

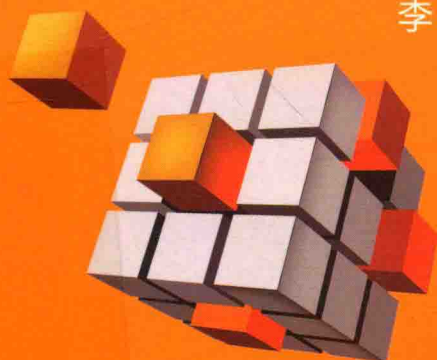


21世纪经济与管理精编教材
金融学系列

量化投资实验

Experiments of Quantitative
Investment

张元萍◎主 编
李 颖 赵 阳◎副主编



北京大学出版社
PEKING UNIVERSITY PRESS

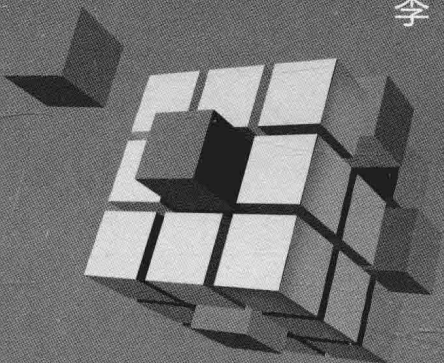


21世纪经济与管理精编教材
金融学系列

量化投资实验

Experiments of Quantitative
Investment

张元萍◎主 编
李 颖 赵 阳◎副主编



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

量化投资实验/张元萍主编.—北京:北京大学出版社,2017.6

(21世纪经济与管理精编教材·金融学系列)

ISBN 978-7-301-28366-0

I. ①量… II. ①张… III. ①投资学—实验—高等学校—教材 IV. ①F830.59

中国版本图书馆 CIP 数据核字(2017)第 105828 号

书 名 量化投资实验

LIANGHUA TOUZI SHIYAN

著作责任者 张元萍 主编 李 颖 赵 阳 副主编

责任编辑 张 燕

标准书号 ISBN 978-7-301-28366-0

出版发行 北京大学出版社

地 址 北京市海淀区成府路 205 号 100871

网 址 <http://www.pup.cn>

电子信箱 em@pup.cn QQ:552063295

新浪微博 @北京大学出版社 @北京大学出版社经管图书

电 话 邮购部 62752015 发行部 62750672 编辑部 62752926

印刷者 北京大学印刷厂

经 销 者 新华书店

787 毫米×1092 毫米 16 开本 9.5 印张 186 千字

2017 年 6 月第 1 版 2017 年 6 月第 1 次印刷

定 价 26.00 元

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。

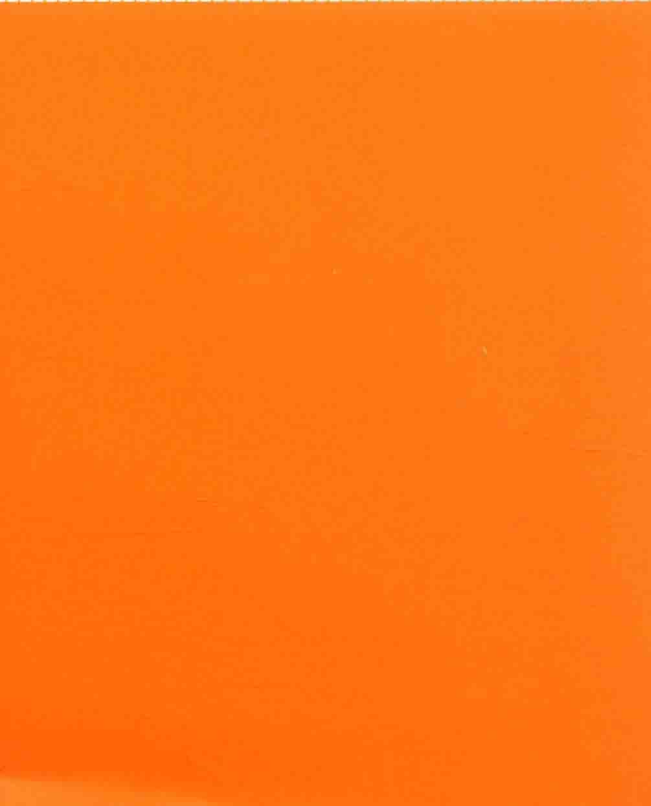
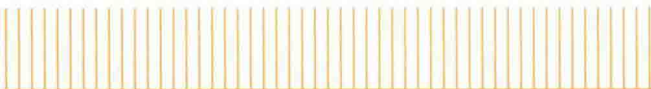
版权所有,侵权必究

举报电话:010-62752024 电子信箱: fd@pup.pku.edu.cn

图书如有印装质量问题,请与出版部联系,电话:010-62756370



张元萍，经济学博士、教授、博士生导师，天津财经大学金融系金融工程教研室主任。1999年享受国务院政府特殊津贴。兼任中国软科学学会理事，天津数量经济学会常务理事，天津金融学会理事。主要研究方向为金融工程、投融资理论与实践。为本科生讲授投资学、金融风险管理、数理金融等课程。



延伸阅读

金融学 / 房燕

国际金融 / 韩莉

数理金融基础 / 张元萍

保险投资学 / 蔡华

企业并购实验教程 / 黄璐

投资银行业务 / 俞姗

商业银行业务管理（第二版） / 杨宜 张峰

中小企业投融资管理 / 杨宜 张峰

责任编辑：张 燕

封面设计：知行北航

010-88380310

前 言

量化投资是将投资理念及策略通过具体的指标、参数,代入设计的模型中,运用模型在历史交易数据中进行回测检验和未来预测,实现数量化证券标的估值的方法。量化交易投资方法在我国证券交易市场的应用较晚,但在华尔街的数十年量化交易应用中,量化投资的算法交易取得了明显优于传统交易策略的超额收益。随着国内投资标的的扩容和交易量的激增,加上互联网的发展,量化投资交易逐渐在我国兴起。

量化投资技术几乎覆盖了投资的全过程,包括量化选股、量化择时、期货套利、统计套利、算法交易、资产配置、风险控制等。全书共分九章,分别为量化投资基础及实验操作平台、量化选股实验模块、量化择时实验模块、统计套利实验模块、ETF 套利交易实验模块、算法交易实验模块、人工智能实验模块、数据挖掘实验模块、支持向量机实验模块。每章均介绍了该模块中涉及的理论知识,并通过一组实际案例习题详细讲解量化投资相应技术方法、程序设计及软件操作。每个案例操作还配有视频演示。

本教材适用于金融工程专业全日制本科生量化投资实验教学课程使用。进行量化投资实验需要掌握编程语言和量化投资平台的操作方法,本实验教材采用的程序语言是 Matlab,量化投资案例分析采用国泰安 QIA 量化投资平台进行模型检验和回测。

本教材由天津财经大学张元萍教授任主编,天津财经大学李颖、深圳国泰安教育技术股份有限公司北京分公司总经理赵阳任副主编。第 1、2、6 章由天津财经大学张元萍、封笑笑编写,第 3、5、8 章由天津财经大学李颖、史德坤编写,第 4、7、9 章由天津财经大学李颖、赵亿编写。在本书出版过程中得到了北京大学出版社的支持,在此表示衷心的感谢。

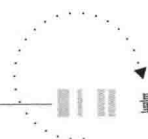
编 者

2017 年 3 月 1 日

目 录

第 1 章 量化投资基础及实验操作平台	1
1.1 量化投资的基本理念和特点	1
1.2 量化投资的主要策略和方法	4
1.3 量化投资实验的知识储备和基本要求	7
1.4 量化投资实验操作平台	10
第 2 章 量化选股实验模块	26
2.1 实验目的与要求	26
2.2 实验基础知识	26
2.3 实验步骤及示例	28
2.4 实验操作录像	51
第 3 章 量化择时实验模块	52
3.1 实验目的与要求	52
3.2 实验基础知识	52
3.3 实验步骤及示例	54
3.4 实验操作录像	65
第 4 章 统计套利实验模块	66
4.1 实验目的与要求	66
4.2 实验基础知识	66
4.3 实验步骤及示例	70
4.4 实验操作录像	82
第 5 章 ETF 套利实验模块	83
5.1 实验目的与要求	83
5.2 实验基础知识	83
5.3 实验步骤及示例	87

5.4 实验操作录像	97
第 6 章 算法交易实验模块	98
6.1 实验目的与要求	98
6.2 实验基础知识	98
6.3 实验步骤及示例	99
6.4 实验操作录像	105
第 7 章 人工智能实验模块	106
7.1 实验目的与要求	106
7.2 实验基础知识	106
7.3 实验步骤及示例	111
7.4 实验操作录像	117
第 8 章 数据挖掘实验模块	118
8.1 实验目的与要求	118
8.2 实验基础知识	118
8.3 实验步骤及示例	122
8.4 实验操作录像	127
第 9 章 支持向量机实验模块	128
9.1 实验目的与要求	128
9.2 实验基础知识	128
9.3 实验步骤及示例	137
9.4 实验操作录像	143
参考文献	145



第1章 量化投资基础及实验操作平台

量化投资是整个金融市场发展的一个必然阶段，也是市场有效性在金融市场上不断进步和发展的必然趋势。由于量化投资交易策略的严谨性、交易回溯的可逆性、数据的直观性，量化投资已经成为全球金融投资的手段和方法之一。

1.1 量化投资的基本理念和特点

1.1.1 量化投资的基本理念

量化投资通常是指使用数学公式、根据过去的数据来判断将来价格走势而进行投资的一种方法。换句话说，它是一种比较机械的，不需要人来进行判断、干预的投资方法。

量化投资就是利用计算机技术并采用一定的数学模型去实现投资理念、实现投资策略的过程。传统的投资方法主要有基本面分析法和技術分析法两种。与它们不同的是，量化投资主要依靠数据和模型来寻找投资标的和投资策略。

量化投资模型很多是基于基本面因素，同时考虑市场因素、技术因素等。量化投资是一种主动投资策略，主动型投资的理论基础就是市场是非有效的或是弱有效的，基金经理可以通过对个股、行业及市场的驱动因素进行分析研究，建立最优的投资组合，试图战胜市场从而获取超额收益。

量化投资不是靠个人感觉来管理资产，而是将适当的投资思想、投资经验甚至直觉反映在量化模型中，利用计算机帮助人脑处理大量信息，帮助人脑总结归纳市场的规律，建立可以重复使用并反复优化的投资策略（经验），并指导我们的投资决策过程。量化投资模型都必须经历不断的跟踪检验、优化、实证等过程。量化投资是一个不断改进的过程，最重要的就是投资者的投资思想，包括对投资的理解、理念、经验，所有模型都是建立在这些思想上的。

量化投资方法虽然起源于一百年前，但是在过去三十年中迅速发展。目前据估计，量化投资金额占全球投资总额的 30%。在量化投资的历史上，詹姆斯·西蒙斯（James Simons）和他的“大奖基金”取得了高额的收益率，连巴菲特都望尘莫及。由此，量化投资逐渐进入人们的视野。量化投资策略以正确的投资理念为根本，通过各种因素的分析，以全市场的广度、多维度的深度视角扫描投资机会，在中国市场的应用将更凸显出其优势。

1.1.2 量化投资的特点

量化投资的最大特点就是定量化和精细化。“一切用数据说话”，是量化投资决策的基石。量化投资的特点包括纪律性、系统性、妥善运用套利思想、及时性、准确性、分散化等几个方面。

1. 纪律性

所有的决策都是依据模型做出的。纪律性首先表现在依靠模型和相信模型，每一次决策之前，首先要运行模型，根据模型的运行结果进行决策，而不是凭感觉进行决策。要严格执行量化投资模型所给出的投资建议，而不是随着投资者情绪的变化而随意更改。纪律性的好处很多，可以克服人性的弱点，如贪婪、恐惧、侥幸心理，也可以克服认知偏差。

2. 系统性

系统性具体表现为“三多”。首先表现在多层次，包括大类资产配置、行业选择、精选个股三个层次；其次是多角度，定量投资的核心投资思想包括宏观周期、市场结构、估值、成长、盈利质量、分析师盈利预测、市场情绪等多个角度；最后就是多数据，即海量数据的处理能力。人脑处理信息的能力是有限的，在一个很大的资本市场，强大的定量投资的信息处理能力能捕捉更多的投资机会，拓展更大的投资机会。

3. 妥善运用套利思想

定量投资正是通过寻找估值洼地，通过全面、系统性的扫描捕捉错误定价、错误估值带来的机会。定性投资经理大部分时间在琢磨哪一个企业是伟大的企业，哪个股票是可以翻倍的股票；与定性投资经理不同，定量基金经理大部分精力花在分析哪里是估值洼地，哪一个品种被低估了，买入低估的，卖出高估的。

4. 及时性和准确性

及时性是指及时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。准确性是指准确客观评价交易机会，克服主观情绪偏差，妥善运用套利的思想。

5. 分散化

在控制风险的前提下，量化投资充当着准确实现分散化投资目标的工具。也可以说，量化投资是靠概率取胜。这表现为两个方面：一是量化投资不断地从历史中挖掘有望在未来重复的历史规律并且加以利用，这些历史规律都是有较大概率获胜的策略；二是依靠筛选出股票组合来取胜，而不是靠一只或几只股票取胜，从投资组合理念来看也是捕获大概率获胜的股票，而不是押宝到单只股票上。

1.1.3 量化投资的操作手法

1. 统计分析

在一开始收集标的的数据进行分析的时候，工作量是巨大的。比如我们想看看宏观经济环境和股票的表现有没有相应的规律，就要研究政府政策和各类宏观经济指标与股票走势、板块走势之间的关系。如果我们发现其中有某些规律，那么还要进行回测，看看这个规律是不是具有代表性，然后才能进行下一步的数学建模。而通常能做这样大型的量化分析的主要是一些量化基金管理团队，几个人甚至几十个人一起用计算机检验收集的数据是否有效，能否成为建模的对象，从众多的数据中找出一些逻辑上会影响到股价走势，或许预示股价走势的数据。这是量化分析中的一个简单而有效的思路。

2. 数学建模

近年来，量化投资作为一种投资方法受到越来越多的关注。公式之所以能够赚钱，背后通常是一些很简单、很常识的东西，而数学建模就是把所收集到的合格的数据用数学的语言表达出来。

3. 虚拟操盘

再好的理论也要通过实践来检验效果。所有建成的模型都要放于虚拟盘中模拟一段时间才能真正为投资服务。而这个模拟期少则半年，多则几载，因为在模拟的过程中还需要对模型进行不断的更新和修正才能使其更加完美。任何一个模型都不可能是完美的，虚拟操盘时要测试模型是否能够达到80%以上的正确率，如果达到了80%就说明这个模型是成功的，如果没有达到80%则证明这个模型在某些地方还是需要改进一下，或许根本不成立。这是一个循环往复的过程，也是通向成功必需的过程。

1.2 量化投资的主要策略和方法

1.2.1 量化投资的主要策略

量化投资的主要策略包括量化选股、量化择时、股指期货套利、商品期货套利、统计套利、期权套利、算法交易、资产配置等。

1. 量化选股

量化选股就是采用某种数量方法判断某个公司是否值得买入的行为。如果该公司满足了该方法的条件，则放入股票池；如果不满足该方法的条件，则从股票池中剔除。量化选股的方法有很多种，总的来说，可以分为公司估值法、趋势法和资金法三大类。

2. 量化择时

随着计算机技术、混沌、分形理论的发展，人们开始将股票的市场行为纳入非线性动力学的研究范畴。众多的研究发现，我国股市的指数收益中，存在经典线性相关之外的非线性相关，从而拒绝了随机游走的假设，说明股价的波动不是完全随机的，它貌似随机、杂乱，但在其复杂表面的背后，却隐藏着确定性的机制，因此存在可预测成分。

3. 股指期货套利

股指期货套利是指利用股指期货市场存在的不合理价格，同时参与股指期货与股票现货市场交易，或者同时进行不同期限、不同（但相近）类别股票指数合约交易，以赚取差价的行为。股指期货套利主要分为期现套利和跨期套利两种。股指期货套利的研究主要包括现货构建、套利定价、保证金管理、冲击成本、成分股调整等内容。

4. 商品期货套利

商品期货套利盈利的逻辑原理基于以下几个方面：相关商品在不同地点、不同时间都对应着一个合理的价格差价；由于价格的波动性，价格差价经常出现不合理；不合理必然要回到合理；不合理回到合理的这部分价格区间就是盈利区间。

5. 统计套利

统计套利是利用证券价格的历史统计规律进行套利，是一种风险套利，其风险在于这种历史统计规律在未来一段时间内是否继续存在。统计套利在方法上可以分为两类，一类是利用股票的收益率序列建模，目标是在组合的 β 值等于零的前提下实现 α 收益，

我们称之为 β 中性策略；另一类是利用股票的价格序列的协整关系建模，我们称之为协整策略。前者是基于日收益率对均衡关系的偏离，后者是基于累计收益率对均衡关系的偏离。

6. 期权套利

期权套利是指同时买进和卖出同一相关期货，但不同敲定价格或不同到期月份的看涨或看跌期权合约，希望在日后对冲交易部位或履约时获利的交易策略。期权套利的交易策略和方式多种多样，是多种相关期权交易的组合，具体包括水平套利、垂直套利、转换套利、反向转换套利、跨式套利、蝶式套利、飞鹰式套利等。

7. 算法交易

算法交易又被称为自动交易、黑盒交易或者机器交易，它指的是通过使用计算机程序来发出交易指令。在交易中，程序可以决定的范围包括交易时间、交易价格，甚至最后需要成交的证券数量。根据各个算法交易中算法的主动程度不同，可以把算法交易分为被动型算法交易、主动型算法交易、综合型算法交易三大类。

8. 资产配置

资产配置是指资产类别选择，即投资组合中各类资产的适当配置及对这些混合资产进行实时管理。量化投资管理将传统投资组合理论与量化分析技术结合，极大地丰富了资产配置的内涵，形成了现代资产配置理论的基本框架。资产配置一般包括两大类别、三大层次，两大类别分别为战略资产配置和战术资产配置，三大层次分别为全球资产配置、大类资产配置及行业风格配置。

1.2.2 量化投资的主要方法

量化投资涉及很多数学和计算机方面的知识和技术，总的来说，主要有人工智能、数据挖掘、小波分析、支持向量机、分形理论和随机过程这几种。

1. 人工智能

人工智能（artificial intelligence, AI）是研究使用计算机来模拟人的某些思维过程和智能行为（如学习、推理、思考、规划等）的学科，主要是利用计算机实现智能的原理，制造类似于人脑智能的计算机，使计算机能实现更高层次的应用。金融投资是一项复杂的、综合了各种知识与技术的学科，对智能的要求非常高。所以人工智能的很多技术可以用于量化投资分析中，包括专家系统、机器学习、神经网络、遗传算法等。

2. 数据挖掘

数据挖掘 (data mining) 是从大量的、不完全的、有噪声的、模糊的、随机的数据中提取隐含在其中的、人们事先不知道的, 但又是潜在有用的信息和知识的过程。与数据挖掘相近的同义词有数据融合、数据分析和决策支持等。在量化投资中, 数据挖掘的主要技术包括关联分析、分类/预测、聚类分析等。

3. 小波分析

小波 (wavelet) 这一术语, 顾名思义, 就是小的波形。所谓“小”是指它具有衰减性; 而称之为“波”则是指它的波动性, 即其振幅正负相间的振荡形式。小波分析在量化投资中的主要作用是进行波形处理。任何投资品种的走势都可以看做一种波形, 其中包含了很多噪声信号。利用小波分析, 可以进行波形的去噪、重构、诊断、识别等, 从而实现对未来走势的判断。

4. 支持向量机

支持向量机 (support vector machine, SVM) 方法是通过一个非线性映射, 把样本空间映射到一个高维乃至无穷维的特征空间中 (Hilbert 空间), 使得在原来的样本空间中非线性可分的问题转化为在特征空间中的线性可分的问题, 简单地说, 就是升维和线性化。支持向量机方法特别适合于进行有关分类和预测问题的处理, 这就使得它在量化投资中有了很大的用武之地。

5. 分形理论

分形理论既是非线性科学的前沿和重要分支, 又是一门新兴的横断学科。作为一种方法论和认识论, 其启示是多方面的: 一是分形整体与局部形态的相似, 启发人们通过认识部分来认识整体, 从有限中认识无限; 二是分形揭示了介于整体与部分、有序与无序、复杂与简单之间的新形态、新秩序; 三是分形从某个特定层面揭示了世界普遍联系和统一的图景。

由于这种特征, 分形理论在量化投资中得到了广泛应用, 主要可以用于金融时序数列的分解与重构, 并在此基础上进行数列的预测。

6. 随机过程

随机过程 (stochastic process) 是一连串随机事件动态关系的定量描述。研究随机过程的方法多种多样, 主要可以分为两大类: 一类是概率方法, 其中用到轨道性质、随机微分方程等; 另一类是分析的方法, 其中用到测度论、微分方程、半群理论、函数堆

和 Hilbert 空间等, 实际研究中常常两种方法并用。另外, 组合方法和代数方法在某些特殊随机过程的研究中也有一定作用。研究的主要内容有多指标随机过程、无穷质点与马尔可夫过程、概率与位势及各种特殊过程的专题讨论等。其中, 马尔可夫过程很适合用于金融时序数列的预测, 是在量化投资中的典型应用。

1.3 量化投资实验的知识储备和基本要求

进行量化投资实验的基本要求, 是要具备一定的数学建模基础和高级编程语言基础。计算机语言的种类很多, 本书主要介绍 C++ 语言、Matlab 语言和 R 语言。重点掌握 Matlab 语言。

1.3.1 C++ 语言

C++ 语言是一种优秀的程序设计语言, 它在 C 语言的基础上发展而来, 但它比 C 语言更容易为人们学习和掌握。C++ 以其独特的语言机制在计算机科学的各个领域得到了广泛的应用。面向对象的设计思想是在原来结构化程序设计方法基础上的一个质的飞跃, C++ 完美地体现了面向对象的各种特性。目前 C++ 语言是金融工程领域的标准语言, 擅长数值计算, 应用比较广泛。

(一) 程序编写步骤

1. 用 C++ 语言编写程序

用高级语言编写的程序称为“源程序” (source program)。C++ 的源程序是以“.cpp”作为后缀的 (cpp 是 c plus plus 的缩写)。

2. 对源程序进行编译

为了使计算机能执行高级语言源程序, 必须先用一种被称为“编译器” (compiler) 的软件 (也称编译程序或编译系统), 把源程序翻译成二进制形式的“目标程序” (object program)。

编译是以源程序文件为单位分别编译的。目标程序一般以“.obj”或“.o” (object 的缩写) 作为后缀。编译的作用是对源程序进行词法检查和语法检查。编译时对文件中的全部内容进行检查, 编译结束后会显示出所有的编译出错信息。一般编译系统给出的出错信息分为两种: 一种是错误 (error), 一种是警告 (warning)。

3. 将目标文件连接

在改正所有的错误并全部通过编译后，得到一个或多个目标文件。此时要用系统提供的“连接程序”（linker）将一个程序的所有目标程序和系统的库文件以及系统提供的其他信息连接起来，最终形成一个可执行的二进制文件，它的后缀是“.exe”，是可以直接执行的。

4. 运行程序

运行最终形成的可执行的二进制文件（.exe 文件），得到运行结果。

5. 分析运行结果

如果运行结果不正确，应检查程序或算法是否有问题。

（二）C++ 语言的优点

C++ 语言具有以下主要优点：

（1）C++ 语言设计成静态类型、和 C 语言同样高效且可移植的多用途程序设计语言。

（2）C++ 语言可以支持多种程序设计风格（程序化程序设计、资料抽象化、面向对象程序设计、泛型程序设计）。

（3）C++ 语言给程序设计者提供更多的选择，并尽可能与 C 语言兼容，借此提供一个从 C 语言到 C++ 语言的平滑过渡。

（4）C++ 语言避免了平台限定，普遍适用于大多数场合。

（5）C++ 语言的开发成本较低，并且代码量越大，这个优势越明显。

（6）C++ 语言无需复杂的程序设计环境。

（7）出于保证语言的简洁和运行高效等方面的考虑，C++ 语言的很多特性都是以库（如 STL）或其他的形式提供的，而没有直接添加到语言本身里。

（8）C++ 语言在一定程度上可以和 C 语言很好地结合，甚至大多数 C 语言程序是在 C++ 语言的集成开发环境中完成的。C++ 语言相对于众多面向对象的语言，具有相当高的性能。

（9）C++ 语言引入了面向对象的概念，使得开发人机交互类型的应用程序更为简单、快捷。很多优秀的程序框架如 MFC、QT、wxWidgets 等使用的都是 C++ 语言。

1.3.2 Matlab 语言

Matlab 是一种高级技术计算语言和交互式环境，主要用于算法开发、数据可视化、

数据分析以及数值计算。

与其他编程软件相比, Matlab 在算法应用上具有以下优势:

1. 简单易用

Matlab 语言是一个高级的矩阵/阵列语言, 语法特征与 C++ 语言相似, 语言简单且书写格式与科技人员对数学表达式的书写格式一致, 更利于非计算机专业人员的使用。Matlab 语言具有移植性、可拓展性强的特点, 编程周期短、效率高。

2. 强处理能力

Matlab 是一个包含大量计算算法的集合, 拥有众多运算函数, 包括矩阵、特征向量、快速傅里叶变换等复杂函数, 可以方便地实现计算功能。经过各种优化和容错处理, Matlab 可替代底层编程语言, 减少工作量, 并能解决包括各类运算求解和多维数组操作以及建模动态仿真等问题。

3. 强图像处理

Matlab 的数据可视化功能将向量和矩阵用图形表示出来, 独有的图形光照处理、色度处理以及四维数据的表现等, 为科学计算和工程绘图提供了完善的图形处理功能。同时 Matlab 还保证了用户不同层次的要求, 加入了图形对话, 改善了图形界面 (GUI)。

4. 应用程序接口

应用程序接口 (API) 帮助 Matlab 实现了与其他高级编程语言 (C 语言、Fortran 语言等) 进行交互的功能。实现方式是通过调节动态链接库 (DLL), 进行与 Matlab 文件的数据交换。

5. 开源工具箱支持

数百个内部函数的主包和三十几种工具包为 Matlab 实现函数计算、可视化建模仿真、文字处理以及专业性的功能提供了技术支持, 目前也已加入多个支持金融建模的工具箱。Matlab 的开放性使得用户可以根据自身需求, 对源程序进行修改或加入自己的编程, 构造新的专用工具箱。

1.3.3 R 语言

R 语言主要用于统计分析、绘图的语言和操作环境。R 语言最初是由来自新西兰奥克兰大学的 Ross Ihaka 和 Robert Gentleman 开发的 (也因此被称为 R 语言), 现在由