而投入大量时间和精力，那么你的第一冲动可能不是拿起电话，把这件事告诉一大堆记者。毕竟，许多想法资金容量是有限的：只有这么多人可以在价格开始大幅度变化前交易，随后获利的机会就会消失。

然而，我们一直可以在公开媒体上看到分析师研报的提及。事实上，我们甚至在一定程度上依赖这些报道，因为它让我们几乎不用付出任何成本就能窥探到分析师研究的世界。如何解释这个现象呢？原因可能包括：

(1) 股票分析师与他们所负责公司之间的一些会议是对公众开放的，并且经常被金融媒体报道。例如盈利电话会议，会议的记录可以公开获得。一个可能的原因（至少在美国）是 SEC 有公平披露的规定（称为 FD）（www.sec.gov, 2015），其规定发行人向投资专业人士披露的非公开信息也必须提供给所有的市场参与者，包括数十亿美元规模的大银行分析师以及参与投资的其他人员。

(2) 适度的曝光率对股票分析师职业生涯有益无害。媒体关注和采访可能会增加投资者对分析师研报的需求。经过市场验证的高质量建议会使分析师的职业发展进入快车道（例如在 2008 年市场触底时的买入建议）。当然，只有在分析师们对大众开放他们的一些研究时，上述的情况才有可能发生。

总结

- (1) 分析师的研究成果可能可以通过金融媒体获取；
- (2) 分析师在达成他们最终建议的过程中使用的方法和推理可能是 Alpha 研究人员的灵感来源之一；
- (3) 分析师的建议和目标价位可能会产生交易信号；
- (4) 正向偏见和羊群效应是需要注意的两个要点。

第 24 章

期货交易简介

作者：Rohit Agarwal

期货交易包括股指期货、商品期货、外汇期货和债券。

Alpha 产生的数量与所包含金融工具的数量及信息的深度成正比。相对于股票市场样本空间，期货市场样本空间的金融工具肯定要少很多。这意味着我们需要进行更深入的研究，从而提升信息的深度，产生大量的 Alpha。

与股票不同，期货合约的多样性使得单一特征可能无法与市场中其他所有合约相比较，但这种困难可以通过市场更大流动性来得以补偿。

以下是在期货市场研究中发现的最有趣、最突出和最有用的信息或现象的概览。

1. 商品期货交易委员会发布的交易员持仓报告

美国每周五公布交易员持仓报告（见图 24.1）。知道“聪明的钱”在购买哪些产品，然后跟随它去操作是非常有价值的。该报告为我们提供了不同市

场参与者的持仓明细，例如商业交易者（大企业/生产者）、非商业交易者（大型投机者）和无须报告交易者（小型投机者）。

```

Disaggregated Commitments of Traders-All Futures Combined Positions as of January 14, 2014
:
: Reportable Positions
:-----:
: Producer/Merchant : Swap Dealers : Managed Money : Other Reportables :
: Processor/User : Long : Short : Spreading : Long : Short : Spreading : Long : Short : Spreading :
:-----:-----:-----:-----:-----:-----:-----:-----:-----:-----:
WHEAT-SRW - CHICAGO BOARD OF TRADE (CONTRACTS OF 5,000 BUSHELS)
CFTC Code #001602 Open Interest is 422,664
: Positions
: 69,239 79,610 94,197 15,094 13,557 95,741 152,136 44,882 38,695 32,458 32,917
:
: Changes from: January 7, 2014
: -8,010 1,077 626 1,083 1,197 6,007 -9,357 1,712 -3,748 1,405 11,273
:
: Percent of Open Interest Represented by Each Category of Trader
: 16.4 18.8 22.3 3.6 3.2 22.7 36.0 10.6 9.2 7.7 7.8
:
: Number of Traders in Each Category Total Traders: 384
: 72 72 15 8 16 40 89 56 62 64 62
:-----:-----:-----:-----:-----:-----:-----:-----:-----:-----:

```

图 24.1 2014 年 1 月份最后一周的交易员持仓报告

资料来源：美国商品期货交易委员会 2014 年 1 月份数据。

2. 市场季节性

季节性是一年中某些时候市场向某一既定方向移动的趋势。这种现象因为天气周期的变化，在农产品期货市场上表现得很明显。但是，它不仅限于在农产品期货市场，也可能因周期性需求、消费、库存或供应曲线而出现。

图 24.2 显示了天然气储量或库存的季节性变化。随着冬季的到来，天然气的需求由于在家庭供暖中的使用而增加，从而使库存量减少。

但过去的表现并不能保证未来也会一样。

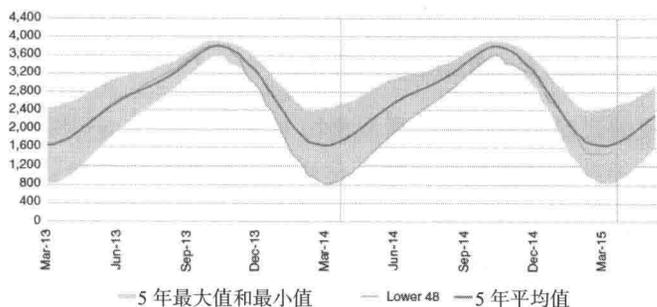


图 24.2 和 5 年最大值和最小值相比的地下储存的工作气体值

资料来源：美国能源信息管理局 2014 年 7 月数据。

3. Risk On与Risk Off

有时市场情绪是积极的，投资者态度乐观而且愿意承担更多的风险以获得更好的回报。这种环境被称为“Risk On 期”，因为投资者在这种环境中总是在承担风险。另外，有时候投资者是悲观的，并尝试通过出售其在风险资产中的头寸，从而将资金转移到现金头寸或低风险的避风港（如美国国债）来降低风险，这种环境被称为“Risk Off 期”（见图 24.3）。

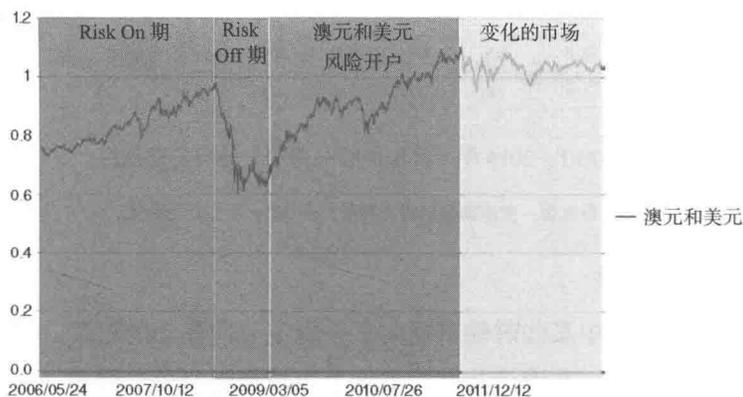


图 24.3 Risk On 期/Risk Off 期澳元和美元的汇率曲线

这种投资者行为——在 Risk On 期购入有风险的资产，在 Risk Off 期转换为无风险的资产，增加了不同资产类别之间的相关性。

成功构建 Risk On/Risk Off 策略的步骤包括判断每个时间区间（每天/周/月/季度）内，市场是 Risk On 还是 Risk Off，然后将不同的资产分为 Risk On 资产或者 Risk Off 资产，然后给它们头寸。

VIX 是反映市场风险的一个重要指标（见图 24.4），它是由标准普尔 500 期权的隐含波动性构成的波动指数。传统上，股票价格和波动率之间的相关性为负。因此，VIX 高或在升高中意味着货币从股票市场转投到更安全的资产上，宣布进入了 Risk Off 时期。VIX 本身也是一种可交易的期货工具，被许多人用来从下跌市场中获益。

4. 升水/贴水

当近期期货比远期期货便宜时，会产生一个向上倾斜的价格曲线，这种

合约被称为升水（见图 24.5）。出现这种现象可能是因为商品的存储成本或者持有成本。

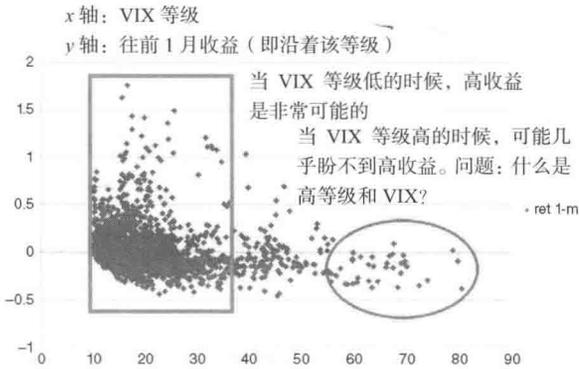


图 24.4 市场收益在不同 VIX 等级下的分布

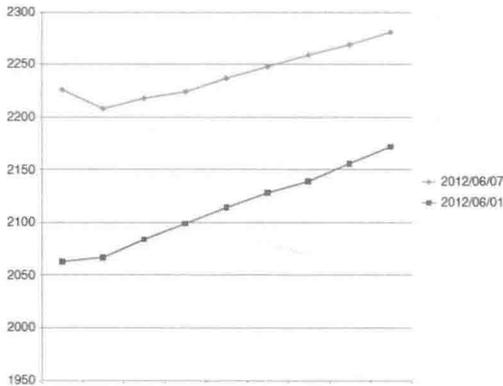


图 24.5 在升水情况下可能的曲线分布

有时近期期货比远期期货更贵，会产生一个向下倾斜的价格曲线，这种合约被称为贴水。

一些贸易商通过在升水的时候卖出合约、在贴水的时候购入合约赚钱。

虽然这些都只是在期货交易中广泛研究的主题，但适用范围可以扩展到更多领域。

第 25 章

远期外汇和期货中的 Alpha

作者：Richard Williams

在外汇和期货市场中寻找 Alpha 是一件能带来巨大切实利益的事情。但正如本书前面提到的，也会有很多挑战。在本章中，我们将讨论在过去几年里不断试错总结得出的解决这些挑战的行之有效的处理技巧和思维方式。

1. 关键的市场特征

期货是用来让交易者在标的市场中获得价格变化对应的敞口，而不用斥资购买标的资产对应的仓位的一种金融产品。

在 Alpha 研究的范围内，短期的远期外汇也有相同的功能去锁定相对货币价值。这就意味着，对于套期保值者和投机者来说，期货和远期外汇都是很便捷的金融工具，因此我们需要去研究它们最显著的特性。

2. 标的资产的因子敞口

既然期货和远期外汇对于标的资产有相同的敞口，那么影响期货和远期

外汇价格的因子就能影响标的资产。这个简单的观点有着重要的启示：有一些市场交易者明显专注于特定的期货和远期外汇，他们无非是想通过控制特定因子的风险来进行套期保值。商品期货市场里就有很好的例子，实物商品的生产者和消费者会对特定商品的风险进行套期保值。

例如，广泛使用农业期货来控制风险的农民和食品生产企业，与广泛利用能源期货来对冲未来燃料成本的航空公司之间就鲜有共同之处。每一组交易者都有自己特有的风险限额、风险容忍度和交易行为，也会相应地对不同的市场行为有相应的定性。

在区别这些市场用户之间的差异时，我们发现在许多组织中传统的模式是将相同标的资产分为一组，然后由每组对应的专家来管理不同类别的期货和货币交易。这进一步强化了“行业”之间的差异，因为不同的交易者、交易场所和业务种类都有着不同的期货和远期外汇。

3. 金融工具分组的结果

前面讨论的金融工具分组，让我们在探索 Alpha 的过程中同时面临着机遇和挑战。其中一个最明显的挑战是，当我们考虑将联系更紧密的金融工具划分在更小的组里时，实际交易中可用来选择的标的就更少了。

因此，我们要求每种金融工具有更好的结构，以便和更大的金融工具集整合形成相同的结果。每种金融工具 Alpha 质量的提高，会更容易被其他的市场参与者知晓，从而导致 Alpha 的适用时间变短。

相反，因为我们区分了相似金融工具并使之成为一个集合，所以我们对所有金融工具的集合中，哪些集合会有特定的 Alpha 有了更好的预判。在集合中，依次删除每种金融工具，并重新确定其关系的测试也变得更有意义。这些金融工具的集合通常会成为我们测试候选 Alpha 的基本单位。

4. Alpha测试的基本清单

我们从 Alpha 的核心想法开始，第一步，先确定我们认为的 Alpha 会出现的板块，以及出现的时间段。以美国能源市场为例，讨论一个和富含石油与天然气的美国海岸的极端天气相关的 Alpha。

我们在考虑数据之前，先确认一下相关的金融工具：石油、天然气及其衍生物；还要确认一下数据具体会在哪个时间段产生影响，这样就需要确认暴风雨（极端天气）会持续多久、生产会因为天气而中止多久，以及重新开工需要多久（在这两个方面有许多观点，但我们仍然可以使用极端天气和商品供应之间的隐含因果关系来缩小范围）。

现在可以通过得到过去天气的预测和主要能源合同的价格变化来完成预期的测试，并使用部分抽样内的历史数据集来测试这两个数据集之间的关联。

然后匹配一个简单的统计模型并测试其稳健性，在匹配过程中会对参数做出调整。一个好的稳健性测试需要覆盖相似的资产（但预期变化不会太大）。

在上述例子中，布伦特原油会是一个恰当的选择。原油是一个全球性的市场，所以我们预期倘若美国中断供应的话，会带来一些溢出效应。然而，在欧洲交付的石油并不是美国所供应原油的完美替代品，因此我们期望其影响力可以有所减弱。此外，我们也可以用抽样内测数据再做测试。

在测试过 Alpha 有望出现的地方后，现在可以反过来测一下，我们期望哪些地方和 Alpha 没有关系。

就本节所述这个例子而言，它是专门针对一个行业的，所以套用到不同的行业，如工业金属或债券期货，我们预测不会对其发生任何影响。这一步骤在发现错误编码或指定的统计测试时效果非常好。

根据上述结果，我们现在可以在自己的抽样外测数据库中测试 Alpha 了。这些小规模的抽样外测数据库可以避免无意的过度拟合，因此在整个流程中它们是非常关键的。

5. 小结

以标的资产为基准，将金融工具划分为不同板块，对于普通和特殊组的市场参与者来说是非常重要的，它为不同板块间及同一板块里的金融资产提供了呆板的弱关联。如果你愿意花点时间来考虑各个板块的因子敞口如何能与你正在考虑的想法相对应，那么你将找到一个针对你的想法展开测试的可行之处。

第四部分

新工具 WebSimTM

- 第 26 章 WebSimTM 简介
- 第 27 章 Alpha 与 WebSimTM 的原理
- 第 28 章 理解 WebSimTM 的工作流程
- 第 29 章 用户帮助手册
- 第 30 章 输出结果说明与 Alpha 信号库
- 第 31 章 Alpha 教程
- 第 32 章 常见问题解答

第 26 章

WebSimTM 简介

作者：Jeffrey Scott

过去，我们运用操作复杂的专业模拟平台开发 Alpha。这些模拟环境允许研究者运用包括基本面数据和价格/交易量数据在内的多种数据集合去验证 Alpha。

数据集包括多种公司指标，如季度利润、现金流、资产收益和负债，以及价格/交易量数据。如开盘价格与收盘价格、特定时间间隔内的最高价与最低价、成交量加权平均价、日交易量等。

虽然模拟平台最初十分有效，但还是有以下局限性：

- 由于只有公司网才可以进入系统，研究员局限于 WorldQuant 办公室的全职雇员；
- 绝大多数研究员需要使用 C++ 语言进行编程。

近几年，我们已经开发出庞大的 Alpha 数据库，可以在交易策略中使用。然而，我们的目标是为证券投资经理最大程度地扩展 Alpha 的数量和种类。上述局限性抑制了 Alpha 的增长速度。

为了克服以上局限性，我们开始思考更多的可能性：

- 假如将模拟平台放在云中，作为网络应用程序，将会产生什么样的效果？
- 假如我们引入兼职研究员、咨询员，将会产生什么样的效果？
- 假如我们破除编程语言方面的门槛，向更广泛的大众开放，将会产生什么样的效果？
- 假如我们允许懂得最多、最有动力和求知欲的人（比如大学学生）进入，将会产生什么样的效果？

正因为有了这样的考虑，WebSim™ 应运而生……

WebSim™ 是基于网络的专用模拟平台，是互联网模拟引擎的功能核心，人们可以用任意浏览器全天候对其进行使用。

打破物理地理局限性，意味着使用者可以分布在世界的任意角落，只要接入互联网即可。甚至在校大学生也可以在他们的业余时间利用 WebSim™ 开发 Alpha，没有任何时间和地点要求。

使用者可以根据他们的交易策略在 WebSim™ 上开发 Alpha，之后，利用真实的历史数据进行回测，从而了解他们的 Alpha 策略在过去市场上的效果。

为了尽量避免使用结构化的编程语言，WebSim™ 使用更为简单的命令流来输入 Alpha 表达式，具体参见图 26.1。

使用者可以按照此类命令流输入简单公式或者复杂公式，使用的数据元素包括基本面或量价数据。具有深厚编程功底的使用者，可以选择用 Python 脚本和数据库来创建 Alpha，如图 26.2 所示。

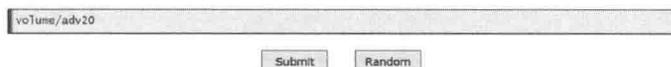


图 26.1 使用命令流输入 Alpha 表达式

```

Python Code

# Load data and assign alpha value

# Load the data 'close' into the variable closePrice. 'closePrice' is a numpy matrix.
closePrice=dr.GetData("close")

# the 'generate' function is used by websim to assign stock weights.
# 'di' is today's date and 'alpha' is the vector to be filled
def Generate(di,alpha):

    # For delay = 1, use yesterday's data values to assign today's alpha value.
    # To access data stored in a variable use dataName[date, instrument]. For alpha use alpha[instrument].
    alpha[:] = 1./closePrice[di-delay,:])
    # The above statement is equivalent to the expression '1/close'

# Using multiple data at the same time

# Different data needs different variables
closePrice=dr.GetData("close")
highPrice=dr.GetData("high")

def Generate(di,alpha):
    # Use vectorisation wherever possible. Avoid using loops as they are slow.
    # Here we use numpy's builtin subtraction function (numpy.subtract is called automatically)
    alpha[:] = closePrice[di-delay,:] - highPrice[di-delay,:])
    # The above statement is equivalent to the expression 'close - high'

```

图 26.2 利用 Python 创建 Alpha

可以使用基本面数据、价格/交易量数据创建 Alpha，也可以将二者结合起来创建 Alpha。WebSim™ 提供多种全局设定，使用者可以根据自己的选择设置参数，如用于检测的股票数目、用于回测的时间段等。

表 26.1 中给出了一系列可以在 WebSim™ 表达式中使用的数据字段。表中字段并不完备，仅为简要参考。

除了数据字段之外，使用者可以使用多种数学运算器创建表达式，具体可以参考表 26.2。同样，表中内容并不完备，仅为简要参考。

使用者可以即时获得并观看 WebSim™ 的模拟结果，如整体 P&L、夏普比率和日换手率等因子，从而可以判断 Alpha 的性能表现（见表 26.3）。通过调整 Alpha 的参数可以提高其性能，从而优化自己的策略。

在指示栏中，WebSim™ 使用者创建的 Alpha 在排行榜上会持续更新，使用者可持续关注自己的 Alpha 表现。使用样本外测的方法可以获得高质量 Alpha，这对进行更深入的分析 and 研究十分重要。

表 26.1 WebSim™ 表达式简要字段列表

| 数据字段 | 描述 | Python 用法 |
|------|-------|--------------------|
| open | 每日开盘价 | dr.GetData("open") |

续表

| 数据字段 | 描 述 | Python 用法 |
|--------------------|---------------|----------------------------------|
| close | 每日收盘价 | dr.GetData("close") |
| high | 每日最高价 | dr.GetData("high") |
| low | 每日最低价 | dr.GetData("low") |
| vwap | 每日成交量加权平均价 | dr.GetData("vwap") |
| volume | 每日成交量 | dr.GetData("volume") |
| returns | 每日收益 | dr.GetData("returns") |
| adv20 | 过去 20 天平均日交易量 | dr.GetData("adv20") |
| sharesout | 每日流通股 | dr.GetData("sharesout") |
| sales | 季度销售额 | dr.GetData("sales") |
| sales_growth | 季度销售增长值 | dr.GetData("sales_growth") |
| sales_ps | 每股季度销售额 | dr.GetData("sales_ps") |
| income | 季度净收入 | dr.GetData("income") |
| eps | 每股收入 | dr.GetData("eps") |
| cashflow | 季度现金流 | dr.GetData("cashflow") |
| cashflow_op | 运营所得季度现金流 | dr.GetData("cashflow_op") |
| cashflow_fin | 金融活动所得季度现金流 | dr.GetData("cashflow_fin") |
| cashflow_invst | 投资所得季度现金流 | dr.GetData("cashflow_invst") |
| cashflow_dividends | 分红所得季度现金流 | dr.GetData("cashflow_dividends") |
| assets_curr | 流动资产 | dr.GetData("assets_curr") |
| assets | 总资产 | dr.GetData("assets") |
| equity | 普通股股东权益 | dr.GetData("equity") |
| debt_lt | 长期债务 | dr.GetData("debt_lt") |
| debt_st | 短期债务 | dr.GetData("debt_st") |
| debt | 总债务 | dr.GetData("debt") |
| liabilities_curr | 流动负债 | dr.GetData("liabilities_curr") |
| liabilities | 债务 | dr.GetData("liabilities") |
| EBITDA | 税息折旧及摊销前利润 | dr.GetData("EBITDA") |
| capex | 资本支出 | dr.GetData("capex") |
| operating_income | 运营收入 | dr.GetData("operating_income") |
| operating_expense | 运营费用 | dr.GetData("operating_expense") |
| cogs | 销货成本 | dr.GetData("cogs") |
| bookvalue_ps | 每股价值 | dr.GetData("bookvalue_ps") |

表 26.4 创造 Alpha 表达式的部分数学运算器

| 运算符号 | 描述 | 返回类型 |
|---------------------------------|--|------|
| +, -, *, /, ^ | 加, 减, 乘, 除, 幂 | 向量 |
| <, ≤, >, ≥, ==, != | 逻辑对比运算 | 向量 |
| , &&, ! | 逻辑或, 与, 非 | 向量 |
| cond?expr1:expr2 | 如果条件为真, 则输出 1; 否则输出 2。例如, close<open?close:open | 向量 |
| Rank(x) | 在所有参数中给 x 评级, 返回平均分配在 0.0 ~ 1.0 之间的浮点数 例如, 指定 6 只股票的收盘价[20.0,15.6,10.0,5.7,50.2,18.4], rank(close) returns [0.8,0.4,0.2,0.0,1.0,0.6] | 向量 |
| Min(x,y) | 向量 x 与 y 的平行最小值 (类似于 R 语言中的 pmin 函数)。这需要 2 个向量作为输入值, 返回值为单个向量。结果中的第一个元素是第一个向量的最小值, 第二个元素是第二个向量的最小值, 依次类推 | 向量 |
| Max(x,y) | 向量 x 与 y 的平行最大值 (类似于 R 语言中的 pmax 函数)。这需要 2 个向量作为输入值, 返回值为单个向量。结果的第一个元素是第一个向量的最大值, 第二个元素是第二个向量的最大值, 依次类推 | 向量 |
| StdDev(x,n) | 过去 n 天 x 向量值的标准差。注意: n 必须小于 256 | 标量 |
| Correlation(x,y,n) | 过去 n 天 x 向量和 y 向量相关性的值。注意: n 必须小于 256 | 标量 |
| Tail(x,lower,upper,newval) | 如果 x 值在上界和下界之间, 则为 x 设为 newval | 向量 |
| Ts_Min(x,n) | 过去 n 天的最小 x 值。注意: 这与最小值不同 | 向量 |
| Ts_Max(x,n) | 过去 n 天的最大 x 值。注意: 这与最大值不同 | 向量 |
| Sum_i(expr,var,start,stop,step) | 遍历变量 (分步从开始到结束) 与每次迭代的计算表达式 (可能包含变量的表达式), 然后对所有值求和。例如: sum_i (Delay (close, i)*1, 1,2,4,1) 等同于 Delay(close,2)×2+Delay (close,3) ×3+Delay (close,4)×4 | 标量 |
| Call_i(expr,var,subexpr) | 将表达式里变量替换为子表达式, 然后计算表达式。如 call_i(x+4,x,2+3) 等同于 (2+3)+4 | 向量 |
| Sign(x) | Returns 1 if $x>0$, -1 if $x<0$, 0 if $x=0$ | 向量 |
| Signedpower(x,e) | Sign(x)*(abs(x))^e | 向量 |
| Pasteurize(x) | 信号去噪。如果为无穷值或不在全域中, 则设为 NaN | 向量 |
| Log(x) | 自然对数 | 向量 |
| Ts_RanR(x,n) | 对过去 n 天中同样虚拟仪器中的 x 值排序, 返回当前值排列顺序。排列值在 0.0 ~ 1.0 之间, 与 Rank(x) 中的描述一致 | 向量 |
| Ts_Skewness(x,n) | 计算过去 n 天 x 的偏度 | 标量 |
| Ts_Kurtosis(x,n) | 计算过去 n 天 x 的峰度 | 标量 |
| Ts_Moment(x,k,n) | 计算过去 n 天 x 的 k 阶中心矩 | 标量 |

续表

| 运算符号 | 描 述 | 返回类型 |
|--------------------|--|------|
| IndNeutralize(x,y) | 根据 y 确定的分组中和 Alpha 中的 x。例如, IndNeutralize(x,industry) 与 IndNeutralize(x,subindustry) 分别可以中和产业和细分产业。使用 IndNeutralize(x,1) 来中和市场 | 向量 |

表 26.3 WebSim™ 模拟总结报告

| 年份 | 买入量 | Long Count | Short Count | Pnl | 夏普率 | 稳定率 | 收益率 | 回报率 | 换手率 | 利润 (bpm) |
|-----------|-------|------------|-------------|-------|------|------|--------|-------|--------|----------|
| 2009 | 20.0M | 1311 | 1212 | 1.86M | 3.55 | 2.13 | 19.67% | 1.18% | 54.76% | 7.18 |
| 2010 | 20.0M | 1395 | 1302 | 1.35M | 3.11 | 1.54 | 13.40% | 1.38% | 54.71% | 4.90 |
| 2011 | 20.0M | 1385 | 1341 | 1.15M | 1.94 | 0.88 | 11.38% | 2.65% | 55.41% | 4.11 |
| 2012 | 20.0M | 1407 | 1332 | 948K | 2.44 | 1.00 | 9.48% | 1.14% | 56.19% | 3.37 |
| 2013 | 20.0M | 1426 | 1293 | 706K | 1.89 | 0.67 | 7.01% | 1.18% | 55.61% | 2.52 |
| 2009—2013 | 20.0M | 1386 | 1297 | 6.02M | 2.55 | 1.19 | 12.10% | 2.65% | 55.34% | 4.37 |

第 27 章

Alpha 与 WebSim™ 的原理

作者：WebSim™ 团队

伴随着信息技术、信息处理、自动交易等的发展，市场不断靠近 100%有效市场，但从未达到。我们从这种非有效性中搜寻套利机会。

预测股票的未来回报非常困难。但由噪声数据导致的单只股票问题不会伤害到整个市场策略。每天预测许多只股票的过程被称为统计套利，也就是 WebSim™ 所做的事情。

统计套利也被称为 Alpha。一个好的统计模型通常包含大量的证券（许多是短期持有的），从而在长期运作中获得正向收益。

1. Alpha和WebSim™

Alpha 是对金融工具（如股票、远期和约等）有预测功能的数学模型，可使用 WebSim™ 来模拟。这个模型由数据（如收盘价、开盘价、成交量等）和数学表达式（+/-、StdDev(x,n)、回归等）组成。

WebSim™ 是个基于网页的全球金融市场模拟器，有助于 Alpha 研究。

WebSim™ 输入为 Alpha 表达式或者 Python 代码，可打印出 PnL 作为输出。根据输入表达式，对金融工具每天的历史数据进行计算，由此就相应地创建了证券组合。WebSim™ 根据每种金融工具表达式的值对其进行投资，也会基于所打印出的 PnL 图，选择合适的空头或者多头头寸，如图 27.1 所示。

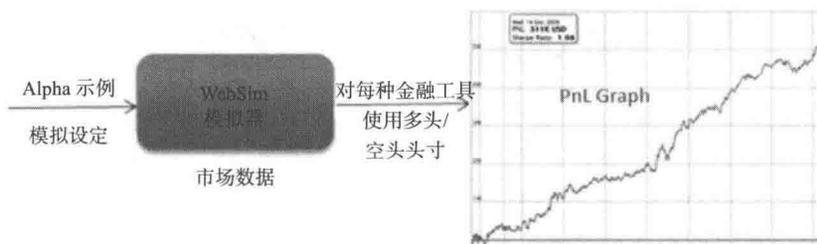


图 27.1 在 WebSim™ 模拟器中产生的 PnL 图

这种 PnL 产生于历史数据，所以被称为样本内 PnL。在 Alpha 被提交后，会用一段实时（样本外）数据对其进行其他标准的检查。在这种情况下产生的 PnL 如图 27.2 所示，被称为样本外 PnL。

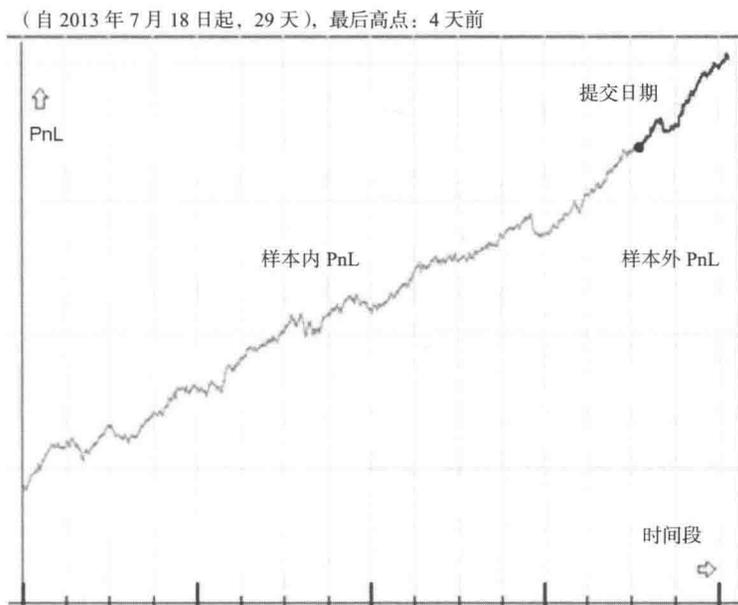


图 27.2 在 WebSim™ 中产生的样本外 PnL 图

在将不同 Alpha 组合为证券组合时，不同 Alpha 的权重也是不同的。当计算损益时，这些权重就被调整到仓位规模，它们不是恒定的，总是基于当前信息和一些变量的变化历史（如价格、成交量等）而变化。

2. Alpha来源

可以在网络上通过研究论文、金融期刊和技术指标找到 Alpha 相关的看法。

技术指标被用于分析短期价格波动，它们派生自股票/资产的演变的价格活动。通过查看过去的模式，可预测证券未来的价格水平或者大概的价格走势。常用的技术指标有：相对强度指标（RSI）、资金流量指标（MFI）、指数平滑异同平均线（MACD）、布林带（Bollinger Bands）等。你可以从一些网站如 Stock Charts、Incredible Charts 等找到这些技术指标的解释、公式和相应的推演。

SSRN 是发现 Alpha 创意的好地方。此外 Seeking Alpha 和 Wilmott 网站也不错。博客也是搜索 Alpha 的不错的渠道，如 Epchan, AuTraSy 等。

需要就如下概念进行探讨。

- 价格移动和技术指标；
- 波动率测量：历史波动率、隐含波动率、波动率指数、日内波动率等；
- 成交量与价格：成交量与绝对价格变化正相关等；
- 短期和长期趋势。

用户应该注意到以上对任何第三方站点的引用只是为方便用户而已。用户应该阅读并遵守使用这些网站的条款。

3. 中性化

中性化是指每一组都有了最初的 Alpha 数值后，在每一组内进行归一化（所有数值都减去平均值）的操作。“组”可以是整个市场，也可以是行业或子行业板块（基于美国标准产业分类代码或北美工业分类系统代码）。这样做

不是为了选择某个组，而是为了将股票的未來收益进行相关处理。通过中性化操作，整个证券组合有了中性的头寸（一半看多，一半看空）。这些工作可以确保证券组合躲避市场波动并且消除一些错误信号。

4. 样本空间

模拟样本空间包含了金融工具及时间数据。WebSim™ 只交易流通股并提供了 TOP3000、TOP2000、TOP1000 等标准的样本空间。

TOP- N 样本空间为在指定指标情况下，在过去三个月内平均指标最高的 N 只股票。比如指定流动性的情况，TOP3000 是在三个月期内流动性最高的 3000 只股票的集合；TOP2000 是在三个月期内流动性最高的 2000 只股票的集合；依次类推，TOP2000 是 TOP3000 股票的子集。

第 28 章

理解 WebSim™ 的工作流程

作者：WebSim™ 团队

假设市场数据组成一个矩阵，每一行代表一个数据，每一列代表一只股票。在 TOP3000 样本空间中的股票收盘价数据的矩阵如表 28.1 所示。

表 28.1 在 TOP3000 样本空间中股票收盘价数据的矩阵

| 工具日期 | MSFT | HOG | AAPL | GOOG | PG | ... |
|----------|--------|-------|--------|--------|-------|-----|
| 20100104 | 30.95 | 25.46 | 214.01 | 626.75 | 61.12 | ... |
| 20100105 | 30.96 | 25.65 | 214.38 | 623.99 | 61.14 | ... |
| 20100106 | 30.77 | 25.59 | 210.97 | 608.26 | 60.85 | ... |
| 20100107 | 30.452 | 25.80 | 210.58 | 594.10 | 60.52 | ... |
| 20100108 | 30.66 | 25.53 | 211.98 | 602.02 | 60.44 | ... |
| ... | ... | ... | ... | ... | ... | ... |

上述股票成交量的矩阵如表 28.2 所示。

表 28.2 在 TOP3000 样本空间中股票成交量数据的矩阵

| 工具日期 | MSFT | HOG | AAPL | GOOG | PG | ... |
|----------|-------------|-------------|-------------|-------------|-------------|-----|
| 20100101 | 3.84142e+07 | 2.90391e+06 | 1.76332e+07 | 1.95796e+06 | 9.19087e+06 | ... |
| 20100102 | 4.97589e+07 | 2.80537e+06 | 2.14966e+07 | 3.00786e+06 | 8.65051e+06 | ... |
| 20100103 | 5.81823e+07 | 3.2833e+06 | 1.97199e+07 | 3.98063e+06 | 9.90891e+06 | ... |
| 20100104 | 5.05643e+07 | 2.52213e+06 | 1.70403e+07 | 6.41802e+06 | 8.97275e+06 | ... |
| 20100105 | 5.12013e+07 | 3.52372e+06 | 1.59956e+07 | 4.72474e+06 | 8.46696e+06 | ... |
| ... | ... | ... | ... | ... | ... | ... |

如果你在 WebSim™ 首页输入一个 Alpha 表达式，设置模拟期为 5 年，那么 WebSim™ 将会以最近 5 年市场和股票的数据矩阵输入 Alpha 表达式来进行评估。

对每一个时间间隔（天），表达式将访问所有数据，每天生成一行新的矩阵数据，直到到达最近的一天。

表达式的作用是将输入矩阵输出为权重的向量。

输出结果为样本空间里的金融工具所对应的权重的向量，如表 28.3 所示。

表 28.3 在 TOP3000 样本空间中每一只股票对应的 Alpha 权重值

| 工具 | MSFT | HOG | AAPL | GOOG | PG | ... |
|----------|--------|--------|---------|--------|---------|-----|
| Alpha 权重 | 0.2423 | 0.5675 | -0.4745 | 0.4734 | -0.5684 | ... |

每天都会计算表达式的数值，将结果以向量形式输出，向量的值反映出你想买（或者卖）的金融工具（股票或者合约）的多少。股票对应的数值不是你想买的股数，而是相对于其他股票的持仓权重。正数表示买入，负数表示卖出，NaN（非数值）将意味着此金融资产没有被分配到权重（也就是没有被分配到资金）。当 Alpha 仓位为 NaN 时，也就不会有损益。基于当天仓位，WebSim™ 会计算出对应损益并提供给用户。在默认情况下，WebSim™ 会使权重规范化，并设置证券组合的市值为 2000 万美元。

1. 模拟设置面板

Alpha 有资产类别、延迟等特征，在模拟开始前，通过模拟设置面板可以指定类别。设置面板（齿轮图标）可以在 WebSim™ 首页的右上角找到。不同的设置参数在图 28.1 中给出。



图 28.1 在 WebSim™ 模拟中的参数设置

- **资产类别:** WebSim™ 目前支持两种金融工具: 股票和期货。其他资产类别如货币、指数股票型基金 (ETF) 等将会很快加入进来。
- **区域和样本空间:** WebSim™ 目前可以仿真美国、欧洲、亚洲各个主要市场。样本空间是由 WebSim™ 准备的一套交易工具。比如, “US:TOP3000” 代表在美国市场的最有流动性的前 3000 只股票, 而 “US: EQ Index Futures” 代表的是美国股指期货。
- **延迟:** 指的是与决策时间相关的数据的可得性。延迟 1 (默认设置) 意味着 Alpha 将会使用昨天的数据 (价格、回报等)。延迟 0 意味着 Alpha 将会用今天的数据。尽管提前知道今天的收益是好事, 但我们必须把延迟包括在内以使得模拟贴近现实。
- **衰减:** 结合今天和 n 天之前的值, 实现过去 n 天的线性衰减函数, 执行如下操作:

$$\text{Decay_linear}(x,n) = (x[\text{date}] * n + x[\text{date} - 1] * (n - 1) + \dots + x[\text{date} - n + 1]) / (n + (n - 1) + \dots + 1)$$

- **中性化:** 是使用策略对市场、行业、子行业进行中性操作。当针对市场中性化时, 它执行以下操作:

$$\text{Alpha} = \text{Alpha} - \text{mean}(\text{Alpha})$$

基本上, 它使 Alpha 向量的平均值为零, “空仓” 是因为 “敬畏市场”。换句话说, 多头仓位和空头仓位完全相互抵消, 使我们的策略为市场中性。

当针对 “行业” 或 “子行业” 中性化时, 在 Alpha 向量中的所有金融工具

被倒逼对应行业或子行业更小的段 (bucket) 里来, 然后对每个段做中性处理。

- **最大权重:** 证券组合里, 所有金融工具 (股票或期权) 的最大权重。当设置为零时, 没有限制。
- **单位检查:** 它确保与表达式里单位 (美元等) 一致。假设 Alpha 是个无意义的数学表达式, 如: $adv20+close$, 其实是成交量+价格, 这个 Alpha 表达式不能被视为一个好的表达式。WebSim™ 打开单位检查功能后, 可以在程序内部检查单位是否一致, 从而发现这种 Alpha。
- **模拟持续时间:** 表示回测需要的历史数据。你可以将股票类设为 2 ~ 5 年, 期货类设为 5 ~ 10 年。
- **代码语言:** 选择 “Expression.” 在线输入一个 Alpha 表达式, 选择 “Python.” 输入 Alpha 的 Python 脚本。

2. 运行你的第一个 Alpha

要运行你的第一个模拟, 首先打开 WebSim™ 主页。单击右上角的齿轮图标, 打开设置面板。选择 “5 年” 为模拟持续时间, “US:TOP3000” 为区域和样本空间, “NONE” 为中性, 保存设置。在 Alpha 表达式文本框中, 暂时输入 “1” (这样结果向量会为所有股票分配同样的权重)。模拟结果页面会显示 Alpha 的累计利润。而且图表还可以被缩放到更短时间区间 (如一个月、三个月、一年等)。

Alpha 的累计利润图和夏普比率图如图 28.2 和图 28.3 所示。每一年累计的模拟结果 (回报、夏普比率、下降等) 如表 28.4 所示。

累计利润图包括两个图, 一个是对应时间的损益, 另一个是对应时间的夏普比率。表 28.4 中显示了每年累计值以及整个模拟周期内的结果。好的 Alpha 应该有持续增长的收益, 高年度回报、高夏普比率、高获利天数百分比和高单位资金利润, 而且换手率和回撤都要低。更重要的是在累计利润图里不能有大幅波动。如果标准差很低, 那么相应的波动就小。如果图中显示高波动率/无序, 那么尽管收益很高, 人们也会认为此 Alpha 不够好。

好 Alpha 满足如下条件:

- 对于延迟为 0 的 Alpha, 夏普比率应该大于 3.95, 对于延迟为 1 的

Alpha，夏普比率应在 2.5 以上；

- 换手率低（50%或者更低）；
- 回撤小于 10%。



图 28.2 累计利润图

如表 28.4 所示（已经被缩放到显示 2011 年 1 月到 10 月），对于 Alpha 表达式 1，这是个平均的模拟结果。从图 28.3 可看出，波动率是高的。从表 28.4 中可以看到 2011 年损益下降，夏普比率落到阈值以下，回撤巨大，每美元交易的收益是负的。因此，由于缺乏连续的正收益，这是个一般的模拟结果。

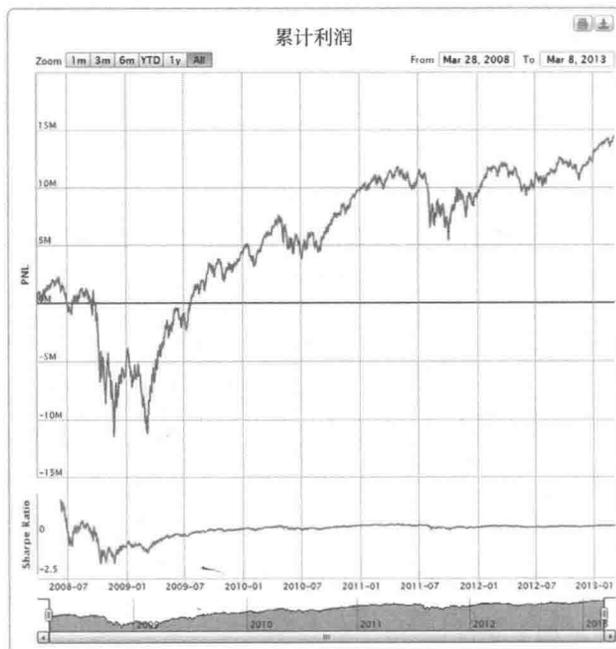


图 28.3 夏普比率图

表 28.4 累计计算的模拟结果（收入、夏普比率、回撤等）

| 年份 | 买入量 | PNL | 年收益率 | 夏普比率 | 最大时效 | 收益日百分比 | 日换平率 | 每笔交易的利润 |
|-----------|-------|-----------|---------|-------|---------|--------|-------|----------|
| 2008 | 2.0E7 | -5.04E6 | -65.33% | -0.66 | 138.47% | 49.74% | 0.50% | -26.09 ¢ |
| 2009 | 2.0E7 | 9.87E6 | 98.30% | 1.39 | 74.18% | 55.38% | 0.53% | 37.36 ¢ |
| 2010 | 2.0E7 | 5.56E6 | 55.17% | 1.22 | 37.28% | 57.54% | 0.32% | 34.59 ¢ |
| 2011 | 2.0E7 | -878.02E3 | -8.78% | -0.15 | 66.91% | 54.00% | 0.33% | -5.36 ¢ |
| 2012 | 2.0E7 | 3.33E6 | 33.43% | 1.05 | 29.27% | 51.00% | 0.31% | 21.91 ¢ |
| 2013 | 2.0E7 | 1.99E6 | 108.00% | 4.53 | 7.23% | 71.74% | 0.30% | 73.13 ¢ |
| 2008—2013 | 2.0E7 | 14.83E6 | 29.87% | 0.48 | 138.47% | 54.39% | 0.39% | 15.42 ¢ |

3. 其他Alpha示例

下面再以可以得到的市场数据举例。我们使用 close（日收盘价）和 sharesout（当日流通股）组成 Alpha 表达式。现在再次打开 WebSim™ 主页面，选中设置齿轮按钮，选择模拟持续时间为“5 年”，区域和样本空间为“US:TOP3000”，中性为“none”，保存设置。在 Alpha 文本框中，输入“sharesout*close”，然后提交表达式。当日流通股乘以股票的收盘价表示股票的市值。由于表示式是按每一天、每一只股票来计算的，损益看起来会类似于商业基准，以雅虎财经能找到的 Russell 3000 为基准，确保你开始了一个新的浏览器（WebSim™ 之外）来搜寻指数，结果会类似图 28.4。

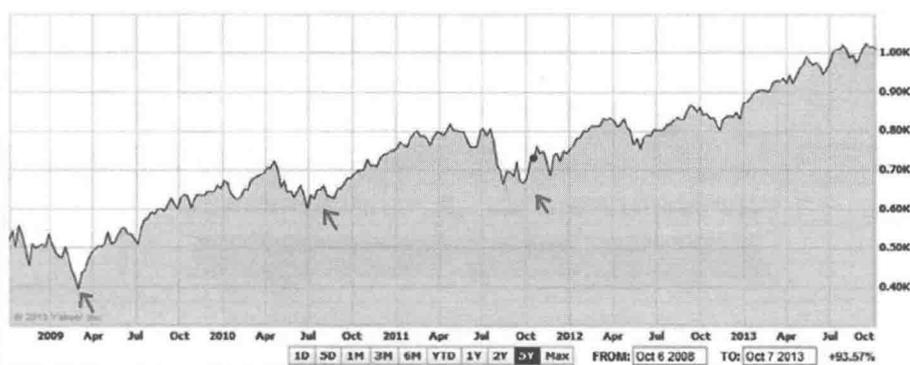


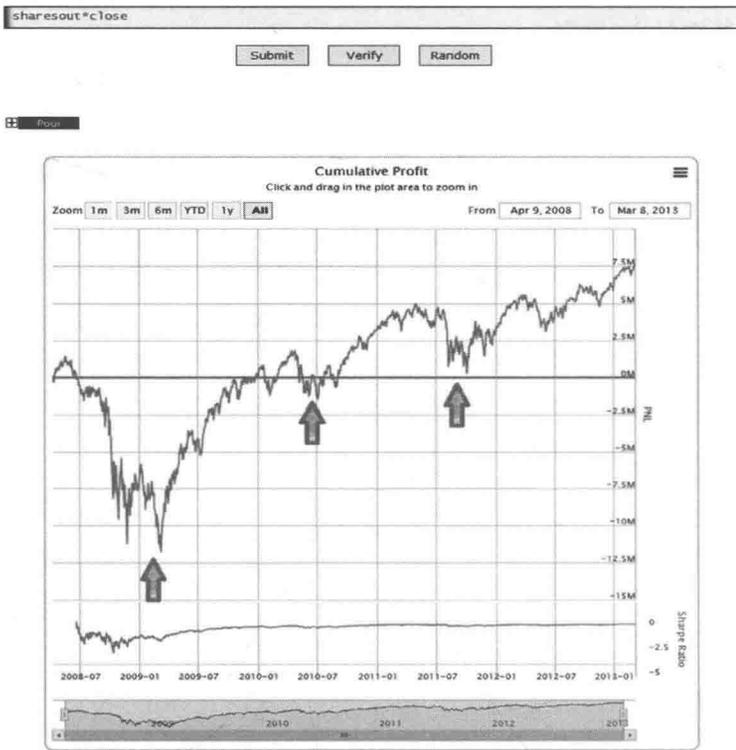
图 28.4 Russell 3000 指数图

从图 28.4 中可以看到市场中的三次下降，一次在 2009 年 4 月附近，第二

次在 2010 年 7 月，第三次在 2011 年 10 月。现在，如果与 WebSim™ 的模拟结果比较（见图 28.5），就会发现两张图在同样的月份都有所下降。

让我们再仔细看图 28.5 中的模拟结果，每年仓位为 2000 万美元。在 2009 年，产生的收益是 675 万美元，以百分比表达的年度收益率是 67.01%，这是在 5 年内产生的最好收益率。再看夏普比率，在 2010 年和 2011 年有显著的下降（见图 28.5）。在 2011 年的最大回撤对应于图 28.5 中所示的下降。

你可以试着通过杂志、网站和博客来搜寻 Alpha 表达式，或者以本章中给出的表达式例子作为起点。



| 年份 | 买入量 | PNL | 年收益率 | 夏普比率 | 最大时效 | 收益日百分比 | 日换平率 | 每笔交易的利润 |
|-----------|-------|--------|---------|-------|---------|--------|-------|-----------|
| 2008 | 20.0M | -6.65M | -89.81% | -0.96 | 126.47% | 51.35% | 2.15% | -8.37 bpm |
| 2009 | 20.0M | 6.75M | 67.01% | 1.15 | 53.25% | 55.95% | 1.64% | 8.20 bpm |
| 2010 | 20.0M | 3.15M | 31.27% | 0.80 | 32.44% | 55.56% | 1.08% | 5.79 bpm |
| 2011 | 20.0M | -170K | -1.65% | -0.03 | 47.23% | 54.76% | 1.21% | -0.28 bpm |
| 2012 | 20.0M | 3.11M | 31.12% | 1.14 | 24.58% | 52.00% | 0.97% | 6.45 bpm |
| 2013 | 20.0M | 1.45M | 78.65% | 3.61 | 6.23% | 65.22% | 0.91% | 17.23 bpm |
| 2008-2013 | 30.0M | 7.68M | 25.60% | 0.92 | 131.27% | 54.48% | 1.35% | 3.23 bpm |

图 28.5 Russell 3000 指数的 WebSim™ 模拟结果图

第 29 章

用户帮助手册

作者：WebSim™ 团队

本章是构建 Alpha 时用到的所有操作符与数据字段的参考指南。

1. 可用的市场数据

表 29.1 列出了所有可以在 Alpha 表达式中使用的市场数据。

表 29.1 股权的价格成交量数据

| 数据名 | 描述 | Python 用法 |
|-----------|---------------|--------------------------|
| open | 每日开盘价 | dr.GetData ("open") |
| close | 每日收盘价 | dr.GetData ("close") |
| high | 每日最高价 | dr.GetData ("high") |
| low | 每日最低价 | dr.GetData ("low") |
| vwap | 每日成交量加权平均价 | dr.GetData ("vwap") |
| volume | 每日成交量 | dr.GetData ("volume") |
| returns | 每日收益 | dr.GetData ("returns") |
| adv20 | 过去 20 天平均日交易量 | dr.GetData ("adv20") |
| sharesout | 每日流通股 | dr.GetData ("sharesout") |

开盘价、最高价、最低价、收盘价、成交量、成交量加权平均价和日流通股等数据已经针对企业变动行为（例如分配股息、股票拆分、反向拆分等）做了复权，可以用来进行 Alpha 计算。相对而言，原始价格和成交量数据不能直接用在 Alpha 计算中，因为会使 Alpha 不稳定。

这可以通过一个例子来解释：假设 GOOG（Google）在 4 月 2 日按 2:1 拆分，其当时价格是 1200 美元。

在 4 月 3 日，价格可能为 600 美元。如果没有复权，则单独的原始价格会扭曲价格行为。

向前复权的数据也必须每天更新。考虑到效率，WebSim™ 只会对一定天数内的数据进行向前复权。这个固定的时间长度被定义为“回看日”。

公司基本面数据如表 29.2 所示，是基于公众公司按监管机构要求填写的季度/年度报告。

表 29.2 股份公司基本面数据

| 数据名 | 描述 | Python 用法 |
|--------------------|-------------|------------------------------|
| sales | 季度销售额 | dr.GetData("sales") |
| sales_growth | 季度销售增长值 | dr.GetData("sales_growth") |
| sales_ps | 每股季度销售额 | dr.GetData("sales_ps") |
| income | 季度净收入 | dr.GetData("income") |
| eps | 每股收入 | dr.GetData("eps") |
| cashflow | 季度现金流 | dr.GetData("cashflow") |
| cashflow_sales | 销售所得季度现金流 | dr.GetData("cashflow_sales") |
| cashflow_op | 运营所得季度现金流 | dr.GetData("cashflow_op") |
| cashflow_fin | 金融活动所得季度现金流 | dr.GetData("cashflow_fin") |
| cashflow_invst | 投资所得季度现金流 | dr.GetData("cashflow_invst") |
| cashflow_dividends | 分红所得季度现金流 | dr.GetData("dividends") |
| assets_curr | 流动资产 | dr.GetData("assets_curr") |
| assets | 总资产 | dr.GetData("assets") |
| equity | 普通股股东权益 | dr.GetData("equity") |
| debt_lt | 长期债务 | dr.GetData("debt_lt") |
| debt_st | 短期债务 | dr.GetData("debt_st") |
| debt | 总债务 | dr.GetData("debt") |

续表

| 数据名 | 描述 | Python 用法 |
|--------------------|------------|----------------------------------|
| liabilities_curr | 流动负债 | dr.GetData("liabilities_curr") |
| liabilities | 债务 | dr.GetData("liabilities") |
| EBITDA | 税息折旧及摊销前利润 | dr.GetData("EBITDA") |
| pref_dividends | 优先红利 | dr.GetData("pref_dividends") |
| capex_fix | 固定资产资本支出 | dr.GetData("capex_fix") |
| capex | 资本支出 | dr.GetData("capex") |
| operating_income | 运营收入 | dr.GetData("operating_income") |
| operating_expense | 运营费用 | dr.GetData("operating_expense") |
| cogs | 销货成本 | dr.GetData("cogs") |
| bookvalue_ps | 每股价值 | dr.GetData("bookvalue_ps") |
| return_assets | 资产回报 | dr.GetData("return_assets") |
| return_equity | 股东权益回报 | dr.GetData("return_equity") |
| inventory | 总库存 | dr.GetData("inventory") |
| inventory_turnover | 季度库存周转率 | dr.GetData("inventory_turnover") |

2. 可用的运算符

表 29.3 列出了可以用在 Alpha 表达式中的运算符。

表 29.3 运算符

| 运算符 | 描述 | 返回类型 |
|----------------------|---|------|
| +, -, *, /, ^ | 加、减、乘、除、乘方 | 向量 |
| <, <=, >, >=, ==, != | 逻辑比较符 | 向量 |
| , &&, ! | 逻辑或、与、非 | 向量 |
| cond?expr1:expr2 | 如果条件为真，则输出 1；或者输出 2。比如，close<open?close:open | 向量 |
| Rank(x) | 对同一天内不同股票的 x 排序，并归一成 0 至 1 的等差数列 比如，指定 6 只股票的收盘价[20.2,15.6,10.0,5.7,50.2,18.4],rank(close)返回 [0.8,0.4,0.2,0.0,1.0,0.6] | 向量 |
| Min(x,y) | 向量 x 和 y 的最小值（类似于 R 中的 pmin 函数）。需要 2 个向量输入值，返回值为单个向量。结果中的第一个元素是在所有实际输入参数的第一个向量的最小值，结果中的第二个元素是在所有实际输入参数的第二个元素的最小值，依次类推 | 向量 |

续表

| 运算符 | 描述 | 返回类型 |
|----------------------------|---|------|
| Max(x,y) | 向量 x 和 y 的最大值 (类似于 R 中的 pmax 函数)。这需要 2 个向量输入值, 返回值为单个向量。结果中的第一个元素第一个向量的最大值, 结果中的第二个元素是第二个向量的最大值, 依次类推 | 向量 |
| StdDev(x,n) | 过去 n 天向量 x 的标准差。注意 n 必须小于 256 | 标量 |
| Correlation(x,y,n) | 过去 n 天向量 x 、 y 相关性。注意 n 必须小于 256 | 标量 |
| Sum(x,n) | 过去 n 天向量 x 值的总和。注意 n 必须小于 256 | 标量 |
| Covariance(x,y,n) | 过去 n 天向量 x 、 y 的协方差。注意 n 必须小于 256 | 标量 |
| CountNans(x,n) | 过去 n 天向量 x 的非数值的数量。注意 n 必须小于 256 比如, CountNans(close-open)^0.5,22), 如果 (close>open), 那么 (close-open)^0.5 就不是非数值, 或者如果 (close<open), 那么 (close-open)^0.5 就是一个非数值。因此, 上述代码基本就是统计在过去 22 天内有多少次收盘价小于开盘价 | 标量 |
| Abs(x) | 绝对值 | 向量 |
| Delay(x,n) | n 天前的 x 值。注意 n 必须小于 256 | 向量 |
| Step(x) | 对于所有参数, 今天是 x , 昨天是 $x-1$, 依次类推。Step(x) 为每一个参数创建一个向量, x 的值是今天, $x-1$ 代表昨天, 依次类推 比如, Step(1250) 会为实际输入参数创建如下数列: [...,-3,-2,-1,0,1,2,..., 1249,1250], 最后一个值对应“今天。” 假设输入表达式: sum(Step(5)*close,5), 会这样计算, 5*close[今天]+4*close[今天-1]+...+1*[今天-4], 也就是计算过去 5 天一系列收盘价的加权平均值 | 向量 |
| Delta(x,n) | $x[\text{date}]-x[\text{date}-n]$, 注意 n 必须小于 256 | 向量 |
| Decay_linear(x,n) | 过去 n 天的衰减 $\text{Decay_linear}(x,n)=x[\text{date}]*n+x[\text{date}-1]*(n-1)+\dots+x[\text{date}-n-1]/(n+(n-1)+\dots+1)$ | 向量 |
| Decay_exp(x,f,n) | 过去 n 天的指数衰减函数, 这里 f 是平滑因子, 可是被赋予一个小于 1 的值。Decay_exp(x,f,n)=(x[date]+x[date-1]*f+...+x[date-n-1]*(f^(n-1)))/(1+f+...+f^(n-1)) | 向量 |
| Product(x,n) | 过去 n 天向量 x 的乘积 $\text{Product}(x,n)=x[\text{date}]*x[\text{date}-1]*\dots*x[\text{date}-n-1]$ | 标量 |
| Tail(x,lower,upper,newval) | 如果 x 值在 lower 和 upper 之间, 则将 x 值设为 newval | 向量 |
| Ts_Min(x,n) | 过去 n 天 x 的最小值。注意与 Min 不一样 | 向量 |
| Ts_Max(x,n) | 过去 n 天 x 的最大值。注意与 Max 不一样 | 向量 |

续表

| 运算符 | 描述 | 返回类型 |
|---------------------------------|--|------|
| Sum_i(expr,var,start,stop,step) | 遍历变量（分步从开始到结束）与每次迭代的计算表达式（可能包含变量的表达式），然后对所有值求和 比如：Sum_i(Delay(close,i)*i,i,2,4,1)将等同于 Delay((close,2)*2+Delay(close,3)*3+Delay(close,4)*4) | 标量 |
| Call_i(expr,var,subexpr) | 在表达式中用 subexpr 代替 var，然后计算表达式 比如 Call_i(x+4,x,2+3)将等同于 (2+3)+4 | 向量 |
| Sign(x) | 如果 x 大于 0，则返回 1；如果 x 小于 0，则返回 -1；如果 $x=0$ ，则返回 0 | 向量 |
| SignedPower(x,e) | $\text{Sign}(x) * (\text{Abs}(x))^e$ | 向量 |
| Pasteurize(x) | 信号去噪。如果为无穷值或不在股票空间中，则将其设为 NaN | 向量 |
| Log(x) | 自然对数 | 向量 |
| Ts_Rank(x,n) | 过去 n 天对同一只股票中的 x 值进行排序，接着返回当前值的排列顺序。排序值在 0.0 与 1.0 之间，与 Rank(x)中的描述一致 | 标量 |
| Ts_Skewness(x,n) | 计算过去 n 天 x 值的偏度 | 标量 |
| Ts_Kurtosis(x,n) | 计算过去 n 天 x 值的峰度 | 标量 |
| Ts_Moment(x,k,n) | 计算过去 n 天 x 的 k 阶中心矩 | 向量 |
| IndNeutralize(x,y) | 根据 y 确定的分组来中和 Alpha 中的 x 。比如，IndNeutralize(x,industry)和 IndNeutralize(x,subindustry) 分别可以中和产业与细分产业。用 IndNeutralize(x,1)来中和市场 | 向量 |
| Scale(x) | 缩放 Alpha x 使得其仓位大小为 1,也就是在所有工具中 abs(x)的总和为 1。为缩放到不同的仓位大小，如 1000，使用 Scale(x)*1000 | |

注意：以上作为函数参数的 x 、 y 都是向量，使用时用市场数据替代。

另外注意，这里的 256 指的是回看的天数。比如，如果计算过去 10 天的平均 CPS 值，则在模拟中计算第一天的数据时，需要提取前 10 天的收盘价。由于 10 小于默认的 256 天的回看天数，就能提取到对应数值，如果要计算过去 257 天的平均收盘价，就不能提取到模拟周期的第一天，即实现在第 257 天的数据。

第 30 章

输出结果说明与 Alpha 信号库

作者：WebSim™ 团队

本章介绍了其他一些用于反映 Alpha 性能的因子,对所有的 Alpha 信号进行追踪,并罗列模拟中的错误与警告信号。

1. 夏普比率分类

如图 30.1 所示,在模拟结果页面,在 Alpha 表达式的正下方、累计利润图的上方有一个反映 Alpha 表现的标签。如表 30.1 所示,根据 Alpha 夏普比率和适应性,标签有如下几个等级:超好、优、好、一般、差和极差。

The image shows a screenshot of a simulation result page. At the top, there is a label '优' (Excellent) in a dark box. Below it is a table with five columns: '百分等级' (Percentile), '总成交量' (Total Volume), '夏普比率' (Sharpe Ratio), '收益率' (Return Rate), and '适应率' (Adaptability Rate). The table contains three rows of data: '全部' (All) with 187 volume, 100% Sharpe, 98% return, and 95% adaptability; '延迟一天' (Delayed one day) with 187 volume, 100% Sharpe, 98% return, and 95% adaptability; and '美国 TOP 3000 延迟一天' (US TOP 3000 delayed one day) with 175 volume, 100% Sharpe, 98% return, and 95% adaptability.

| 百分等级 | 总成交量 | 夏普比率 | 收益率 | 适应率 |
|---------------------|------|------|-----|-----|
| 全部 | 187 | 100% | 98% | 95% |
| 延迟一天 | 187 | 100% | 98% | 95% |
| 美国 TOP 3000 延迟一天 | 175 | 100% | 98% | 95% |

图 30.1 Alpha 模拟结果示例页面

表 30.1 根据夏普比率的等级划分

| 等 级 | 延迟为 1 的 Alpha 的夏普比率 | 延迟为 0 的 Alpha 的夏普比率 |
|-----|---------------------|---------------------|
| 超好 | >4.5 | >6 |
| 优 | >3.5 | >5.25 |
| 好 | >3.0 | >4.50 |
| 一般 | >2.5 | >3.95 |
| 差 | ≥1 | ≥1 |
| 极差 | <1 | <1 |

请注意上述分类可根据公司的情况进行调整。用户可以从 WebSim™ 的帮助页面获取最新的定义。

如果你单击“+”看扩展结果，就会发现一个百分等级表，是用户提交信号的历史模拟数据的等级划分表。

- 所有：（第一行）指所有的历史模拟。
- 延迟-1/0：（第二行）指有同样延迟设置的历史模拟。
- 美国：前 N 名 延迟-1/0：（第三行）表示有同样延迟和样本空间设置的历史模拟。

相应的夏普比率、回报率和适应性的分级排序可以在表格中看到。

2. 夏普比率和信息比率

夏普比率和信息比率可以相互替代。

信息比率（IR）衡量模型的预测能力，指的是证券组合在基准回报（通常是一个指数）之上的回报和这些回报波动性的比率。也有人说这是衡量产生与基准相比较的额外回报的能力，但是也会去识别 Alpha 的一致性。IR 会判断 Alpha 是每个月都可以勉强战胜基准还是几个月胜过基准很多。IR 越高，Alpha 就越一致，完全一致是一种理想状态。夏普比率是 IR 统计的年化版本，也就是说， $\text{Sharpe} = \sqrt{252} \times \text{IR} \sim 15.8 \times \text{IR}$ ；这里 252 是美国一年交易日（市场开放的天数）的平均值。在样本外时，合格 Alpha 在延迟为 0 的情况下夏普比率至少为 3.95，在延迟为 1 的情况下夏普比率至少为 2.5。

3. 回报和适应性

回报是一只股票在特定时期的盈利或损失。回报由收入和相关投资的资

本收益组成。回报=年化损益 ÷ 仓位规模的一半。好的 Alpha 应该明显产生高回报的夏普比率。

Alpha 的适应性是收益、换手率和夏普比率的函数。好的 Alpha 有高适应性。

4. 模拟结果

如图 30.2 所示，累计利润图可以对应 1、3、6 个月、截止到现在、1 年和所有时间的 Alpha。在图形区，可以通过单击进行放大和拖动；也可以改变损益计算的开始与结束日期。在损益图的正下方是夏普比率图（夏普比率 vs 时间）。要确保损益图有向上的趋势，夏普比率高，回撤最小。

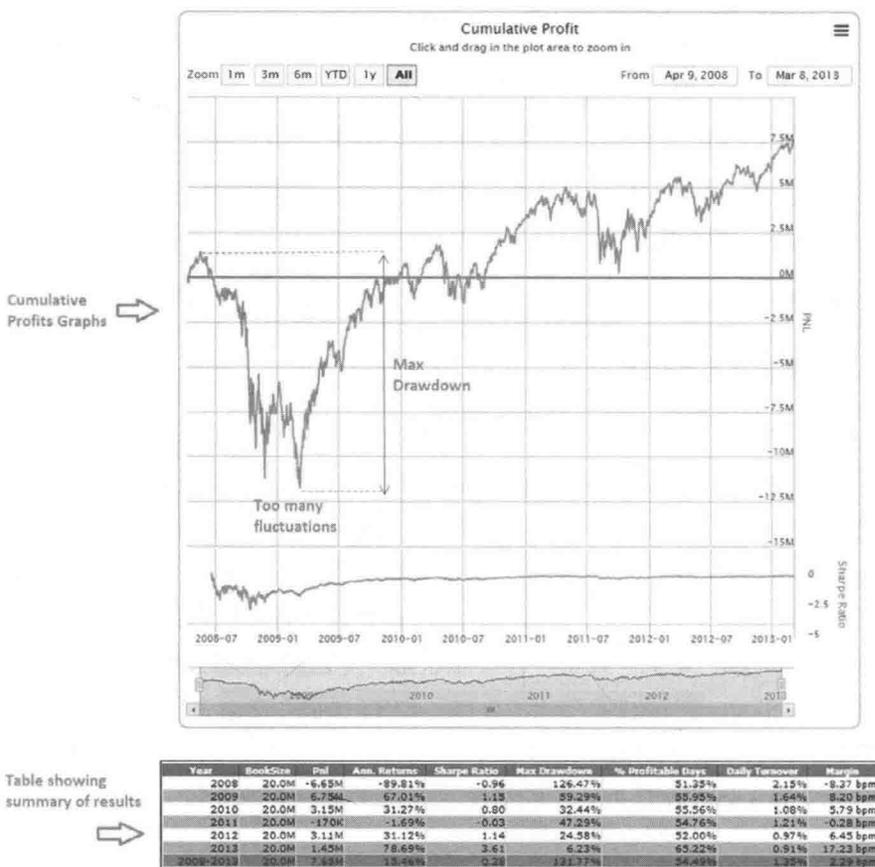


图 30.2 累计利润图

图 30.3 所示为对 Alpha 模拟结果的总结。

- **年**：这是数据模拟的年份。最后一行显示该 Alpha 的总表现。
- **仓位规模**：仓位规模指的是模拟仓位的大小，表示这是在模拟中每个工具上做多和做空的资金量。仓位规模是常量，被设置为 2000 万美元。因此在整个模拟中只有 2000 万美元可用。

WebSim™ 假设你用 1000 万美元将投资资产总额增加到 2000 万美元，这被称为用杠杆。结果（如回报、夏普比率）的计算基于 1000 万美元。

| 年份 | 购入量 | 损益 | 年收益率 | 夏普比率 | 最大回撤 | 盈利天数百分比 | 日换手率 | 单位美元的交易利润 |
|-----------|-------|---------|--------|------|-------|---------|--------|-----------|
| 2008 | 2.0E7 | 1.66E6 | 21.47% | 3.10 | 2.76% | 58.03% | 40.12% | 0.11 c |
| 2009 | 2.0E7 | 2.42E6 | 24.03% | 5.12 | 1.70% | 60.32% | 38.56% | 0.12 c |
| 2010 | 2.0E7 | 1.33E6 | 13.15% | 4.20 | 2.04% | 60.71% | 38.65% | 0.07 c |
| 2011 | 2.0E7 | 1.28E6 | 12.67% | 3.27 | 3.46% | 56.35% | 39.05% | 0.06 c |
| 2012 | 2.0E7 | 1.18E6 | 11.84% | 4.53 | 0.66% | 62.00% | 39.15% | 0.06 c |
| 2013 | 2.0E7 | 89.71E3 | 4.88% | 2.19 | 0.50% | 56.52% | 37.40% | 0.03 c |
| 2008-2013 | 2.0E7 | 7.98E6 | 13.96% | 3.71 | 3.46% | 55.44% | 38.00% | 0.08 c |

图 30.3 对 Alpha 模拟结果的总结

- **损益**：即盈利和亏损（又叫 PnL）是持仓和交易产生的资金（也就是交易者在这一年中亏损或挣到的钱），用美元表示。公式如下：

$$\text{每日损益} = (\text{仓位} \times \text{每日回报率}) \text{ 的求和}$$

这里的每日回报率是按金融工具的收盘价计算的：（今天的收盘价 ÷ 昨天的收盘价）- 1

- **年收益率**：已用资本的收益率。

$$\text{年度收益率} = \text{年化损益} \div \text{仓位规模的一半} \times 100\%$$

表示在观察期内你获利或损失的以百分比呈现的总量。

- **最大回撤**：用每天损益的最大连续下降值除以仓位规模一半得出的最大损失百分比。这表示现在是损益（最大亏损的钱）的低点。也就是说现在 P&L（损失最大）是最低点。
- **夏普比率**：夏普比率 = （在观察期间内的平均回报） ÷ 回报的标准差
- **盈利天数百分比**：获利天数和总交易天数的比率。
- **日换手率**：表示交易的频率。计算方法为交易量与持有股份价值的

比值。

$$\text{每日换手率} = \text{交易量} \div 2 \text{ 倍市值}$$

好的 Alpha 换手率低。

- 单位美元的交易利润：衡量交易的盈利能力。

$$\text{单位美元的交易利润} = \text{损益} \div \text{交易的美元}$$

该数字与交易成本无关。

5. 我的Alpha页面

图 30.4 展示到目前为止的表达式总结。

| Name | Creation | Simulation Duration | Universe | Entry | Exit | Multiplier | Volatility | Sharpe | Returns | PIV | Turnover | Drawdown | Margin |
|-------------|---------------------|---------------------|------------|-------|------|------------|------------|--------|---------|-------|----------|----------|----------|
| 22154326011 | 2014-02-10 09:49:34 | 5 | US_TOP3000 | 0 | 1 | long | 0 | 3.4 | 10.33% | 4.714 | 35.87% | 0.48% | 0.55 bpm |
| 22174442607 | 2014-02-17 09:17:44 | 5 | US_TOP3000 | 1 | 1 | market | 0 | 2.35 | 10.63% | 5.984 | 40.87% | 1.21% | 0.54 bpm |
| 22174439049 | 2014-02-17 09:34:19 | 5 | US_TOP3000 | 1 | 1 | market | 0 | 2.65 | 14% | 5.468 | 42.7% | 4.34% | 0.55 bpm |
| 22174439060 | 2014-02-17 09:51:17 | 5 | US_TOP3000 | 5 | 1 | market | 0 | 2.65 | 13.73% | 6.334 | 34.34% | 4.05% | 0.55 bpm |
| 22154326116 | 2014-02-11 09:21:21 | 5 | US_TOP3000 | 0 | 1 | market | 0 | 2.5 | 13.96% | 6.284 | 40.37% | 4.48% | 0.57 bpm |
| 22094604619 | 2014-02-09 09:44:44 | 5 | US_TOP3000 | 0 | 1 | market | 0 | 2.49 | 14.27% | 6.044 | 40.12% | 6.13% | 0.55 bpm |
| 22154326423 | 2014-02-10 09:44:46 | 5 | US_TOP3000 | 0 | 1 | market | 0 | 2.43 | 15.7% | 6.954 | 39.47% | 6.1% | 0.58 bpm |
| 22094603674 | 2014-02-09 09:31:16 | 5 | US_TOP3000 | 0 | 1 | market | 0 | 2.45 | 14.29% | 6.054 | 40.17% | 6.14% | 0.55 bpm |

图 30.4 模拟表达式的总结

在该页面提供的控件是：

- 收藏：可以通过单击右上角的 按钮来勾选复选框标记收藏的 Alpha。用于收藏的 Alpha 的流程同样适用于取消收藏的 Alpha。Alpha 当前的状态为显示为 的符号。
- 删除：可以通过勾选复选框后单击右上角的 按钮来删除 Alpha。
- 过滤器：用于查看所有 Alpha 的子集。可以用来过滤的参数有很多：名称（支持正则表达式）、样本空间、衰减、夏普比率、换手率、评级、代码、模拟时间、中性化、收益、回撤、创建时间、延迟、截断、损益和边际利润。（注意：边际利润是损益除以交易美元数量。这和在结果页面的“单位美元的交易利润”是一样的。单位 bpm 代

表基点（利润率）。

- **显示的数量：**使用“显示 N 个输入”的下拉菜单限制在一个页面显示数量。你可以使用“上一个/下一个”按钮或数字按钮在页面间导航。
- **查看代码/模拟：**将鼠标悬停在  按钮来查看表达式的代码。可以单击  按钮来模拟表达式，查看细节结果。
- **排序：**表格中每一列都可以列位排序。单击列表头部按照增序和降序来排列。

样本内：

此处有 Alpha 信号从首次模拟到最近的所有表现的总结，也就是说，Alpha 样本内的表现是从所有回测历史数据中得到的。这也就是模拟时，在结果页面看到的内容。显示的数据有：名称、创建时间、模拟时间、样本空间、衰减、延迟、中性化、截断、夏普比率、收益、损益、回报和边际利润。

样本外：

在样本外，延迟为 0 的 Alpha 应该有最小 3.95 的夏普比率，延迟为 1 的 Alpha 才是合适的。样本外测是 Alpha 提交后的表现，是 Alpha 在“真实世界”的表现。样本外标签显示 Alpha 在不同时间段的夏普比率（表现）：30 天、40 天、60 天到 1 年。我们对所选择的 Alpha 运行样本外模拟，结果如图 30.5 所示。显示数据为：名称、测试日期、一段时间的夏普比率（如 Sharpe 30 表示一个月，Sharpe 360 表示 1 年）、总样本外夏普比率、和你的 Alpha 的相关性统计（阈值从 0.5 到 0.7）和评级（比例从 0 到 100）。可以根据不同时间段的夏普指数追踪 Alpha 的样本外表现（见图 30.5）。

My Alphas

In-Sample Out-Sample

Show / hide filters

Show 20 entries
Showing 1 to 5 of 5 entries

| Name | Test Date | Sharpe30 | Sharpe40 | Sharpe60 | Sharpe90 | Sharpe110 | Sharpe150 | Sharpe210 | Sharpe360 | 0.5 Sharpe | Correlation to 0.7 | Rating |
|--|------------|----------|----------|----------|----------|-----------|-----------|-----------|-----------|------------|--------------------|--------|
|  021274056431 | 2014-05-09 | 0.443109 | 0.448218 | N/A | N/A | N/A | N/A | N/A | N/A | 0.228585 | 0 | 0.00 |
|  021114206116 | 2014-05-09 | 0.440798 | 0.289555 | N/A | N/A | N/A | N/A | N/A | N/A | 0.228103 | 0 | 0.00 |
|  021274488997 | 2014-05-09 | 0.002582 | 0.145156 | N/A | N/A | N/A | N/A | N/A | N/A | 0.130593 | 0 | 0.00 |
|  021274076369 | 2014-05-09 | 0.000214 | 0.207222 | N/A | N/A | N/A | N/A | N/A | N/A | 0.121933 | 0 | 0.00 |
|  021274490706 | 2014-05-09 | 0.10147 | 0.122828 | N/A | N/A | N/A | N/A | N/A | N/A | 0.11032 | 0 | 0.00 |

图 30.5 不同时期的夏普比率

相关性：

理解相关性的理念很重要的。我们已经有成千上万的 Alpha。如果你的 Alpha 与现有模型的相关性太高（比如相关性大于 0.7），那么你的模型对现有信号库来说或许不会有增值。同时从相关性上也可以明白 Alpha 的性质和种类。

到目前为止，在 WebSim™ 中，Alpha 的“样本外”的全部内容就是：Alpha 及与内部 Alpha 池之间相关性的计算，基于样本内和“样本外”结果以及样本外的夏普比率对 Alpha 进行排序评级。

夏普比率 X 字段表示样本外中 X 天内的夏普比，起始点为自从 Alpha 信号开始样本外测试之后第一次有收益的那一天。

6. 错误与警告

错误与警告包括：

(1) Syntax error in expression（表达式中的语法错误）

请确保你的表达式是符合逻辑的。只有表 30.2 和表 30.3 所示的符号（操作符和关键词）和数字（0~9）允许出现在 Alpha 表达式中。

表 30.2 操作符

| | |
|-------|--|
| 一元运算符 | 排列，绝对值，符号，对数，比例，级别 |
| 二元运算符 | +, -, *, /, ^, <, <=, >, >=, !=, , &&, Min, Max, StdDev, Ts_min, Ts_max, Ts_Rank, Ts_Kurtosis, Ts_Skewness, SignedPower, Sum, Delay, Delta, Product, Decay_linear, CountNans |
| 三元运算符 | ?, Correlation, Call_i, Ts_Moment, Decay_exp |
| 四元运算符 | Tail |
| 五元运算符 | Sum_i |

表 30.3 数据字段

| | |
|----------|--|
| 价格成交量数据域 | 开盘价，收盘价，最高价，最低价，成交量，回报，20 天移动平均价，流通股，成交量加权平均价，交易时间 |
| 基础数据域 | 销售额，销售额增长速度，每股销售额，收入，每股净收入，现金流，销售现金流，运营现金流，融资现金流，投资现金流，库存，资产，流动资产，股权，负债，长期负债，短期负债，债务，短期债务，利息折旧摊销前利润，优先股利，现金流股利，资本支出，固定资本支出，营运收入，营运费用，销售成本，每股价值，资产回报，股权回报，库存，库存周转 |

(2) Cannot retrieve simulation status (不能获取模拟状态)

这是服务器之间通信临时出现的问题。这条信息与用户的行为无关。

(3) Error simulating alpha (模拟 Alpha 的错误)

Python 模式下的语法错误或运行时错误。

(4) 因 Python 安全机制模拟中断运行。例如：在 Python 中输入“import sys”提交，会显示错误：Error: Simulation terminated due to Python security breach: “Forbidden to import sys.”

这些都是因为在 Python 代码中运行了被禁止的行为迫使 WebSim™ 阻止了模拟，这些行为包括试图打开和关闭文件，以及输入受限词汇，如“os”等，在 WebSim™ 上的 Alpha 开发中并不需要。

(5) 未完成的提交

错误：Error: “You have an outstanding submission, please wait 572 seconds to finish.”

当你提交一项表达式或 Python 代码而另外的模拟正在同步运行时，这个错误会发生。为跳过等待期，你必须退出后再次登录 WebSim™。

7. 快速摘要

图 30.6 所示的流程图展示了一项 Alpha 的生命周期。

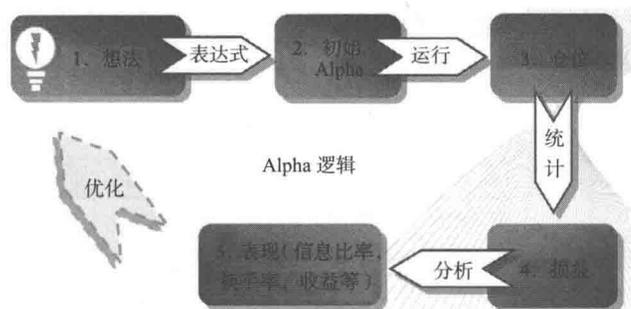


图 30.6 Alpha 的生命周期

首先，必须在互联网上仔细阅读博客、期刊和研究论文来形成一个想法。其次，在 WebSim™ 中输入 Alpha 表达式，然后在原始 Alpha 结果上进行操作（像截断、中性化、衰减）。WebSim™ 会模拟在设置面板中选择的所有股票的投资（看多或看空），然后产生损益。最后，在模拟结果页面，显示信号计算后的结果（夏普比率、换手率、收益）。如果这个 Alpha 不好，就优化此 Alpha。这个过程一直循环，直到产出适合的 Alpha 产品。

第 31 章

Alpha 教程

作者：WebSim™团队

本章为初学者提供了可用的表达式和 Python 的示例，还有对常见 Alpha 示例的描述，用来讨论在建立 Alpha 时应遵循的良好做法。

1. Alpha表达式示例

用不同的样本空间、延迟、中性化和其他的设置测试表 31.1 中的 Alpha 表达式。

表 31.1 Alpha 表达式示例

| 表 达 式 | 描 述 |
|------------------------------|--|
| 1/close | 将股票每日收盘价的倒数作为股票权重 对于每日收盘价较低的股票占权重大的情况，后面也有示例 后面有更多的通过表达式赋予大权重（更多资金分配）的示例 |
| volume/adv20 | 使用当日成交量和过去 20 天成交量的平均值的比值作为权重 |
| Correlation(close, open, 10) | 使用过去 10 天每日收盘价和每日开盘价的关联度作为权重 |
| open | 使用每日开盘价作为股票权重 |
| (high + low)/2 - close | 使用每日最高价和最低价的平均值与每日收盘价的差额作为权重 |

| 表达式 | 描述 |
|---------------------------------------|--|
| $vwap < close ? high : low$ | 如果成交量加权平均价小于每日收盘价，则采用每日最高价作为权重，否则就采用每日最低价作为权重 |
| Rank(adv20) | 使用最近 20 天的平均日交易量作为权重 |
| $Min(0.5*(open+close),vwap)$ | 取每日收盘价和每日开盘价的均值和成交量加权平均价的小数值作为权重 |
| $Max(0.5*(high+low),vwap)$ | 取每日最低价和每日最高价的均值和成交量加权平均价的大数值作为权重 |
| $1/StdDev(returns, 22)$ | 使用过去 22 天股票收益的标准差的倒数作为权重 |
| Sum(sharesout, 5) | 使用过去 5 天流通股的总和作为权重 |
| Covariance(vwap,returns, 22) | 使用过去 22 天成交量加权平均值和收益的协方差作为权重 |
| $1/Abs(0.5*(open+close) -vwap)$ | 使用每日开盘价和每日收盘价的平均值和成交量加权平均值的差额的倒数作为权重 |
| Correlation(vwap, Delay(close, 1), 5) | 使用成交量加权平均值和过去 5 天里，delay 为 1 时的每日收盘价的的相关度作为权重 |
| Delta(close, 5) | 使用每日收盘价和 5 日前收盘价的差额作为权重 |
| Decay_linear(sharesout*vwap, 5) | 使用 5 日内每日流通股和成交量加权平均值的乘积的线性衰减作为权重 |
| Decay_exp(close, 0.25, 5) | 使用过去 5 天平滑因子为 0.25 时收盘价的指数衰减作为权重 |
| Product(volume/sharesout, 5) | 使用过去 5 天里成交量和流通股的比值的乘积作为权重 |
| Tail(close/vwap, 0.9,1.1, 1.0) | 在收盘价和成交量加权平均值的比值小于 0.9 或者大于 1.1 的时候使用该比值数据作为权重，否则取 1 作为权重 |
| Sign(close-vwap) | 如果收盘价大于成交量加权平均值，使用 1 作为权重，否则使用 -1 |
| SignedPower(close-open, 0.5) | 使用收盘价和开盘价之差的绝对值的平方根作为权重 |
| Pasteurize(1/(close-open)) | 使用收盘价和开盘价之差的倒数的巴氏处理数据作为权重 (如果金融工具不在样本空间内或者数值为无限大，则设为 NaN) |
| Log(high/low) | 使用最高值和最低值比值的自然对数作为权重 |
| IndNeutralize(volume*vwap, 1) | 使用市场中性的成交量和成交量加权平均值的乘积作为权重 |
| Scale(close^0.5) | 使用收盘价的平方根的归一化数值作为权重 (经过归一化处理资金规模为 1) |
| Ts_Min(open, 22) | 使用过去 22 天里最低的开盘价作为权重 |
| Ts_Max(close, 22) | 使用过去 22 天里最高的开盘价作为权重 |
| Ts_Rank(volume, 22) | 使用过去 22 天里成交量的时间序列上的排名作为权重 |
| Ts_Skewness(returns, 11) | 使用过去 11 天里收益的偏离作为权重 |
| Ts_Kurtosis(returns, 11) | 使用过去 11 天里收益的峰值作为权重 |

续表

| 表达式 | 描述 |
|--|--|
| Ts_Moment(returns, 3, 11) | 使用过去 11 天里的三阶中心矩作为权重 |
| CountNans((closeopen)^0.5, 22) | 使用过去 22 天里, 收盘价和开盘价之差为 Nan 的次数作为权重 |
| Step(1250)*close | 使用收盘和 step (1250) 的乘积作为权重 |
| Sum_i(Delta(close,i),i,4,6,2) | 使用 i 从 4 到 6, 步长为 2 情况下 Delta (收盘价, i) 的和作为权重 |
| Call_i(Ts_Rank(x,5),x, close>vwap ? close : high) | 使用 Ts_Rank(x,5)作为权重 如果收盘价大于成交量加权平均值, x 为每日收盘价, 否则为每日最高价 |

2. 怎么用Python创建Alpha

若要在 WebSim™ 里用 Python 创建 Alpha, 则必须有很好的 Python 编程能力。

3. 数据库导入

记住, 用户提交的任何 Python 代码, 都会由 WebSim™ 通过如下方式进行预处理:

```
import scipy as sp          #导入 Python 数字模块支持库:scipy
from numpy import *        #导入 NumPy 的数组
import scipy.stats as ss   #导入统计相关函数
delay = 1                  #设置 Alpha 对应延迟天数
dr= WQSim_DataRegistry.Instance() #dr 初始化为 registry 类型
                               #使用 dr.GetData() 去访问市场数据
valid = dr.GetData_m_b("USA:TOP3000") # GetData_m_b() 访问指定的
空间样本, m 对应矩阵(matrix), b 对应布尔值(boolean), 这个结果会返回布尔
数值的矩阵, 从而说明金融资产在特定天里是否正常
```

4. 访问数据

WebSim™ 支持通过使用 Python 代码在后端访问数据, 需要通过 GetDataon 做数据的记录。例如:

```
shares_outstanding=dr.GetData("sharesout") # 获得流通股数据
ltDebt=dr.GetData("debt_lt") # 长期债务数据
closePrice=dr.GetData("close") # 收盘价数据
```

5. 生成函数、指针和Alpha的表达式

如前所述, 市场数据可以被认为是一个数值矩阵, 矩阵为每只股票的每

一个交易日提供可用的数据。时间对应时间指针变量 di ，股票对应金融工具指针变量 ii ，如表 31.2 所示。

表 31.2 将日期对应到日期索引 di ，股票对应到股票索引 ii

| 时 间 | di (时间索引) | 金融产品 | ii (股票索引) |
|----------|-------------|-----------|-------------|
| 20100101 | 0 | MSFT (微软) | 1 |
| 20100102 | 1 | AAPL (苹果) | 2 |
| 20100103 | 2 | PG (宝洁) | 3 |
| 20100104 | 3 | GOOG (谷歌) | 4 |
| 20100107 | 4 | AA (美铝公司) | 5 |
| 20100108 | 5 | K | 6 |
| ... | ... | ... | ... |

市场数据将以矩阵的形式排列，如收盘价，如表 31.3 所示。

表 31.3 以矩阵形式存在的市场数据

| 金融资产 时 间 | MSFT ($ii=0$) | HOG ($ii=1$) | AAPL ($ii=2$) | GOOG ($ii=3$) | PG ($ii=4$) | |
|--------------------|--------------------|-------------------|--------------------|--------------------|------------------|-----|
| 20100104($di=0$) | 30.95 | 25.46 | 214.01 | 626.75 | 61.12 | ... |
| 20100105($di=1$) | 30.96 | 25.65 | 214.38 | 623.99 | 61.14 | ... |
| 20100106($di=2$) | 30.77 | 25.59 | 210.97 | 608.26 | 60.85 | ... |
| 20100107($di=3$) | 30.452 | 25.8 | 210.58 | 594.1 | 60.52 | ... |
| 20100108($di=4$) | 30.66 | 25.53 | 211.98 | 602.02 | 60.44 | ... |
| ... | ... | ... | ... | ... | ... | ... |

现在看 2010 年 1 月 7 日, 苹果(AAPL)的收盘价, 我们需要使用 `close(3,2)` 得到数值。

使用 `Generate` 函数的唯一目的(在代码中可以实现)是将 Alpha 含有权重的向量的结果适用于每一只股票。`Generate` 函数每天都会执行 Alpha 表达式(好像是根据时间日期索引进行的循环), 参数为 di (是对应当前日期的时间索引)和 Alpha (需要被填充的结果向量)。例如:

```
closePrice = dr.GetData("close")
def Generate(di, alpha): # Alpha 表达式从这里开始
```

数据可以通过 `dataname[di-delay,ii]` 的方式进行访问。

`Dataname[时间索引, 金融工具索引]` 会返回特定金融工具在指定日期的对应数值。可以使用 `Alpha[金融工具索引]=表达式` 的做法, 将变量赋予 Alpha, 例如:

```
Alpha[ii] = dataname[di, ii]
```

用收盘价数据的例子可参考如下:

```
closePrice = dr.GetData("close")
def Generate(di, alpha):
    alpha[:] = 1./closePrice[di-delay, :] #本句等同于 '1/close'
    # “:” 指挥使用所有的金融工具, 从 0 到最后一个
```

由于我们使用了列表切片功能 (“:”), 它能在所有的金融工具上执行表达式 (即矩阵中所有的列)。

另一个使用收益数据定义 `alpha[:]` 的例子如下:

```
returnsMat = dr.GetData("returns")
def Generate(di, alpha):
    alpha[:] = -(returnsMat[di-delay, :]) # 收益总是趋向于平均值的回归
```

在一定时间窗口期内访问金融工具的数据, 例如 n 天, 可使用 `dataname[di-delay-n: di-delay, ii]`, 示例如下:

```
closePrice = dr.GetData("close")
def Generate(di, alpha):
    alpha[:] = mean(close[di-delay-10:di-delay, :], axis = 0)
    #取过去 10 天的平均值
    alpha[:] = where(valid[di, :], alpha[:], nan) # 正确性检查
```

需要注意的是, 用户选择时间窗口期应少于回测天数 (默认设置为 256 天)。这个值可以使用 Python 函数来检索: `Build.GetBackdays()`。

最后一行是用来检查数据的合规性的。这样确保不在样本空间内 (TOP 3000) 的数值不会被使用。请注意, 在 Python 程序开头的时候就有的 `valid` 变量定义, 随后也会对其进行解释。

6. Python的Alpha示例

下面是关于如何使用 Python 的几个常见示例。

(1) 同时使用多个数据

接下来的示例会展示如何同时访问和使用多个数据。能使用向量的地方一定要尽可能多地使用向量，因为循环的速度很慢。本例使用了 NumPy 内嵌的数学函数 (`numpy.subtract` 会自动被载入)，在本例中向量是通过“close-high”进行赋值的。

```
#针对不同的数据需要使用不同的变量
closePrice=dr.GetData("close")
highPrice=dr.GetData("high")
def Generate(di,alpha):
alpha[:] = closePrice[di-delay,:] - highPrice[di-delay,:]
```

(2) 使用自定义的函数

如下 Python 代码显示怎么去定义一个函数并且使用它。也会用到 NumPy 的函数 `numpy.where()`，这样会替代循环的方式进行向量的比较：

```
closePrice=dr.GetData("close")
lowPrice=dr.GetData("low")
def Generate(di,alpha): #可以调用 np_max (defined below)
alpha[:] = np_max(closePrice[di-delay:], lowPrice[di-delay-
1, :]) #等同于 'Max(close, Delay(low, 1))'
def np_max(data1, data2):
return where(data1 > data2, data1, data2) # NumPy 执行向量的比较
```

(3) 错误的 Alpha 赋值方法

```
closePrice=dr.GetData("close")
def Generate(di,alpha):
alpha[:] = closePrice[di-delay,:] # 和 close 函数一样
alpha = ones(alpha.shape[0]) # 用 NumPy 的 ones 函数给 Alpha 所
有向量赋值为 1
# 注意“alpha”后面并没有使用[:], 这样不会更改已经存在的 Alpha 所指向对
象的值, 而是给 Alpha 分配一个新的对象。这种赋值方法会导致对 Alpha 的向量值在
WebSim 中的丢失
alpha[:]=1./closePrice[di-delay,:] # 和 1/close 函数一样
# 这种方法没有问题, 因为在前面已经有了预先定义, 最终结果是“close”而不是
“1/close”
```

(4) 有效矩阵的应用

有效矩阵是有效金融工具的列表(比如, TOP 3000 例有 3000 种金融工具), 而且自动被创建。

```
closePrice=dr.GetData("close")
def Generate(di,alpha):
    alpha[:] = 1./closePrice[di-delay,: ]
    alpha[:] = onlyValid(alpha, di) # 使用之前定义好的函数过滤不合格金融工具

def onlyValid(x, di):
    myValid = valid[di-delay, :] # myValid 向量只有昨日的合规数据
    return where(myValid, x, nan) # 我们使用 NumPy 的 where() 函数来筛选出失效的金融工具。使用 numpy.nan() 来筛选金融工具。“0”是有效的数值
```

以上的有效性检查函数, 都可以插入到所有 Alpha 代码的结尾中, 如下:

```
alpha[:] = where(valid[di-delay,: ], alpha[: ], nan)
```

(5) 使用在 SciPy 中的统计函数

可以在 SciPy.org 网址查到 SciPy 中统计函数的列表。我们将在此处使用 SciPy rankdata() 函数。此函数给 Alpha 权重排序并且可以妥善处理这些关系。

```
high=dr.GetData("high")
def Generate(di,alpha):
    alpha[:] = ss.rankdata(high[di-delay, :]) #使用 SciPy 的 rank, 等同于 Rank(high)
    alpha[:] = where(valid[di-delay,: ], alpha[: ], nan)
```

上述 Alpha 表达式的排序(数值为从 0 到 1)调整至更高的段位。

下面示例演示了 SciPy 中 scipy.kurtosis() 函数对 11 天的收盘价数据处理的方法:

```
closePrice = dr.GetData("close")
def Generate(di,alpha):
    alpha[:] = ss.kurtosis(close[di-delay-11:di-delay,: ], axis = 0)
    #使用 SciPy 的 kurtosis 函数, 等同于 kurtosis(close,11)
    alpha[:] = where(valid[di-delay,: ], alpha[: ], nan)
```

下面的示例演示了 SciPy 中 scipy.skewness() 函数对 10 天的收益数据处理的方法:

```

returns = dr.GetData("returns")
def Generate(di, alpha):
    alpha[:] = ss.skew (returns[di-delay-10:di-delay,:], axis = 0)
                    #使用SciPy的skewness函数,等同于skewness(returns,10)
    alpha[:] = where(valid[di-delay,:],alpha[:],nan)

```

(6) 使用 NumPy 可用的统计函数

可以在 NumPy.org 网址查到 NumPy 中统计函数的列表。

下面的 Alpha 示例中展示了 NumPy 中均值和标准差函数 `numpy.std()`:

```

closePrice = dr.GetData("close")
def Generate(di, alpha):
    alpha[:] = mean(close[di-delay-5:di-delay,:], axis = 0)/std
(close[di-delay-5:di-delay,:], axis = 0) # 等同于 [Sum(close,5)/5]/
StdDev(close,5)
    alpha[:] = where(valid[di,:],alpha[:],nan)

```

下面的 Alpha 示例中展示了 NumPy 的最大值函数 `numpy.amax()` 的使用方法:

```

closePrice = dr.GetData("close")
def Generate(di, alpha):
    alpha[:] = amax(close[di-delay-20:di-delay,:], axis = 0)
                    # 找到过去 20 天最高收盘价
    alpha[:] = where(valid[di,:],alpha[:],nan)

```

(7) 访问和使用行业数据

```

closePrice = dr.GetData("close")
industry = dr.GetData("industry") # 访问 industry 行业数据
def Generate(di, ti, alpha):
    ind = unique(where(industry[di-delay, :] > 0, industry[di-
delay, :], 0))
    indclose = zeros(ind.shape[0]) # 初始化 NumPy 数组值为 0
    for i in xrange(ind.shape[0]):
        indclose[i] = mean(where((industry[di-delay, :] == ind[i]) *
(valid[di-delay, :]), closePrice[di-delay, :], 0))
        alpha[:] = where( (industry[di-delay, :] == ind[i]) *
(valid[di-delay,:]), indclose[i], alpha[:])

```

这里的行业数据是指所有有效行业的索引的数组。行业包括林业、金属矿产业、电力技术业、肉类加工业、纺织业、印刷业等,都有对应的索引。上面的例子演示了行业数据是怎么被访问和使用的。

表 31.4 展示了金融工具索引（对应 `industry[di-delay,ii]` 的数值）、行业索引和收盘价。

表 31.4 金融工具索引、行业索引（对应 `industry[di-delay,ii]` 的数值）和收盘价

| ii | industry[di-delay,ii] | close[di-delay,ii] |
|----|-----------------------|--------------------|
| 0 | 2 | 1.0 |
| 1 | 1 | -2.0 |
| 2 | 0 | 5.0 |
| 3 | 1 | -3.0 |
| 4 | 1 | -4.0 |
| 5 | 2 | 5.0 |
| 6 | 0 | 2.0 |

对于上述 Alpha 示例，`close` 数组中每个数值的计算方法如下：

因为 `indclose[i] = 所有这个行业金融工具的 closePrice[di-delay,ii] 的平均值`，

```
indclose[0] = (5.0+2.0)/2 = 3.5 (shown in table)
indclose[1] = (-2.0-3.0-4.0)/3 = -3
indclose[2] = (1.0+5.0)/2 = 3
```

然后通过 `indclose[industry[di-delay,ii]]` 对 `alpha[ii]` 进行赋值。

请注意，你也可以通过 `dr.GetData()` 来访问和使用各板块是行业的数据。

(8) 简单线性回归模型

在下面的代码中，5 天的收盘价和成交量加权平均值作为训练样本，都被用来计算回归权重 `w1` 和 `w2`：

```
Model formula: close[di] = w1 * close[di-1] + w2* vwap[di-1]
closePrice=dr.GetData("close")
vwapPrice=dr.GetData("vwap")
def Generate(di,alpha):
    CloseX=zeros((5,closePrice.shape[1]))          # 初始化 NumPy 数组
    CloseY=zeros((5,closePrice.shape[1]))
    VwapX=zeros((5,vwapPrice.shape[1]))
    for dnum in xrange(5):
        CloseX[dnum,:]=closePrice[di-delay-dnum-1,:]
        VwapX[dnum,:]=vwapPrice[di-delay-dnum-1,:]
        CloseY[dnum,:]= closePrice[di-delay-dnum,:]
    for ii in xrange(alpha.shape[0]):
```

```
SampleX=hstack((CloseX[:, [ii]],VwapX[:, [ii]]))
                # NumPy 函数 hstack() 将数组全部水平排列 (或竖直排列)
ParaX=LinearRegres (SampleX,CloseY[:, [ii]])
alpha[ii]=float (ParaX[0])*closePrice[di-delay,ii]+float (ParaX
[1])*vwapPrice[di-delay,ii]    #等同于表达式: w1 * close[di-1] + w2*
vwap[di-1]
alpha[:] = where(valid[di,:],alpha[:],nan)
def LinearRegres (xArr,yArr):
    xMat=matrix(xArr)
    yMat=matrix(yArr)
    xTx= xMat.T*xMat
    res=linalg.det(xTx)          # NumPy 的线性代数功能用来计算数组的行列
    if res==0.0 or isnan(res):  #如果 res 为 0 或者 nan
    return matrix([[nan],[nan]])
    else:
    return xTx.I*(xMat.T*yMat)  # matrix.I 实现矩阵的反转, matrix.T 实
现矩阵的转置
```

第 32 章

常见问题解答

作者：WebSim™团队

本章收集了一系列在 WebSim™ 用户社区提出的问题。

WebSim™

1. WebSim™ 是否适用于每一款浏览器？

答：官方支持的桌面浏览器都可以做到，如：IE9 及其以上版本浏览器；火狐 3.6 及其以上版本浏览器；谷歌浏览器。

Safari 5.1 及其以上版本的浏览器，和大多数版本的 Opera 浏览器，虽然没有官方支持，但也都可以运行得相当好。所有可以运行上述浏览器的桌面操作系统都可以运行 WebSim™。对手机系统 Android 和 iOS 系统还只是部分支持（这里不做保证）。

2. 可以同时提交多个 Alpha 吗？

答：不能在同一个会话里提交多个 Alpha。但如果有多多个会话，则每个会

话都可以单独运行一个 Alpha。可以通过在多个浏览器或计算机上登录来实现多会话。

3. 如何从表达式结果中得到每个股票的头寸？

答：图 32.1 显示了应用在 Alpha 上的操作/转换序列。

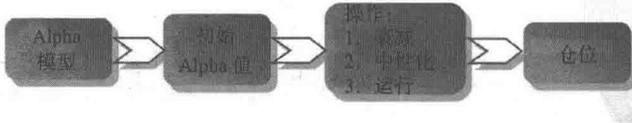


图 32.1 应用在 Alpha 上的操作/转换顺序

4. 发现这里只有有限的一些公司财务比率，比如税息折旧及摊销前利润 (EBITDA)、净资产收益率 (ROE) 以及资产收益率 (ROA) 等。是否可以提供其他的财务比率？还是需要我们自己来计算？

答：所有当前可用的数据序列都罗列在 FAQ 页上。

5. 我不小心双击了提交按钮，然后弹出来一个错误提示：“你尚有一个提交未完成，该提交 599 秒后完成，请等待。”我想知道，是否可以中止运行中的模拟？

答：目前还不能中止运行中的模拟。但是你可以先退出会话，然后重新登录。这样做可以结束你正在进行的模拟，你还可以使用“Escape”键。我们以后会为我们的模拟窗口添加中止按钮。

6. 当我在表达框里输入时，我会看到一些建议。它们是来自我自己的输入历史还是来自所有用户的输入历史？

答：这些建议来自用户自己的输入历史。

7. 可以给自己的 Alpha 拟定特定的名称吗？

答：是的，你可以自定义你的 Alpha 信号名称，只需要去“My Alphas”页面上单击“Alpha Id”按钮，重命名就可以了。

8. 你能在我的 Alpha 表里将“逻辑上相似”的 Alpha 组合在一起吗？目

前，可以按创建日期和其他字段进行排序，但我想表明的是，给基于同一基本想法的不同版本的表达式分类，并将它们组合在一起使它们看起来更加清晰。

答：不能，Alpha 逻辑上的相似性是不够明显的。的确，使用其他分类方式来给 Alpha 分组要比根据日期分组有用得多，这也值得我们付出更多精力去研究。

9. 为什么我会在“启动……”阶段被卡在模拟进度的页面上？为什么 Python 代码在代码输入框中没有高亮显示语法？

答：这两个问题都可以通过在浏览器设置中启用 JavaScript 来修复。

10. 我在注册 WebSim™ 账号时遇到问题，应该和谁联系？

答：在大多数情况下，都可以直接查询用户报告常见问题列表，上面有产生的可能原因和修复方法，如表 32.1 所示。

表 32.1 用户提到的常见问题、可能原因和解决方法

| 常见问题 | 可能原因 | 解决方法 |
|--------------------------|------------------------------------|---|
| 在注册时，单击“我同意”按钮，但是返回到欢迎界面 | cookies 被禁用 | 在浏览器设置中允许 cookies |
| 用户没有收到账户确认邮件 | 邮件被邮件服务器认定为垃圾邮件或者被挪到了不常用的收件箱的子文件夹里 | (1) 检查垃圾邮箱和收件箱的子文件夹； (2) 如果还不行，就尝试用比较流行的基于 Web 的邮件服务商，比如 Gmail 和 Yahoo |
| 在代码输入界面没有语法高亮 | JavaScript 被禁用 | 在浏览器设置中允许 JavaScript |

注：上述第三方链接仅为方便用户。用户应仔细阅读并严格遵守这些网站上的使用条款。

11. 我的配置文件在哪里？如何更改我的密码？

答：你可以在 My Account 上查看你的个人资料；在 Change Password 上更改你的密码。

12. 我想知道，Alpha 表达式中收益变量的内涵是什么？我是指，return 是指哪只股票的收益率？顺便问一下，有没有办法改变默认的资产分类（如我想用我的 Alpha 来交易像“谷歌”和“苹果”那样的高科技公司股票，而不

去交易像“沃尔玛”这样的零售公司股票)？

答：WebSim™ 不计算特定股票或者用户自定义的金融工具结合（股票或者合约）的收益率。相反，它计算的是一个流动金融工具整个样本空间的收益率。WebSim™ 会给你机会选择不同的样本空间。例如，如果你选择 USTOP 3000 股票，WebSim™ 将根据表达式计算美国地区流动性排在前 3000 名的股票。遗憾的是，你不能工作于样本空间的一个子集或指定行业/子行业。在这种情况下，你能做的就是使用设置版面来改变样本空间设置。

WebSim™ 使用这种方式来确保你建立一个稳健的 Alpha 模型，而不是偏向于某一些股票。这种方法确保你的 Alpha 是市场/行业/子行业中性的。

13. 信号作者是否可以在“My Alphas”页面为每次模拟上添加标签/注释（或许是名称模拟而不是数字）？这对搜索 Alpha 将会非常有帮助。例如，你可以设置一个标签名“#factor”。这样就可以简单地尝试用这些标签名来复制它们，而不用每次都打字。如果这个设置已经可用，请告知我。

答：非常感谢你的好建议。目前，如果你单击它，每个 Alpha 的名称区域都可以进行编辑。请尝试重命名你的 Alpha。名称参数可以使用过滤器，所以你可以很容易地搜索到 Alpha。

Alpha 的表达式、市场数据及其函数

1. 什么是 $-1 * \text{Alpha}$ ？它意味着什么？

答：最后的 Alpha 值（中性化等处理之后）表明应该对不同的股票做多（正值）或者做空（负值）的资产池的比例值。因此要弄清楚各种操作将如何影响 Alpha 的权重很简单。 $-1 * \text{Alpha}$ 和 Alpha 会使用同样数量的资金在一只股票上分别卖空和买多。

2. 是否有实现移动平均计算的现有函数？

答：Sum、Sum_i 和 Call_i 等都可以用来构建各种移动平均线。请使用 decay_exp(x,f,n) 函数来计算指数移动平均线。

3. NaN 是什么意思？它等于零吗？

答：NaN 代表的不是一个数字。它用来表示“无效”操作的结果，类似

于零除法，或表示某些数据已损坏或不可用。如果某只股票的 $\text{Alpha} = \text{NaN}$ ，则意味着没有仓位。但是，如果某只股票的 $\text{Alpha} = 0$ ，那么经过衰减和中性化等操作后， Alpha 可能有一个非零值，就意味着有仓位。在不同情况下，如果无法对一只金融工具赋 Alpha 值，则将 Alpha 设置为 NaN 。例如，如果有错误或丢失的数据，则将值设置为 NaN （无效）。

4. 数据多久更新一次？

答：数据每周都会更新。更新之后，将可以使用最新的可用数据。

5. 如何定义 Python 代码中的全局变量？我想使用全局变量，这样可以保存一个值并每天进行变更。

答：下面给出了如何定义和使用全局变量的示例代码：

```
closePrice = dr.GetData("close")
Var = 1 # 全局变量初始化
def Generate(di, alpha):
    global Var # 从函数内部进行访问
    wdi1 = (di-delay)%closePrice.shape[0]
    wdi2 = (di-delay-3)%closePrice.shape[0]
    if Var > 0:
        Var += 1 # 开始使用
    # some more operations
    # alpha expression
```

6. 在帮助部分文档里提到的基本面数据是不可用的。所以，我无法创造一个新的 Alpha 。我该怎么办？

答：很抱歉给你带来不便。我们将尽快修复基本面数据的错误。你可以先用价格和交易量来创造新的 Alpha 。使用这些数据也可以创建一些好的 Alpha 。

感谢你的耐心！

7. 每当我尝试使用 `sp.linalg` 方法时，它都会跳出一个错误提示，“错误：没有收益结果”；调试器上查找不到详情，我已经在我的个人电脑上测试了这段代码，看起来没什么毛病。假如我在代码里面注释掉该行，会显示：`import scipy.linalg as LA`（用户定义的导入语句）和 `evals, evects=LA.eig(matrix(c))`，

代码则会运行得很好。你能检查一下 `sp.linalg` 是否正常工作吗？

答：请参阅答案如下。

a. 到目前为止，`WebSim™` 还不能导入模块。但是，你可以使用 `NumPy` 和 `SciPy` 数据库，因为它们是由 `WebSim™` 动导入的。

所以你需要在代码中删除此句：

```
import scipy.linalg as LA
```

也需要改变 `eig()` 函数的调用为：

```
evals, vecs = sp.linalg.eig(matrix(c))
```

b. `eig()` 函数可以用在 `WebSim™` 当中，但该函数要求矩阵数据里不包括 `NAN` 和 `INF` 字符。由于存在某个公司在某一天里数据不可用的情况（例如，该公司被摘牌），所以在 `WebSim™` 中获得的数据有可能会含有 `NAN`。你需要检查这些 `NAN`，并寻找更合适的数值来代替它们，或者将包含 `NAN` 的股票一起删除。例如，如果你想用 `0` 来更换所有 `NAN`，则可以使用如下语句：

```
c[:] = where(c != c, 0, c)
```

c. 虽然上述两个方法可在技术上使你的代码准确无误，但是它可能需要很长的时间来模拟（甚至是超时）。`eig()` 函数本身就是一个很耗时的操作，实际股票的数目上升到约 7000 只时，它就会试图计算出一个大小约 7000×7000 元素的矩阵特征值。我们正在努力加快 `Python` 模拟的速度。同时，你可以试着修改你的代码来计算一个较小矩阵的特征值，或少调用 `eig()` 函数。

我们知道，要想清除这些错误，就需要在模拟失败时将相关错误信息显示给用户。我们的团队仍在为这个问题做修复。

8. 在这里我只能找到一些数学函数和汇总的 API。有没有更多的资源，可以帮助我们了解更多关于 `WebSim™` 系统与策略的内容？

答：`WebSim™` 中所有函数和可供使用的数据，我们都会在 `FAQ` 部分列出来。请参考帮助部分的案例，来了解如何使用它们。我们正在努力发布更多的函数、数据集和资产类别，也将发布包含 `Alpha` 研究、更多的例子以及插图的 `WebSim™` 官方文件。

在此之前，如果你在完善 WebSim™ 方面有任何建议（如要添加特定函数/新特性等），或在使用 WebSim™ 时碰到任何技术问题，可以随时向我们提出来。

9. 我尝试使用 Python 来构建我的 Alpha。我想要在构造第 n 天的 Alpha 向量时，可以参考第 $n-1$ 天的 Alpha 值。有简单的方法可以做到吗？

答：为了存储 Alpha 历史数据，你需要在 Generate 函数之外，设置一个全局向量。我附上了一个示例代码，Alpha 计算如下：

今日 Alpha 数值 = (当日收盘价的倒数 + 前一日的 Alpha 数值) / 2

请参考如下示例代码：

```
closePrice = dr.Getdata("close")
hist = [] # 存储 Alpha 历史值的向量
start = True # 布尔值，定义开始日期
def Generate(di,alpha): # 如果需要在函数内修改，则全局变量需要定义
    global start
    global hist # 以上三行是为了初始化“hist”向量
    if start:
        hist = zeros(alpha.shape[0])
        start = False # 计算当天的数值
    wdi = (di-delay)%closePrice.shape[0]
    alpha[:] = 1./closePrice[wdi, :] # 用当天的平均值和前日数据
去计算最终 Alpha 数值
    alpha[:] = 0.5 * (alpha[:] + hist[:]) # 将当天的数值保存在历史向
量里
    hist[:] = alpha[:]
```

操作

1. 三种市场中性化方法的区别是什么？如何判断哪一种方法更好？

答：市场中性化的方法决定哪一组用于 Alpha 的中性化，选择所有同一子行业或行业的股票，或更简单地使用整个样本空间为一组都是基于此目的。至于第二个问题，正确的答案取决于 Alpha 的逻辑/公式，所以这里没有具体的答案。结果应指出怎么去操作。

2. 如何添加自定义的 Alpha 操作符，并在 Alpha 表达式中运行它？

答：在表达式输入模式下，这个功能是禁止的。但是你可以用 Python 来实现它。

3. 当输入 Alpha 表达式“1”时，在所有样本空间都没有返回值。那么要如何消除 NaN 呢？

答：如果你在设置下拉菜单中查找（在顶部导航栏右侧的齿轮形状图标上悬停可以显示），就会注意到，默认中性化设置是“子行业”（subindustry）。这就是当你使用“1”作为你的表达式时，你会看到 NaNs 的原因（中性化设置导致了所有股票都没有头寸）。如果将它设置为“None”，那么你会看到预期的结果。

4. CAPM 模型 ($R[i] = B(R[m]) + \text{Alpha}$) 中的 Alpha 与我们尝试计算的 Alpha 相同吗？我们忽略了 Beta，只专注于 Alpha 是因为多空市场的中性策略等相关原因，还是因为其他原因？

答：本公司所给的 Alpha 定义与你研究过的 CAPM 中的 Alpha 完全不一样。

在 WebSim™ 中，Alpha 指的是一个数学模型或策略，形式为 Python 代码或者表达式，它对不同的金融工具（股票）赋予不同的赌注（即权重），并期望从中长期获利。

简单地说，它可以创建一个权重向量，在选定的样本空间中每个权重对应一只股票。根据中性化设置（选择市场、行业、子行业或不选）的不同，这些权重可能是市场中性化的，也可能不是。在模拟期间的每一天里都会创建一个证券组合，然后根据这个证券组合来计算当天的损益。

Alpha 只是一个象征性的名字，不应该与常见的 Alpha 定义相混淆，其常用定义为它是一种对预期收益之外的超额收益或异常收益的度量策略（如 CAPM）。

5. 我曾经用 $\text{Alpha} = \text{close} * \text{sharesout}$ 公式来测试 WebSim™，用的是样本空间的默认值。令我惊讶的是，它给出的结果很不理想，甚至还有一个负的

夏普比率。

这与 Russell 3000 指数过去 5 年里的表现不太一致，如果可以的话请查看是否有异常。

答：默认设置是市场中性化。在这种情况下，为了达到市场中性化，所有股票都被赋予了多或空的头寸。所以请将设置修改为“None”。请注意，因为收益有大幅下滑，所以由此产生的夏普比率和结果都不会太理想。

Alpha 的性能

1. 我如何才能获得一个好的表达式呢？每个比率（如换手率或最大回撤率）的阈值是多少？

答：可进入操作系统进行测试的 Alpha 的标准是，夏普比率须超过一定的阈值。在延迟为 1 和 0 的情况下，Alpha 夏普比率的阈值分别为 2.5 和 3.95。这里也有定性的词来反映性能的状况，如“极好”“好”“差”等，如收益表所示，这些也可根据夏普比率来确定（参见帮助菜单中的常见问题页面）。

2. 模拟评分列为 6 个等级，分别是：超好、极好、好、一般、差、极差。那么 Alpha 必须满足的标准是什么？

答：针对夏普比率有一个可配置的内部阈值。但是，上述 6 个等级是另一个独立的测量标准。

3. 用户可以检查他自己的 Alpha 信号之间的相关性吗？

答：相关性只能通过样本外测来计算。需要几天的时间才会在 My Alphas 页面上显示出结果。结果会以在特定相关性区间内的 Alpha 个数的形式显示在“OS”标签栏。

4. 当用户提交一个新尝试时，收益结果会被存储在什么地方？相关性只有在它经过检测后才可获得，这一点不是很方便。因为当我们改进一个想法时，我们只想尽快检查它与原想法的相关性。

答：目前没有存储日收益结果，仅保存以统计数据形式出现的结果汇总。更多的像你提到的建议一样的、对开发 Alpha 有益的特性都会被添加进来。

5. Alpha 输出向量产生在哪里？

答：WebSim™ 不在模拟结果页中显示 Alpha 向量的内容。

6. 样本外测试的 Alpha 是如何提交的？

答：目前，如果该 Alpha 通过了前面定义的夏普比率阈值，则 Alpha 会自动提交到样本外测。

7. 被选定进行样本外测的 Alpha，要多久才能看到统计数据？

答：估计一周左右可以看到统计数据，因为它是在周末完成的。

8. 样本空间是如何影响 Alpha 的 IR 的？

答：大样本空间的信息比率（即 IR）要比小样本空间的好。但这也并非不总是正确的。有一些 Alpha 在大型股票上表现得更好，原因如下：

a. 大股票的数据是比较干净的，即数据平稳。

b. 大股票的行为与小股票的不同。

虽然大样本空间里的 Alpha 通常比小样本空间里的更容易开发，但仍建议你开发的 Alpha 能在小样本空间里工作。因为小样本空间具有更好的流动性。

9. 好的 Alpha 的特点是什么？如何才能最优化我的结果？

答：思想的独创性是必不可少的。Alpha 应该是稳健的。在样本外测中，延迟为 0 的 Alpha 至少有 3.95 的夏普比率，而 delay 为 1 的 Alpha，夏普比率至少要到 2.5。同时要确保收益率高，换手率小于 40%，并保证最大回撤能维持在最低限度。这不只关乎收益，也包括它带来的影响。在流动的样本空间里，稳健的性能是很必要的。样本外测时的夏普比率像样本内测时性能的降低一样重要。

性能良好的 Alpha 对我们而言是有价值的。如果你在样本外测区域看到一个 Alpha，那么你可以确信它很有价值。

利用套利观点思考问题，可尝试不同的操作（如衰减、中性化等）。可尝试同时更换所有的公式，而不是只改变一个。不要试图将不符合逻辑的表达式拼凑在一起，例如 $adv20 + close$ ，这本质上是“交易量+价格”的组合，没

有任何实际意义。通过打开单元检查设置，这个问题可以在 WebSim™ 内部检查出来。

不要过度拟合参数，那会伤害性能。

注意：过度拟合在这里是指，为获得一个好的样本内测夏普比率，以一种不符合逻辑的方式对 Alpha 的表达式进行轻微改动。如，稍微改变表达式中的常量，将 power 的参数从 2 改变至 2.5，或改变某些板块的静态翻转标志等。以上做法都不可行，因为它将不可避免地导致样本外测试的失败。

注意：所有查询、建议、反馈，请通过 WebSim™ 的联系页面发送电子邮件。对于任何技术问题，用户必须标明他们使用的浏览器及其版本，同时也要告知 WebSim™ 团队最后遇到问题的确切时间，以便团队能够找出相关的日志信息。

第五部分

结 语

第 33 章 卓越宽客的七个习惯

第 33 章

卓越宽客的七个习惯

作者：Richard Hu

宽客通常被认为是“华尔街上的火箭科学家”，这句话不禁让人联想到一个有头脑、受过高等教育并且高薪——可能年入数百万美元的人物形象。在典型的买方量化投资公司中，工作环境是大学式而专业化的，在这里你会有很多机会收获新的知识。因此，许多世界名校的工程及科学专业的毕业生想要成为宽客也就不难理解了。

在过去的 10 年里，我已经管理了成百上千的宽客，也曾在逾 10 个国家的众多招聘活动中与成千上万个有抱负的宽客有过交谈。在交谈过程中，一个被反复提及的问题是：如何才能成为一名成功的宽客？

为了回答这个问题，我们调查、采访了多位顶级宽客。我们发现以下七个习惯是他们取得成功的重要因素。

1. 愿意付出额外的努力

极其成功的宽客们一般都愿意投入必要的额外努力来到达成功的彼岸。如果失败是成功之母，那么额外的努力就是成功之父。

我曾经聘请了一位极聪明的、名校毕业的年轻人。在面试时，他解答我们分析方面的困难考题时的速度相当快，让我们不禁怀疑他之前是否曾见过这些考题。

因为他在面试时表现很好，我们决定花些时间重新给他出一道更加棘手的问题，然而他又以迅雷不及掩耳之势地解决了。我们都特别高兴能找到这样的人才。

然而他进入公司之后，我却发现他有一个严重的毛病——工作态度很不端正。他总是在晚上聚会，并且在外面待到很晚，所以他经常起得很迟，然后在上午 10:00 左右奔进办公室，而其他员工早在 8:00 就开始工作了。

当我与他谈及此事之后，他会先守规矩一段时间。但过不了多久，又会打回到原形，像龟兔赛跑里睡着的兔子一样落后于其他人。他从来没有成为过最出色的那个人。

另外，我也雇用了一些聪明但不一定是最聪明的人。我发现他们是员工中最早来到办公室的，也是最晚离开的——日复一日，月复一月，年复一年。

随着时间的推移，这样的人慢慢地稳步领先于其他人，不断超越别人，最终成为员工中表现最好的。当然，在这些人当中，那些既聪明又十分努力的人自然成功得也很快。倘若要我在长久地加倍努力和高智商之间做选择，那么我一定会选择前者。

2. 总是做出合理的改变，而不只是单纯让模型匹配数据

Alpha 研究是提出假设并做相关实验的实验过程。但很多假设和实验都不会有很好的结果。

当结果不够好时，最简单的事情是调整模型参数，直到它通过某一最小阈值。

这种类型的 Alpha 在真实的市场里运行得并不是很好。成功的宽客们会试图找出导致结果不好的原因，并尽力做出合理的改进。这样一来，Alpha 在真实的市场里才更有可能表现出色。

3. 渴望用新想法做实验

成功的宽客都热衷于探索新的想法和数据。有时，为了探索新领域，需要做更多的准备或基础工作，而成功的宽客们都愿意去做这部分工作。如果一个人等到所有的路障都被移除后再去探索，那么大部分的宝藏早就被别人给挖走了。

4. 做增值工作

在创造 Alpha 时，基于一个现有 Alpha 做轻微改变来构建 Alpha 很简单，而基于一个新想法来构建 Alpha 却很难。但是，对于投资管理来说后者更有价值。成功的宽客们会投入更多的精力去建立基于新想法的 Alpha。

5. 有强烈的紧迫感

最成功的宽客有着强烈的紧迫感去把事情做完。当有一个想法时，他们会迫不及待地去尝试。他们不断思考着 Alpha 的研究并快速跟进，而且做事情非常主动。

6. 形成协同团队

成功的宽客都会尽量与他们喜欢并且信任的其他优秀宽客组成团队。以团队形式讨论 Alpha 想法，分享研究工具，互相帮助以渡过暂时的难关，打破单调孤立的研究模式。通过这种形式的团队合作，他们会变得更有效率，从而获得更多的成果，并为自己营造一种更愉快的工作环境。

7. 设定高目标

对于最后的习惯，我打算引用一句古语：“倘若你瞄准月亮，即使迷失也会落于星辰之间。”

这些就是极其成功的宽客的七个习惯。倘若你能很好地执行它们，你一定会找到最适合你的卓越宽客的荣耀人生。

Finding Alphas

如何透过市场走势发掘别人无法量化的潜在获利机会？

本书从抽象理论到具体实施，阐述了Alpha设计中的基本技术和独特技巧，揭秘了发现数据中的隐藏信号的神奇方法。本书还配备了强大的培训工具：

- ★卓越宽客的七个习惯
- ★Alpha设计中所有关键技术点的处理方法
- ★在进行Alpha信号挖掘和证券交易时，行之有效的方法和诀窍
- ★可以创建和修正Alpha信号的线上模拟器WebSim

通过大量图、表清晰地解释了：

- ★开发、评估和改进Alpha质量的关键步骤
- ★换手率、回测、基本面分析、股票价量、统计套利、过度拟合和Alpha多样性的关键技术细节
- ★在Alpha多样性中，哪些该做，哪些不该做
- ★各种资产类别的Alpha设计，以及新闻和社交媒体对股票收益率的影响

• 关于作者

Igor Tulchinsky是世坤公司（WorldQuant）的创始人兼CEO，该公司是一个跨国私人投资管理机构。他还是世坤基金的创始人，该基金为那些在科学和量化研究领域深造的杰出学子提供奖学金。他还创建了WorldQuant大学，免费提供计量金融硕士的在线学位教育。

上架建议：量化投资

ISBN 978-7-121-33136-7



9 787121 331367 >

定价：69.00元



WILEY



责任编辑：高洪霞

封面设计：李玲