



通联数据
DATAYES



量化投资

交易模型开发与数据挖掘

韩焘 / 著

公开展示高成功率量化交易模型的开发流程，挖掘市场中的投资机会

Quantitative Investment

 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>

量化投资

交易模型开发与数据挖掘

一本好书可以改变人的一生，在量化时代，工具将成为重要的“帮手”，《量化投资：交易模型开发与数据挖掘》一书就是这样的工具。

张伟 “总裁Talk”创始人 知名财经主持人

本书是韩焘以他对金融的深刻理解和对量化交易理论的严谨论证，加上10余年市场上的实盘交易经验融合而成的，实属诚意之作，对广大投资者具有一定的借鉴意义，值得一读。

马华轩 国元期货总经理

韩焘的这本量化投资实战参考书，既具有理论与实践相结合的特点，又增加了当前热门的人工智能在量化投资中的应用内容，与时俱进，值得力荐！

丁鹏 中国量化投资学会理事长

理性逻辑被情绪影响，难逃贪婪与恐惧的决策结果。投资者要想投资成功，量化投资是一个重要途径。本书全面介绍量化投资实务，与当下技术热点结合，对不同阶段的量化投资者都会有所帮助。

秦红 中科汇泽资产管理有限公司总裁

很多投资者都想赚“快钱”，期望快速致富，所以把时间和精力都放在了研究K线理论上，每天还特别认真。但实际上，投资者每天研究K线波动是没有意义的，因为你再聪明、手再快，也“干”不过量化的电脑程序化交易。所以投资者想赚钱只有两条路可以选择：一条是认清投资的本质，回归企业本身的价值；另一条是学习量化交易，摆脱每天看盘的痛苦。否则，大多数人不过是在玩追涨杀跌的“博彩”游戏，成本极其高昂。

吕长顺（财经大V凯恩斯） 北京中和应泰财务顾问有限公司首席研究员

对于每个行业的从业人员来说，都需要多关注当前的科技发展，以提高现有的劳动生产率，投资行业更是如此。无论是机构还是散户，量化交易无疑是未来投资中的必修课。

宋振宇 《证券市场周刊·红周刊》研发部主编 北京电视台《投资者说》栏目长期嘉宾

用数学的方法从混沌中提炼确定性，是量化投资的根本，也是本书所阐述的核心内容，值得一读。

姚念康 IBM大中华区咨询整合部原总经理 金评媒首席科学家



读者服务

微信扫码回复：37586

- 获取博文视点学院20元优惠券
- 获取本书配套代码资源
- 获取精选书单推荐

上架建议：量化投资

ISBN 978-7-121-37586-6



定价：99.00元



责任编辑：刘 伟
封面设计：李 玲



量化投资

交易模型开发与数据挖掘

韩焘 / 著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书是一本针对想在股票、期货和期权等投资市场上获取更多收益的初中级投资者的技术参考书。

本书第1~4章主要讲解量化投资的入门知识,包括量化投资的发展现状、量化投资的开发工具、策略回测、择时与选股策略等内容;第5~7章主要讲解量化对冲策略与数据挖掘,包括数据加载与收益分析、量化投资中数据挖掘的使用等内容;第8~9章主要讲解量化投资中的配置方法,包括资产配置和风险控制,以及量化投资中的仓位决策方法与技巧等内容;第10~11章主要讲解人工智能技术在量化投资中的运用,包括机器学习与遗传算法、人工智能选股模型的使用等内容。

全书内容专业,案例丰富翔实,是作者近10年不断在量化投资与人工智能技术领域探索的最佳结晶。

本书不仅适合初入门的投资者,也适合有一定投资经验且想深入掌握量化操作的投资者使用,还可以作为私募投资机构 and 券商培训机构的参考教材。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

量化投资:交易模型开发与数据挖掘 / 韩焘著.—北京:电子工业出版社,2020.1

(量化交易丛书)

ISBN 978-7-121-37586-6

I. ①量... II. ①韩... III. ①机器学习②数据处理 IV. ①TP181②TP274

中国版本图书馆CIP数据核字(2019)第219785号

责任编辑:刘伟 特约编辑:田学清

印 刷:三河市双峰印刷装订有限公司

装 订:三河市双峰印刷装订有限公司

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:787×980 1/16

印张:30.00 字数:600千字

版 次:2020年1月第1版

印 次:2020年1月第1次印刷

定 价:99.00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到 dbqq@phei.com.cn。

本书咨询联系方式:010-51260888-819, faq@phei.com.cn。

推荐序一

散户亏钱的原因不是不懂基本面，也不是不懂技术面，而是克服不了人性的弱点，而量化交易以先进的数学计算模型代替个人的主观判断，能够避免贪婪、恐惧、侥幸等让投资者亏钱的人性弱点，其在国内外市场上受到了很多人的追捧，这一定也是未来研究股票交易策略的发展方向。

我个人很早就听说过量化交易，但一直没怎么在意，直到2016年我们营业部发行了一款基于量化交易的私募基金，到2018年结束的时候，两年时间该基金盈利超过20%。而每个投资者都知道，2016年—2018年的市场行情是什么样的，很多中小板股票跌幅达60%~70%，从那以后，我对量化交易产生了很大的兴趣。

而对量化交易有更深的认识是来自网上的一段话，大意是“在AlphaGo战胜李世石的那个夜晚，疲惫的李世石早早睡下，而AlphaGo又和自己下了100万盘围棋。第二天当太阳照常升起的时候，李世石还是李世石，而AlphaGo已经变成完全不同的存在，从此以后人类可能再无获胜的机会。人工智能不再是科幻电影里的画面，不再是新闻标题，它正在以一个我们不可想象的速度改变我们的生活”。

受以上两件事的影响，我们营业部确定了大力发展量化交易类私募基金产品的方向，随后调研了许多人工智能和量化交易类的私募基金，并查看了很多相关资料，最后得出一个结论：凡是在总结经验有用的领域，人类可能将永远失去机会。简单来说，对同一种病症，即便是三甲医院的医生，一辈子可能也就能看上万张X光片（我觉得这还是往多了说的），而使用人工智能，可能一晚上就能看上千万张，更重要的是，它还不会因为生活中的琐事而影响自己的判断。

股票交易恰好处在一个总结经验有用的领域中，我认识一个私募界前辈，他从5万元



起家，通过权证、股指期货、股票交易等赢得了几亿元的身家。他给我分享成功经历时，有一段话使我印象非常深刻，“1997年刚开始炒股时，买不起电脑，每周六到报刊亭买《中国证券报》，上面有当时几百只股票的日K线图，周末时就反复看这些图，预测下周可能的走势，等下周再买新的报纸，把真实的走势和当时自己的预测进行比较，看了半年后，基本上就没有再亏过钱，后来赚的几亿元也是当时打下的坚实基础”。因此，我相信人工智能和量化交易将在股票市场中大放异彩。

市场上有很多关于量化交易的参考书，但大家仔细去看就会发现，作者大部分来自计算机行业或高校，这些书理论上没有问题，对想要学习量化交易的投资者也会起到一定的作用，但我觉得总是缺了一些什么——这些人或者不懂交易原理，或者不懂交易心理。

而本书作者，我的好朋友韩焘先生，是一个从散户成长起来的私募基金经理，作为散户他知道普通投资者容易犯哪些错误，他成功就是因为他反思并修正了这些错误；作为基金经理，他懂得机构是如何进行投资决策的。在本书中，他结合自己的投资经验，给出了很多交易策略，可以说其中有一些就是他投资盈利的“真家伙”，很多人可能不相信有人 would 会把自己赖以生存的东西无私地拿出来分享。这里，我想和大家说的是：真正的投资者无论是做价值投资，还是做技术投资，都是非常纯粹并乐于分享的人，如股神巴菲特、传奇基金经理彼得·林奇等，都用其一生的时间给所有股票投资者树立了一个良好的榜样——与投资者分享他们的方法，甚至他们买的股票品种等。可是真正能从中获益的人并不多。究其原因，就是很多人为了方法而方法，没有真正掌握他们所说的方法中的使用技巧，本书将这些技巧进行了详细说明，期望对读者有所帮助。

你相信什么，就会看到什么，最终也会得到什么。我相信，在不远的将来，未来股票市场中量化交易的交易量将达到50%以上，到时候市场上分为两种人：懂量化交易的和不懂量化交易的。如果你是那个不懂量化交易的人，就如同拿着木棒的原始人与一批武装到牙齿的数字化士兵在战斗，你凭什么获胜？

因此，无论你是散户，还是机构投资者，或者和我一样是证券公司的从业人员，只要有志于学习和了解量化交易，这本书你一定要读，因为这是市面上十分接地气的一本量化交易专业类书籍。

张云龙

东北证券北京三里河东路证券营业部
财富总监

推荐序二

量化投资作为一门投资的方法论及应用技术，在国外成熟市场已经有几十年的成功应用与靓丽业绩。应韩焘先生之邀为本书作序，感到很荣幸。国内资本市场形成之初的十几年，由于交易规则、技术条件、投资品种、资产管理等方面均处于起步阶段，量化投资的应用相当有限。随着国内市场的逐步开放与技术进步，尤其是大量诸如金融期货、期权等新的投资品种上市交易，伴随着交易技术的自动化执行，量化投资在策略开发端与交易执行端等投资核心环节，日益受到重视，大量以量化投资为核心业务的资产管理公司也如雨后春笋般涌现出来。

韩焘先生在本书中一一列举了量化投资对于传统投资的优势与改进策略，包括但不限于量化投资在数据样本、精确度、准确度等方面质的提升，以及可供交易的策略类型、数量的大幅度增加。同时，本书也对量化投资的产生、发展与兴盛，在时间维度上进行了阐述，对量化投资在国内外实践应用过程中的一些有代表性的具体案例进行了深入说明。

量化投资是一个系统性的思维、设计、研发与决策实施的过程，包括策略、交易、风险控制等诸多核心内容。本书作者以当下流行的 Python 语言作为量化投资的程序设计语言，结合通联数据公司提供的优矿量化投资平台，对于多种不同类型与目标的量化投资策略进行了详细解说，其中有传统的因子、择时策略，也有基于机器学习的智能化策略，还有风险与资金管理技术。

处于量化投资起步阶段的国内资本市场，投资者面临更多的是机遇，一本好书能为读



者带来正确的投资观与执行方法。韩焘先生笔耕不辍而成此书，也希望本书能指导国内有志于研究量化投资技术、从事量化投资行业的读者进入一片新的领域。

童少鹏

北京市金融发展促进中心
首都经济贸易大学
量化金融研究中心研究员

推荐序三

人工智能作为当前信息社会中的热词，其深度学习、机器学习和神经网络等技术在各个领域都得到了广泛的应用，尤其是在金融领域，应用更为深入。

作为当今的投资者，在常规理财的基础上，最好了解一些多品种、多策略的投资理财方式。例如，基金、股票、外汇等多资产配置手段。小到一件商品在不同商店的差价，大到全球经济一体化背景下的跨国贸易，以及在资本市场进行股票、期货和大宗商品等套利交易，也都是投资的一种手段。投资者要想在投资市场博取利润，必须学会降低投资风险，提升投资收益率。而要想有效地捕捉这种非对称信息下的投资机会，尽可能地降低投资风险，就对数据的统计分析有很大的挑战，由此也使得人工智能显得日益重要。

当下，散户机构化是资本市场的一个发展趋势，机构投资者凭借雄厚的资本实力和丰富的投资经验，借助计算机的运算能力，在人工智能化的趋势下逐渐催生出纷繁复杂的量化投资模型。在这个过程中，无论是公募、私募等机构投资者，还是很多个人投资者，量化投资作为一种专业化的投资方式已经不再陌生，并在资本市场的推动下一步步地扩散并深入人心。量化投资领域中应用较为广泛的人工智能技术的不断发展，进一步推动了量化投资策略的逐步完善，很多机构也从简单的技术选股到多因子选股，再到通过计算机的大数据获取与挖掘，逐步形成了独特而有效的量化投资模型。

量化投资模型一般具备如下特点：

- 能使用多层次的量化模型观察海量数据，进而捕捉投资机会。
- 能够依靠概率取胜，如定量投资从历史数据中挖掘有望在未来重复的历史规律，以及大概率获胜的投资策略。量化投资模型是依靠筛选出的股票组合取胜的，而不是



依靠一只或几只股票取胜，从投资组合的理念来看也是捕获大概率获胜的股票。

- 能严格地执行量化投资模型所给出的投资建议，克服了人性的弱点。
- 能准确客观地评价交易机会，克服主观情绪的一些偏差，通过全面、系统性扫描捕捉错误定价和错误估值带来的机会。
- 能及时而快速地跟踪市场变化，不断发现能够提供超额收益的新统计模型，寻找新的交易机会。

以上这些量化投资的特点，其实也是我们在投资中需要克服的弱点。那么，如何更加有效地克服这些弱点呢？韩焘先生根据多年的投资经验与量化研究，在本书中给出了精彩的答案。

本书开篇先回顾了量化投资研究发展中的几个过程，并在中间给出了量化投资策略设计的相关思路，包括择时、对冲、风险控制和回测等内容，还指出了人工智能中的数据挖掘、神经网络、机器学习等技术应用于相应投资中的问题与解决方法。我们知道，人工智能各种技术的应用，都需要大量的底层数据，作者在介绍数据获取时也说明了各种数据的获取渠道和清洗方法。韩焘先生指出，得益于移动互联网的快速发展与互联网的宽带化，各种物联网技术的快速发展及源源不断产生的数据，都为人工智能催生的量化投资的发展打下了坚实的基础。

韩焘先生的这本书，不失为当前量化技术丛书中的一抹彩虹，相比市面上琳琅满目的书籍，更具实用价值。韩焘先生凭借自身对量化投资知识的多年投资经验和研究，很多观点见解独到，以过去洞察未来，引导读者认清量化投资技术的真正含义，内容深入浅出，既有专业的介绍，又有通俗的语言。特别是他通过多年的实践和研究及生动的案例得出的结论，以及直言不讳地分享在量化投资应用中的一些弯路更让人钦佩，只有理论和实践相结合才能真正及时发现问题，并给出具有可行性的解决方案，由此给读者带来更大的启发。

本书不仅适用于各类初级投资者，对有一定投资基础且想进行资产配置的投资人也是很好的参考书，书中的投资思想和量化投资策略适用于股票、期货、期权等各类资产配置。他山之石，可以攻玉。愿这本书能为投资者带来不一样的体会与感悟，更愿本书能为投资机构提供更多的参考与帮助。

韩 勇

中信建投证券机构业务部副总裁

作者简介



韩焘

现任北京风云略投资管理有限公司基金经理。
从1998年成功申购基金开元开始入市参与基金和股票投资，拥有20多年的A股投资经验和10余年的程序化交易模型开发经验。

自2015年起，在私募、证券类基金公司担任副总经理和基金经理，管理资产近10亿元，拥有丰富的资产管理经验。截至2019年4月，在私募排排网“私募基金排名”的“近半年”时间段内位居全国第24名，“今年来”时间段内位居全国第139名。

主要研究量化分析及模型选股交易，熟练使用C、Java与Python等编程语言。通过多年积累的实战经验提炼出独有的高性能选股因子，设计开发了量化风控模型、量化交易与对冲模型，大大提升了交易成功率。

好书分享



聚焦互联网前沿技术、金融科技图书出版

微博: @6v1206

E-mail: 6v1206@gmail.com



扫码联系编辑

目 录

第 1 章 量化投资入门.....	1
1.1 量化投资概述.....	1
1.2 量化投资与传统投资的比较.....	2
1.2.1 两种投资策略简介.....	2
1.2.2 量化投资相对于传统投资的主要优势.....	2
1.3 量化投资的国外发展现状及国内投资市场未来展望.....	4
1.3.1 量化金融和理论的建立过程.....	4
1.3.2 国外量化投资基金的发展历史.....	5
1.3.3 国内量化投资基金的发展历史.....	8
1.3.4 国内投资市场的未来展望.....	8
1.4 突发汇率、加息、商誉的应对方法.....	9
1.4.1 突发汇率变化和加息的应对方法.....	10
1.4.2 面对商誉减值的应对方法.....	12
第 2 章 量化投资策略的设计思路.....	17
2.1 量化投资策略的研发流程.....	18
2.2 量化投资策略的可行性研究.....	20
2.3 量化平台常用语言——Python.....	22
2.3.1 Python 简介.....	22
2.3.2 量化基础语法及数据结构.....	23
2.3.3 量化中函数的定义及使用方法.....	40



2.3.4	面向对象编程 OOP 的定义及使用方法	43
2.3.5	itertools 的使用方法	48
2.4	量化投资工具——Matplotlib	51
2.4.1	Matplotlib 基础知识	52
2.4.2	Matplotlib 可视化工具基础	56
2.4.3	Matplotlib 子画布及 loc 的使用方法	58
2.5	Matplotlib 绘制 K 线图的方法	61
2.5.1	安装财经数据接口包 (TuShare) 和绘图包 (mpl_finance)	61
2.5.2	绘制 K 线图示例	62
第 3 章	量化投资策略回测	65
3.1	选择回测平台的技巧	65
3.1.1	根据个人特点选择回测平台	66
3.1.2	回测平台的使用方法与技巧	66
3.2	调用金融数据库中的数据	68
3.2.1	历史数据库的调取	68
3.2.2	数据库的分析方法与技巧	72
3.3	回测与实际业绩预期偏差的调试方法	74
3.4	设置回测参数	75
3.4.1	start 和 end 回测起止时间	75
3.4.2	universe 证券池	76
3.4.3	benchmark 参考基准	78
3.4.4	freq 和 refresh_rate 策略运行频率	78
3.5	账户设置	83
3.5.1	accounts 账户配置	83
3.5.2	AccountConfig 账户配置	85
3.6	策略基本方法	88
3.7	策略运行环境	89
3.7.1	now	90
3.7.2	current_date	90
3.7.3	previous_date	91

3.7.4	current_minute	91
3.7.5	current_price	92
3.7.6	get_account	93
3.7.7	get_universe	93
3.7.8	transfer_cash	95
3.8	获取和调用数据	96
3.8.1	history	96
3.8.2	get_symbol_history	103
3.8.3	get_attribute_history	105
3.8.4	DataAPI	107
3.9	账户相关属性	107
3.9.1	下单函数	107
3.9.2	获取账户信息	115
3.10	策略结果展示	120
3.11	批量回测	122
第4章	量化投资择时策略与选股策略的推进方法	125
4.1	多因子选股策略	125
4.1.1	多因子模型基本方法	125
4.1.2	单因子分析流程	126
4.1.3	多因子(对冲)策略逻辑	134
4.1.4	多因子(裸多)策略逻辑	139
4.2	多因子选股技巧	141
4.2.1	定义股票池	141
4.2.2	指标选股	144
4.2.3	指标排序	145
4.2.4	查看选股	146
4.2.5	交易配置	147
4.2.6	策略回测	147
4.3	择时——均线趋势策略	148
4.3.1	格兰维尔八大法则	149



10	4.3.2 双均线交易系统	150
50	4.4 择时——移动平均线模型	151
50	4.4.1 MA 模型的性质	151
50	4.4.2 MA 的阶次判定	153
20	4.4.3 建模和预测	154
69	4.5 择时——自回归策略	155
29	4.5.1 AR(p)模型的特征根及平稳性检验	156
101	4.5.2 AR(p)模型的定阶	158
201	4.6 择时——均线混合策略	163
103	4.6.1 识别 ARMA 模型阶次	164
101	4.6.2 ARIMA 模型	167
	第 5 章 量化对冲策略	174
111	5.1 宏观对冲策略	174
150	5.1.1 美林时钟	175
151	5.1.2 宏观对冲策略特征	178
251	5.2 微观对冲策略：股票投资中的 Alpha 策略和配对交易	178
251	5.2.1 配对交易策略	178
251	5.2.2 配对交易策略之协整策略	185
150	5.2.3 市场中性 Alpha 策略简介	202
134	5.2.4 AlphaHorizon 单因子分析模块	203
130	5.3 数据加载	204
141	5.3.1 uqer 数据获取函数	204
141	5.3.2 通过 uqer 获取数据	209
141	5.3.3 因子数据简单处理	211
241	5.4 AlphaHorizon 因子分析——数据格式化	213
201	5.5 收益分析	214
241	5.5.1 因子选股的分位数组超额收益	214
241	5.5.2 等权做多多头分位、做空空头分位收益率分析策略	217
181	5.5.3 等权做多多头分位累计净值计算	220
181	5.5.4 多头分位组合实际净值走势图	221

5.5.5	以因子值加权构建组合	222
5.6	信息系数分析	223
5.6.1	因子信息系数时间序列	223
5.6.2	因子信息系数数据分布特征	224
5.6.3	因子信息系数月度热点图	225
5.6.4	因子信息系数衰减分析	226
5.7	换手率、因子自相关性分析	227
5.8	分类行业分析	228
5.9	总结性分析数据	231
5.10	AlphaHorizon 完整分析模板	233
第 6 章	数据挖掘	241
6.1	数据挖掘分类模式	241
6.2	数据挖掘之神经网络	242
6.2.1	循环神经网络数据的准备和处理	243
6.2.2	获取因子的原始数据值和股价涨跌数据	243
6.2.3	对数据进行去极值、中性化、标准化处理	246
6.2.4	利用不同模型对因子进行合成	256
6.2.5	合成因子效果的分析 and 比较	269
6.2.6	投资组合的构建和回测	270
6.2.7	不同模型的回测指标比较	282
6.3	决策树	295
6.3.1	决策树原始数据	295
6.3.2	决策树基本组成	296
6.3.3	ID3 算法	297
6.3.4	决策树剪枝	302
6.4	联机分析处理	303
6.5	数据可视化	304
第 7 章	量化投资中数据挖掘的使用方法	306
7.1	SOM 神经网络	306
7.2	SOM 神经网络结构	307