

电子科技大学
UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

专业学位硕士学位论文

MASTER THESIS FOR PROFESSIONAL DEGREE



论文题目 多因子量化投资策略的构建及应用

| | |
|--------|---------------------|
| 专业学位类别 | <u>工商管理硕士</u> |
| 学 号 | <u>201972151341</u> |
| 作者姓名 | <u>李丽佼</u> |
| 指导教师 | <u>赵 武 副教授</u> |
| 学 院 | <u>经济与管理学院</u> |

分类号 _____ 密级 _____ 公开 _____
UDC ^{注1} _____

学 位 论 文

多因子量化投资策略的构建及应用

李丽佼

(作者姓名)

指导教师 赵 武 副教授
电子科技大学 成 都

(姓名、职称、单位名称)

申请学位级别 硕士 专业学位类别 工商管理硕士
提交论文日期 2022 年 3 月 25 日 论文答辩日期 2022 年 5 月 26 日
学位授予单位和日期 电子科技大学
答辩委员会主席 _____
评阅人 _____

注 1: 注明《国际十进分类法 UDC》的类号。

Construction and Application of Multi-Factor Quantitative Investment Strategy

A Master Thesis Submitted to
University of Electronic Science and Technology of China

Discipline **Master of Business Administration**

Student ID **201972151341**

Author **Giselle Li**

Supervisor **Associate Prof Zhao Wu**

School **School of Management and Economics**

摘 要

量化投资产品是指通过应用数量化方式以及程序编程来完成交易市场中的操作，以一定的模型、计算方法、理论体系为基础，实现以获取较为稳定的收益为目标的自动化投资理财模式。2019 年量化多头产品平均收益 33.82%，而同期的 A 股平均收益率为 26.48%，量化投资产品不仅在收益率上比较稳定，同时其市场波动情况较少，在量化产品中极少出现大幅度的回撤。当前随着信息技术的发展，以及人们对信息技术在各领域中应用的接受度不断增加，国内金融市场内针对量化等相关技术的关注度也在不断提升，这也表明了国内的金融投资环境已经发生了重大改变。在此背景下，有必要针对 A 股市场展开量化选股策略的研究，也为投资者进行投资组合策略的决策提供相应的依据。

本文的研究内容包括以下几个方面：首先对量化投资策略的相关研究背景、研究意义、国内外研究现状、研究总体思路等进行概述。随后对本文涉及到的投资理论、资产定价理论、量化投资的模型，Fama-French 三因子模型等进行系统的介绍，为后文的研究提供理论依据。在 Fama-French 三因子模型基础上加入了新浪大数据 100 因子，并提出了这一因子加入后的 Fama-French 模型，并分析了模型中的因子构成以及具体的模型构建与运行步骤。在上述基础上结合沪深 A 股的数据，对加入了大数据因子后的 Fama-French 模型的有效性进行测算，结果发现大数据因子能够有效地解释沪深 A 股，在投资策略的构建中有重要的参考价值。最后利用 Choice 的量化分析功能，对 2016-2021 的股票数据进行测算，优化了投资模型，并基于量化投资模型给出了具体的投资建议。

关键词：Fama-French 模型；大数据因子；量化投资

ABSTRACT

Quantitative investment products refer to the automatic investment and financial management mode that completes the operation in the trading market through the application of quantitative methods and program programming, and realizes the goal of obtaining more stable income based on a certain model, calculation method and theoretical system. In 2019, the average yield of quantitative long products was 33.82%, while the average yield of a shares in the same period was 26.48%. Quantitative investment products not only have relatively stable yield, but also have less market volatility, and there are few significant pullbacks in quantitative products. At present, with the development of information technology and the increasing acceptance of the application of information technology in various fields, the attention to quantification and other related technologies in the domestic financial market is also increasing, which also shows that the domestic financial investment environment has changed significantly. In this context, it is necessary to carry out quantitative stock selection strategy research for the A-share market, and also provide corresponding basis for investors to make portfolio strategy decisions.

The research content of this paper includes the following aspects: firstly, it summarizes the relevant research background, research significance, research status at home and abroad and the general idea of quantitative investment strategy. Then it systematically introduces the investment theory, asset pricing theory, quantitative investment model and Fama French three factor model involved in this paper, so as to provide a theoretical basis for the later research. Then, Sina big data 100 factor is added to the Fama French three factor model, and the Fama French model with this factor is proposed, and the factor composition in the model, as well as the specific model construction and operation steps are analyzed. Based on the above, combined with the data of Shanghai and Shenzhen A shares, the effectiveness of Fama French model with big data factor is calculated. The results show that big data factor can effectively explain Shanghai and Shenzhen A shares, which has important reference value in the construction of investment strategy. Finally, based on the investment data of choice-2016, the quantitative analysis of the investment model is given, and the specific

investment suggestions are given.

Key words: Fama-French model; Big data factor; Quantitative investment

目 录

| | |
|--------------------------------|----|
| 第一章 绪论 | 1 |
| 1.1 研究背景 | 1 |
| 1.2 研究意义 | 3 |
| 1.3 研究综述 | 4 |
| 1.3.1 国外研究现状..... | 4 |
| 1.3.2 国内研究现状..... | 6 |
| 1.3.3 研究评述..... | 8 |
| 1.4 研究主要内容及框架 | 8 |
| 1.4.1 研究内容..... | 8 |
| 1.4.2 研究框架..... | 9 |
| 第二章 因子组合量化投资理论基础 | 10 |
| 2.1 现代资产配置理论 | 10 |
| 2.2 资本资产定价模型（CAPM） | 13 |
| 2.3 Fama-French 三因素模型 | 15 |
| 2.4 套利定价理论（APT） | 15 |
| 2.5 本章小结 | 16 |
| 第三章 因子组合量化投资模型建模过程 | 17 |
| 3.1 因子组合量化投资策略 | 17 |
| 3.2 大数据因子 | 19 |
| 3.2.1 大数据因子基本情况..... | 19 |
| 3.2.2 大数据指数的特点..... | 20 |
| 3.2.3 大数据指数的构造情况..... | 22 |
| 3.2.4 2021 年大数据指数变动情况..... | 24 |
| 3.3 基于大数据因子的 French 模型构建 | 25 |
| 3.4 自变量因子构成 | 25 |
| 3.5 因变量的选取 | 26 |
| 3.6 技术准备 | 27 |
| 3.7 本章小结 | 28 |
| 第四章 实证分析 | 29 |
| 4.1 描述性统计分析 | 29 |

| | |
|---|-----------|
| 4.1.1 行业分析..... | 29 |
| 4.1.2 因子相关性检验..... | 30 |
| 4.1.3 BM 效应、规模效应分析 | 30 |
| 4.2 Fama-Macbeth 回归分析 | 31 |
| 4.3 引入大数据因子改进的 Fama-French 模型适应性分析 | 32 |
| 4.3.1 股票各板块回归系数分析..... | 33 |
| 4.3.2 Fama-French 模型中引入大数据因子更适合中国投资市场 | 40 |
| 4.4 本章小结 | 43 |
| 第五章 量化策略的设计及优化 | 41 |
| 5.1 Choice 量化策略平台 | 41 |
| 5.2 量化投资策略设计和优化 | 42 |
| 5.2.1 量化投资策略设计..... | 42 |
| 5.2.2 量化投资策略的优化..... | 45 |
| 5.2.3 量化投资策略的优化结果..... | 46 |
| 5.3 投资建议 | 47 |
| 5.4 本章小结 | 48 |
| 第六章 结论及展望 | 49 |
| 致 谢 | 51 |
| 参考文献 | 52 |

图目录

| | |
|---|----|
| 图 1-1 2018-2019 年散户投资收益率 | 1 |
| 图 1-2 上投摩根阿尔法混合累计净值 | 3 |
| 图 1-3 本文的研究框架图 | 9 |
| | |
| 图 2-1 效用函数 | 10 |
| 图 2-2 有效前沿 | 12 |
| 图 2-3 穿过有效前沿的射线 | 13 |
| 图 2-4 资本市场线上的市场组合 | 13 |
| | |
| 图 3-1 投资的三个过程 | 17 |
| 图 3-2 美国智能化量化平台管理规模 | 19 |
| 图 3-4 2021 年大数据指数 100 变动情况 | 24 |
| 图 3-5 2021 年沪深 A 股数据变化情况 | 25 |
| 图 3-6 对沪深 A 股中股票的证券板块、所属行业板块进行筛选 | 27 |
| 图 3-7 单个股票数据整理 | 27 |
| | |
| 图 5-1 Choice 量化平台界面 | 41 |
| 图 5-2 2016 年 1 月 1 日—2021 年 12 月 31 日测算结果 | 44 |
| 图 5-3 重新设计后的四因子量化选股模型测算结果 | 46 |

表目录

表 3-1 大数据 100 指数的最新指数成分 21

表 3-2 不同组合的具体含义 26

表 4-1 沪深 A 股市场板块数据描述 29

表 4-2 pearson 相关系数 30

表 4-3 三因素模型按市值比的 6 组合收益率 30

表 4-4 三因素模型按市值比的 6 组合方差 30

表 4-5 四因子模型的检验结果 31

表 4-6 四因子模型中各板块回归系数分析 32

表 5-1 股票收益率影响因子 43

表 5-2 计算超额收益率及排名 43

表 5-3 四因子模型的选股测算收益率结果 44

表 5-4 四因子模型的投资测算收益率结果 46

第一章 绪论

1.1 研究背景

当前在我国股票市场之中投资活动的风险不断加剧，尤其是在近年来面对错综复杂的政治环境、地缘经济、国内外资金流向等复杂因素，股票市场的波动往往并非数个指标、或者单一技术方法能够进行预测的，而应用单一指标、基本面进行判断往往具有较高风险。如下图所示，为我国 2018-2019 年散户投资收益情况：

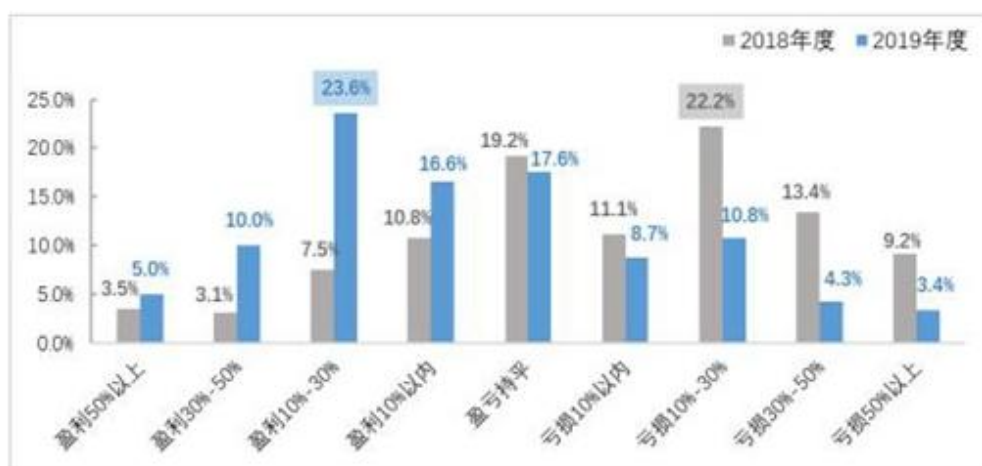


图 1-1 2018-2019 年散户投资收益率

数据来源：新浪财经《2018-2019 年投资者盈亏情况调查》

经上图可发现，在 2018-2019 年期间，我国散户投资收益率与散户投资亏损程度均较高。而在投资市场中，能获得收益并将收益率保持在 10-30%、30% 以上的散户则只有少数（占比 36.5%）。从上图可以看出有超过 54.3% 的散户亏损在 10%-50% 以上。从这一结果也可以看出，散户这种基于基本面或者较为单一的技术方法进行购买的股票组合由于缺乏专业的投资分析策略加持，具有较大的投资风险。

而在此背景下，部分散户、股票投资人开始将视线转移到量化交易上。量化投资产品是指通过应用数量化方式以及程序编程来完成交易市场中的操作，以一定的模型、计算方法、理论体系为基础，实现以获取较为稳定的收益为目标的自动化投资理财模式。通过量化交易能够对市场上的大部分因素进行考察，同时还能够对股票市场未来的走向进行一定程度预测，并能够最终为投资者的决策提供

依据。当前比较主流的量化投资理论主要包括：马克维茨均值-方差理论、CAPM 理论、以及 Fama-French 的多因子选股模型。

量化投资主要包括择时策略、选股策略、资产优化配置策略等：（1）择时策略作为量化投资策略下的第一重要环节，将对证券发展建模分析，从而获得对未来价格涨幅及特定时间证券价格的预估，并判断最佳买入、卖出的时机。（2）选股策略，将股票历史价格、历史对应时间与企业过往生产经营数据指标作为分析要素，并结合量化模型的算法获取较为准确的结果。所以量化投资的特征是结合基本信息、加入各类技术分析方法，并快速获取结论，最终进行股票投资的决策，构建投资组合策略；（3）资产优化配置，对于不同的资产结构其具体优化要参考上述的择时策略、以及选股策略，不可盲目的投入所有的本金，而是要按照自身的择时策略、选股策略在既定的时间点买入与卖出，分配好资金的利用计划、提升资产的利用效率。

而无论是哪种量化分析方法，都需要在量化分析模型的基础上展开：（1）其中马克维茨均值-方差理论中将股票收益方差作为市场风险因素，对关键要素投入股票收益与风险进行衡量，并得出投资策略。（2）CAPM 理论是以资产资本定价作为核心要素建立的分析模型，这一模型也是较早的量化投资模型，能够较好的考察资本投入以及资本投入的效率、回收价值等。（3）Fama-French 的多因子选股模型是当前应用较为广泛的模型，这一模型之中对资产的组合、公司规模、账面市值比等进行了较为详细的研究与分析，这一模型最终能够确认某种投资组合的收益预测情况，进而能够对投资者的决策提供依据。但以上三种分析方式，都是工具性质，量化投资分析的根本意义在于如何科学的指导投资行为。市场中的各类因素、市场本身都在不断的发生变化，单一的技术性分析也难以紧跟市场动态。因此，传统的量化分析模型之中难以对投资者的情绪以及一些市场以外的因素进行测算，这些都使得传统的量化分析与当前的市场情况存在一定的延迟和差异。较为典型的例子，如基金投资，量化因素分析下操作的投资在市场没有较大的外力因素影响的时候，收益率始终保持在较好的稳定情况，但对于突发事件缺乏应对能力。如下图所示为上投摩根阿尔法混合累计净值，这一量化产品在 2018-2020 年上涨幅度超过 46%，但在 2022 年该基金首次出现负增长，一度跌幅达到 17.55%。



图 1-2 上投摩根阿尔法混合累计净值

数据来源：天天基金网 2022 年 3 月 15 日数据

其核心原因在于，传统的量化分析模型之中缺乏对市场更为具体的观察，无论是投资组合、还是公司规模、账面市值比等其实都没有从投资者的角度去考察与分析。众所周知，我国股票尤其是 A 股市场是一个较为感性的市场，散户投资者们缺乏理性，而对这类占比数量较大的散户的情绪进行观察，并纳入到量化投资模型之中，就能够优化量化投资模型，同时也能够更好地为投资者的决策提供依据。因此，本文将在此背景下，基于传统的量化投资模型，引入能够观测散户投资者情绪的因子“大数据因子”，并对这一模型的实际应用适应性展开分析，进行模型的优化与应用，最终提出基于改进后的量化投资模型的股票市场投资建议。

1.2 研究意义

本文的研究意义在于基于我国股市实际发展现状下，结合现实因素构建定价因子模型；通过 Fama-French 三因子模型的改进，引入大数据因子进行分析，更合理地对股票市场进行预测。传统的量化分析模型主要是以市场因素为主要分析目标，如 Fama-French 三因子模型就在 CAPM 的基础上引入了 SMB、HML 这两个因子，但这一模型仍然缺乏对投资者情绪的观察，缺乏对市场技术面以外的观察，最终所获得的分析结果不一定符合我国的股票市场环境。同时，由于大数据

指数能够直接反映出市场中投资者的情绪，通过大数据 100 指数能够直接观察投资者的投资行为，大数据因子的计算逻辑之中考察了一个时间阶段年内用户的交易行为，因此通过大数据 100 的数据变动能够分析投资者的情绪变化情况。因此，引入大数据因子，能够有效弥补 Fama-French 三因子模型在股票市场上的适应性，能够得出更为准确、有效的分析结论，提供给投资者参考，因此本文将对 Fama-French 三因子模型进行改进，引入大数据因子进行优化，并结合实际的市场数据进行分析，为投资者、投资机构展开量化分析投资提供一定的参考。

1.3 研究综述

1.3.1 国外研究现状

如今的 Fama-French 五因子模型，由 Sharpe（1964）提出，但早期模型只是单因子模型。早期模型仅考虑市场风险因素，在不断的发展过程下，模型才逐渐综合企业规模、股票流动性、动量效应、产品盈利能力等因素，模型的资产定价理论同样在不断发展中得以完善。Sharpe 所提出的 CAPM 模型，解释了资产收益与资产风险的关系，并成为现代金融市场价格理论基础。但同时 CAPM 模型并没有对市场出现的市值效应进行有效解释。直到 Banz（1986）发现市值与收益率存在深度关联，但 CAPM 模型难以对超额收益进行完整解释。所以 Fama、French（1933）等在以上的基础研究下提出三因子模型，该模型在市场溢价因子的基础上添加 SMB 与 HML 两种因子对市值效应作出了较好的解释。

Lakonishok(1994)等人后续发现股票未来收益率可以由股票当前价格、所有者权益、账面价值及每股收益进行更好的解释。至此资本资产定价模型由最初的单因子模型开始正式发展成为多因子模型。

Jegadeesh（1995）等发现上期股票与现期股票的收益率存在一定关联，继而开始提出构建买入过去投资收益较好与投资收益较差的企业，从中获取较好的超额收益，这也就是其开展的动量策略。

基于以上研究，Porras D M(1998)以日本股票市场作为分析蓝本，通过股票市场相关横截面收益情况，以三因子模型代入分析，从而验证该模型对股票市场具有有效的适用性。

Chui（1998）同期将三因子模型代入整个亚洲股票市场进行验证，在验证过程中发现，账面市值与市场价值的比率和企业规模有较大联系，而股票平均收益率与市场系数指标的关联性则较小，因此 Fama-French 模型在较大的股票市场环境下同样具备适用性。Vernon J A（2010）以 CAPM 模型与 Fama-French 三因子

模型进行综合运用，发现无论是那种模型的改进，均需要对价值加权因素与众多风险因素进行分解，才能检验模型分析结果是否能够解释企业的最终收益，但模型的使用过程仍较复杂。

Brennan M J (2015) 基于跨期资本资产定价发现 Fama-French 三因子模型的优化,将对超额收益带来更好的解释。随后 Mahnoor S(2017)开始通过 Fama-French 三因素模型代入对达卡证券交易的超额收益进行分析，证实了其优化后 Fama-French 三因子模型的价值。

Drew Naughton 和 Veeraraghavan(2003)在通过我国沪 A 市场进行调研验证模型适用性时发现，超额收益可以通过市场因子和市场规模因子综合得出，但此时账面价值与市场价值比值却不明显。所以 Titman(2004)为了验证模型下所形成的投资动量策略是否会带来良好的超额收益时，发现不同超额收益的差异，可以通过动量因子进行一定解释。

Griffin(2005)继而在三因子模型下加入动量因子形成四因子模型进行研究，并对多个大型交易所深入调研，发现该模型下的动量因子效应可以解释股票市场细小变化，但与宏观经济关联性并不大，但在稳定的经济周期下，投资收益也能一定程度上获得较为稳定的保障。

Lam(2010)在对香港股市进行研究时发现，股市中存在动量效应，这一点与 Gregory (2013) 在英国股市中的研究结果相同，两次研究结果表明四因子模型中动量效应显著存在。Fama French (2013) 基于上述研究成果，对三因子模型进行修改，添加盈利因子和投资因子，而形成五因子模型。该模型的最终成型也完善地解释了超额收益与模型拟合优度。

Ji Z ,Chang V (2020) 将投资者情绪作为一个新的自变量引入 Fama-French 三因素模型 (FFTFM) 中，以增强 FFTFM 的解释力。这两个模型之间的比较表明，情绪因子可以提高模型的总体解释力。Datta S (2020) 结合印度市场的数据，重新评估和比较了 CAPM、Fama 和法国三因素模型在印度环境下的绩效。这项研究显示，在小型行业投资组合中，有八个获得了统计上显著的月平均超额回报。Shaabani J (2020) 指出 Fama-French 模型被广泛用于评估投资组合相对于市场回报的表现。在 Fama-French 模型中，所有因素都是时间序列数据。横截面数据与时间序列数据略有不同。时间序列回归的一个明显问题是，时间序列回归中的 R^2 平方通常非常高，尤其是与横截面数据的典型 R^2 平方相比。 R^2 平方的高值可能会导致误解，即回归模型很好地拟合观测数据，自变量很好地解释了因变量的方差。因此，其在研究中，以两种不同的方式重新审视和验证了 Fama-French 模型，对这一模型的应用提出建议，即 Fama-French 模型应考虑重尾分布，能够提高对股

票数据的解释能力。

1.3.2 国内研究现状

国外研究下的资产定价理论历经半个世纪，最终形成 Fama-French 五因子模型。在对我国关于 Fama-French 五因子模型的研究中，多数研究仅存在证券市场模型适用性的检验。研究中以三因子模型与四因子模型的使用较多。例如在两种模型中加入我国市场环境的定价因子对模型进行改良。但由于我国股票市场发展历史原因，及我国证券交易市场特色，我国证券市场与外部证券市场环境仍存在较大的不同，因此在众多研究中可以发现均对其模型进行了优化，并得出适用于我国市场的研究成果。

陈浪南和屈文洲(2000)利用 A 股市场五年数据，研究证实了 CAPM 模型中的 β 值和超额收益率在本土环境不具备关联性，尽管 β 值在模型中有较突出的表现。黄兴旺等人（2002）按照 Fama-French 模型的因子组成法，在我国市场中进行验证，并未有发现足够的价值。范龙振等人（2002）年在同期研究时，表示模型下呈现的账面效应和市值效应的关联性是存在的，其比值变化较为显著，但不能够通过简单的 CAPM 模型对其作出完整解释。

张祥建等人（2003）在进行模型实证的过程中，将上海证券交易所的所有发行股作为数据基础，研究中将流通市值作为规模因子代入模型，分析结果表明以上海证券交易所为例的我国股票市场公司效应明显。

刘蕾等人（2004）研究时，以上海证券交易所股票按照规模因子、账面价值与市场价值比值代入五年的股票发展中分组对比分析，研究结果表明我国股票证券市场规模效应显著，账面价值与市场价值比值在模型中的有效性能得到检验。

刘沙（2012）将非线性资产定价模型与线性因子模型代入住房消费的研究，结论表示模型在股票溢价、规模溢价等的分析结果好于单因子模型、三因子模型，并且模型分析过程中能够任意代入分析因子。

于是为了验证哪种模型在我国的适用性更高，李泉等人（2015）以 CAPM 模型与 Fama-French 模型同时开展对上证 A 股分析，对比发现 Fama-French 在我国适用性更高。因此其得出结论，金融市场的发展依赖于实体经济，金融市场资本体系发展与体系的完善，需要建立在科学的投资策略下，所以利用更优的模型进行分析，能够有效地实现投资行为长久化。

王源昌等人（2010）年在对规模因子研究时发现，在模型分析结论解释我国股市横截面收益率的表现中，市盈率因子有着更好的效果。刘维奇（2011）以三因子模型代入分析我国股市改革前后两个时期的各项特征，均发现在改革前后期

的分析中,模型对股票市场都有着较高的适用性,股市发展解释与分析结果有着较高程度的贴合。

由以上研究可发现,在利用模型分析解释我国股票市场的发展中,学者在不断通过对模型进行优化,一方面验证模型在我国市场中的适用性、准确性,同时也验证模型使用不能依靠拿来主义,结合我国市场环境的优化才能够对市场各项指标做出合理解释。田利辉(2014)由中美股市 10 年交易数据进行对比分析,发现风险溢价因子更符合我国的股市环境,在我国股市预测之中的准确性较高。李倩等人(2015)进而在对沪 A 股票交易数据分析中,以三因子模型代入分析股票发展周期,较为显著地发现三因子模型对股票衰退期的分析结果贴近程度更高,但账面市值比因子作用不突出。

周骆等人(2016)所进行的实证分析中,表示三因子模型需要进一步优化,才能在我国股票市场环境下对股票横截面收益率有着较好的解释。实证分析中同时指出不同市场背景下的分析结果解释程度不一定相似,这是市场不同所决定的。所以实证分析模型需要结合当地市场背景进行优化,并不需要保持一成不变。

高春亭(2016)同期以 Fama-French 五因子模型对我国 A 股市场进行分析,分析结果对我国证券市场流动性定价等方面做了较高程度的解释。与此同时表明,在优化模型的基础上,五因子模型相对三因子模型具有较高的优势。

欧阳志刚等人(2016)利用沪 A 股月数据,并在三因子模型下加入股票动量因子,证实基于本地市场构建的四因子模型较三因子模型对股票市场解释程度更高。最后宋光辉等人(2017)年在研究中,基于四因子模型添加流动性因子进行优化,实证结果表明流动性因子的加入可以帮助量化模型对市场解释带来提升。

齐岳,周艺丹(2020)基于 Fama-French 三因子模型,创新性地将公司治理水平作为溢价因子考虑到资本资产定价模型当中进行实证分析,对传统的 Fama-French 模型进行了扩展,旨在基于扩展的资本资产定价模型探讨我国上市公司治理水平是否会对股票资产定价产生影响,具有一定的理论意义。付巍巍,张立卫(2020)选取了中国市场 A 股中 100 支股票的相关数据,采用资产组合分类的方法进行了数据处理,计算出市场因子,市值因子(SMB),市值比因子(HML)后进行多元线性回归,并对模型的拟合度和 P 值进行分析,完成 FF 三因子模型在中国股市的数据实证。施昊天(2022)选取上证主板 1618 家上市公司 2015 年 6 月至 2021 年 5 月,共 72 个月的月度收益率为研究对象,验证了 Fama-French 三因子模型在我国股票市场的适用性。

综上所述历经半世纪的发展,Fama-French 因子模型由最初功能单一的 CAPM 模型发展至今,在发展历程中逐渐完善,并给予优化以适用于各个金融市场环境。

而资产定价理论也已是经济、金融学中的热门研究领域。不断地演变与发展让资产定价理论体系不断完善、丰富，基于国外学者的开创与优化下，我国学者的后继研究也有了更为充分的研究理论依据。截至目前，在国内外研究环境中，资产定价理论的有效性已得到充分验证。尽管我国股票市场发展与成熟股票市场体系上尚存差距，但资产定价理论在我国学者研究下，依然证实适用于我国股票市场。Fama-French 三因子模型国内外的理论研究成果也为本文研究带来了坚实的研究基础。

1.3.3 研究评述

综上所述，我国股票市场尚且不够完善，我国股票市场仍旧存在资产定价失真现象，这也是由于在市场投资行为下，尚缺利用资产定价理论而形成有效、合理的投资策略。为获得较好的投资回报，稳定超额收益率，多因子模型在我国股票金融市场的使用仍需积极应用。因此结合前文对 Fama-French 模型的发展及成果阐述，这一模型应该在我国资本投资市场中发挥更好的作用。虽然当前国内外的专家学者对于金融领域的量化投资分析都有了较为系统的研究，无论是量化投资分析的理论基础，还是量化投资的理论模型都有了较为系统的研究。但对于相关量化投资模型在中国 A 股市场的适应性并未做系统的研究，同时也未结合我国 A 股市场对 Fama-French 三因素模型等进一步进行改进研究、改进后的模型适应性研究以及根据改进模型提出相应的投资建议等。因此，本文将在此背景下展开研究，希望能够在 Fama-French 三因素模型的基础上引入大数据因子，并分析优化后模型在 A 股市场中的适应性，据此提出具体的股票投资策略。

1.4 研究主要内容及框架

1.4.1 研究内容

本文研究首先对量化投资策略的研究背景进行描述，并结合国内外对量化投资理论、多因子模型的发展及研究成果，得出本文的研究思路。对本文涉及的投资理论、资产定价理论等进行综述，为后文的研究提供相应的理论基础。随后以 Fama-French 模型作为基础，融合估值、资本结构因子、成长因子、大数据因子等，代入分析并建立回归模型，为本文实证提供有效的分析成果支持。最后本文对量化投资策略的研究以量化投资方案的具体形成、结果运用、实证下各因子的有效性进行评价，研究加入了大数据因子后的改进模型的适应性，并应用这一模型模拟投资决策，为投资者的决策提供基础。

1.4.2 研究框架

本文研究框架下如下图 1-1 具体所示。本文的研究框架主要包括绪论、相关理论综述、因子组合量化投资策略、实证分析、研究结果等五个方面的内容。绪论部分主要阐述了本文的研究背景及意义、研究内容与研究框架；理论综述对现代资产配置、资产定价、Fama-French 三因素模型进行介绍；因子组合量化投资策略中包括因子组合量化投资策略与因子组合量化投资策略的流程，并对大数据因子的选取原因、大数据因子的概念、引入大数据因子的意义等进行分析；随后实证分析股票数据、研究本文构建的引入了大数据因子的 Fama-French 三因素模型在沪深 A 股的适应性，并基于此应用模型，测算模型的效果并实施改进，最终给出相应的投资建议、以及本文的研究结论。

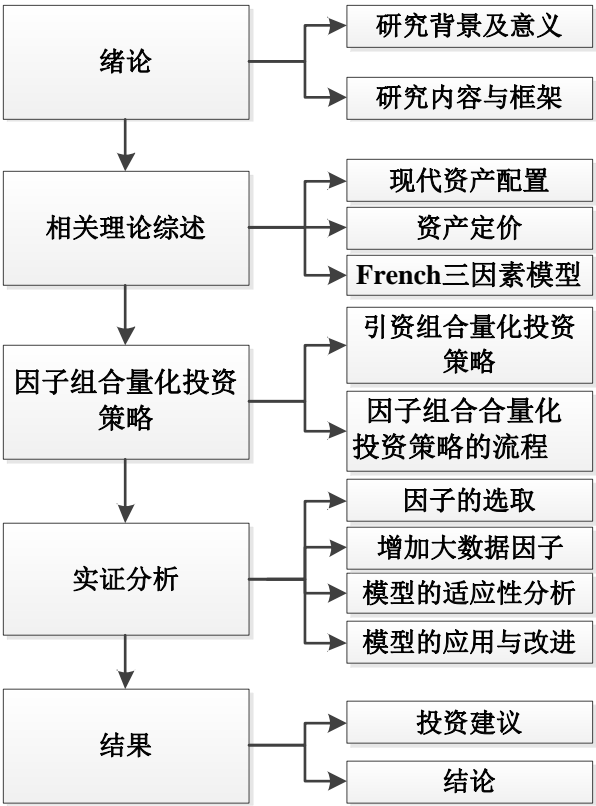


图 1-3 本文的研究框架图

第二章 因子组合量化投资理论基础

本章主要对量化分析相关的理论基础，如现代资产配置理论、资本资产定价模型、Fama-French 三因素模型、套利定价理论（APT）等进行研究与分析，分析各分析理论及模型的概念、特点等，为本文的研究提供相应的理论基础与分析工具。

2.1 现代资产配置理论

现代资产配置理论中，为了获取更大收益，最大程度降低风险，所以在金融投资中将资金进行拆分进行多种产品组合的多元化投资。其中心思想由 Markowitz（1952）提出，在实现合理配置的这一操作过程中，该过程也被称为均值-方差分析，成为金融经济学中最重要的几轮理论之一。理论假设每个人的财富总额用函数 U 表示，那 U 的具体值则代表个人财富将带来的具体效用，对不同的人来说其“效用”是有所差别的，排除非数学性因素，如果不考虑个体之间存在的差别，那么其实大部分的投资者所代表的“效用函数”一般都会具备以下两个方面的特点：

（1）假设 $x \leq y$ ，则 $u(x) \leq u(y)$

（2）假设 $d \geq 0$ ，同时 $x \leq y$ ，则 $u(x+d) - u(x) \geq u(y+d) - u(y)$

得出的效用函数在坐标轴呈抛物线状，在横轴“资产”不断增加的条件下，Y 轴“效用”将会越来越高，但最终会趋于平衡。

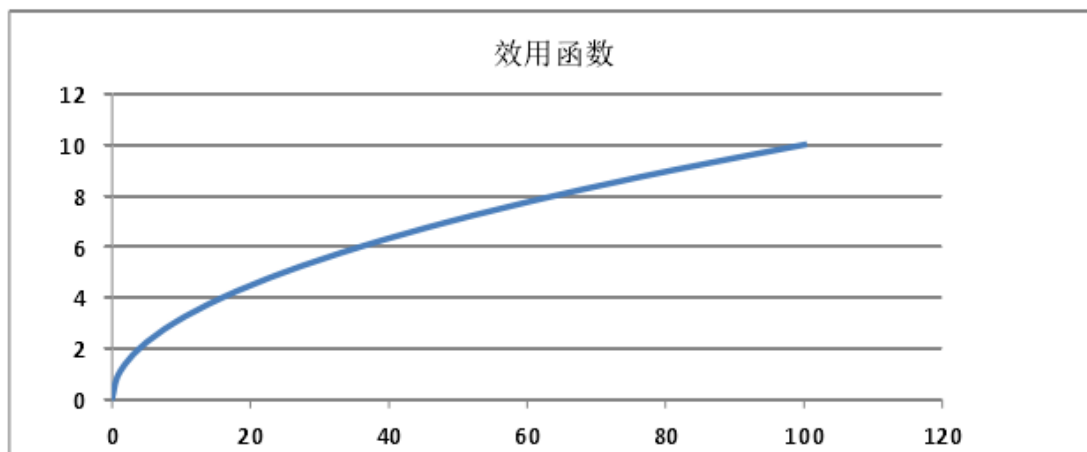


图 2-1 效用函数

在现代资产配置理论之中，对个人投资者的“效用”用 U 来进行表示，而个人投资者在实际的投资活动之中，往往会面临不同的投资选项，这些 n 个投资选项最

终的财富值结果用 $\{X_1, X_2, \dots, X_n\}$ 来表示。因此对于任何一个理性投资者来说，最终投资者肯定会选择 $E[u(x)]$ 中数值最大的来作为投资选项。

由于在效用函数中，效用增幅边际递减，因此在投资者在某一点产生损失时，则相应产生效用的损失，但同时在获取相同效益时，此时投资者获得效用远低于损失，这就是投资对风险产生厌恶的根本原因。

所以在 MPT 模拟过程中，假设投资者对于资产收益与收益的认知存在一个标准误差值。则可以将不同的多个资产或单个资产进行组合呈现在坐标上以反映其收益和风险关系。图中横轴表示收益率的标准差，纵轴表示预期收益率。

而投资者通常会追求风险低、收益高的资产配置方案，这是投资者的基本特征。所以假设在资本收益率固定且不会发生变化下，在最大程度降低风险下，在坐标上表示便为横轴值趋于最小；同样假设资产风险固定不发生变化，那么在最大程度下将收益最大化，反馈在图表上即相当于获取在 Y 轴上的最大值。

在资产配置中，金融资产一般会以 $1, 2, \dots, n$ 进行表示。假设资产用 i 来表示，则 r_i 代表的就是资产 i 的收益率的变动情况， $E[r_i]$ 则表示这一资产可能的预期收益率等， σ_i 表示 r_i 的标准差。综合上述规定，市场收益率方差如果 > 0 则可以将之定义为存在风险的资产，且只要整体收益之中每个收益率都能够确定，就可以将这一资产视为无风险资产，无风险资产的利率为 r_f 。

风险资产配置是某一个存在风险收益的资产组成的投资组合，同时结合现代资产配置理论会对资产的权重进行分配， ω_i 表示一项风险资产在总资产投资组合中所占的比例，且需要满足 $\sum_{i=1}^n \omega_i = 1$ 。

资产组合收益率随机变量为 $r_p = \sum_{i=1}^n \omega_i r_i$ ，预期收益为：

$$E[r_p] = E\left[\sum_{i=1}^n \omega_i r_i\right] = \sum_{i=1}^n \omega_i E[r_i]$$

$$\text{方差计算为 } \text{Var}(r_p) = E[r_p - E[r_p]]^2 = \sum_{i=1}^n \sum_{j=1}^n \omega_i \omega_j \text{Cov}(r_i, r_j)$$

标准差为

$$\sigma_p = \sqrt{\text{Var}(r_p)}$$

此时当预期收益固定，在坐标上找到标准差最小一点，用拉格朗日乘法得出。在最小值上找到对应资产配置标准差最小位置。因为任何特定的 μ 都有对应点，且需要满足：

$$E(r_p) = \mu$$

所以该点还会有一个最小方差值，用图形表示为一条抛物线，如下图所示：

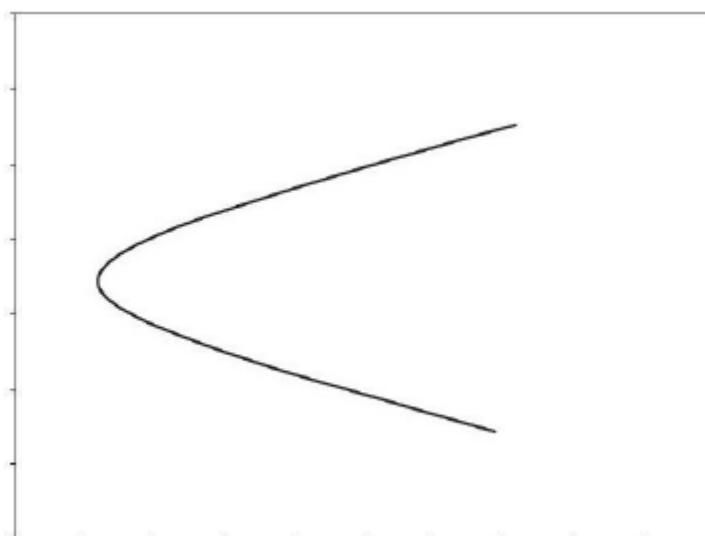


图 2-2 有效前沿

该曲线叫做有效前沿（马科维茨子弹），在图中当曲线上某一点所表示的波动率在最小位置时，则表示曲线中最靠左侧的部分是波动有效位置。当投资者在固定风险与最大预期收益下，选择最大预期收益，则需要排除有效区域的下半区。当在无风险资产配置进方案时，便可以在前沿左侧弥补一部分。

所以假设某 P （资产配置），当 $\alpha \geq 0$ 时， P 配置中的本金 α ，在无风险资产配置中为 $1 - \alpha$ 。当假如 $\alpha \leq 1$ 就存入银行 $1 - \alpha$ 倍的本金，或者购买 $1 - \alpha$ 倍的本金债券，以求无风险资产投资回报。若 $\alpha \geq 1$ 时，则 $\alpha - 1$ 倍的资本，用来作为无风险资产的利率，并会在 P 中对融资资助金以及本金进行配置。因此将 α 作为新的资产投资组合收益的随机变量用 r_α 计算，得出

$$E[r_\alpha] = E[\alpha r_p + (1 - \alpha)r_f] = \alpha E[r_p] + (1 - \alpha)r_f$$

$$\sigma_\alpha = \sqrt{\text{Var}(\alpha r_p) + (1 - \alpha)r_f} = \sqrt{\alpha^2 \text{Var}(r_p)} = \alpha \sigma_p$$

此时

σ_α 与 $[r_\alpha]$ 的关系呈函数形式，将所有 $\alpha \geq 0$ 时出现的点在图形上表示出，可以得到穿过 $(0, r_f)$ 和 $(\sigma_p, E[r_p])$ 这两个点的整条射线，如下图所示：

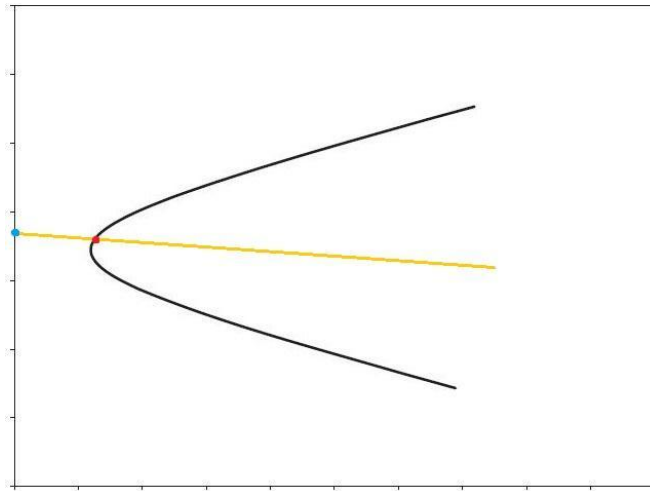


图 2-3 穿过有效前沿的射线

其中蓝色表示无风险投资部分利率，P 为红色，P 所表示在有效前沿上某一点时的资产配置方案。所以黄线表示在连接黄色与红色后，资产配置方案的数据。黄线的斜率为夏普比率，用 $S_p = \frac{E[r_p] - r_f}{\sigma_p - \sigma_{r_f}} = \frac{E[r_p] - r_f}{\sigma_p}$ 得出。

如图 2-4 所示：

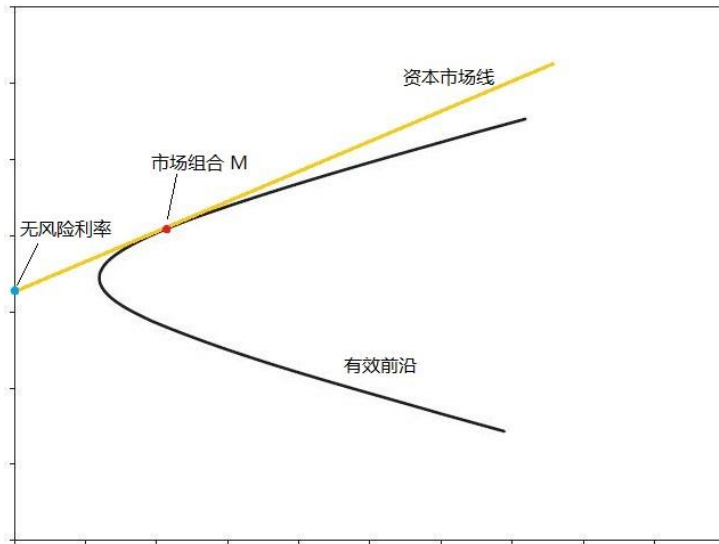


图 2-4 资本市场线上的市场组合

图 2-4 中经无风险利率与市场组合点的射线叫做资本市场线，用 $\mu = r_f + S_M \times \sigma$ 表示。

2.2 资本资产定价模型（CAPM）

Treynor、Sharpe 等人（1964）提出资本资产定价模型（CAPM）。模型建立

在 Markowitz 的资产配置理论之上。在其提出的理论体系之中，针对投资市场之中投资者对风险与收益的评估，是来自于投资者在市场之中存在信息不对称，同时投资者与被投资公司、市场之间都处于完全公开的状态。而在这一情况下投资者最终能够取得较大的收益。投资者为了对于其收益进行最大化，应当在投资者市场上进行合理化的资产配置，这一配置之中包括其现金的持有量、以及投资市场组合 M 。 M 也即代表这一投资者面临的市场风险。

在 CAPM 的框架内，有针对投资风险 S 的计算公式：

$$E[r_s] = r_f + \beta_s \cdot (E[r_M] - r_f)$$

其中 r_s 是组合 S 的收益变量、 r_M 表示市场组合收益、 r_f 是同期市场无风险利率、 β_s 是投资组合对于市场风险的敏感情况。

β_s 准确值为 $\beta_s = \frac{Cov(r_s, r_M)}{Var(r_M)}$ 。 r_f 表示资产的时间价值（无风险利率收益），

$\beta_s \cdot (E[r_M] - r_f)$ 为资产风险收益（对投资者所承担的风险补偿）； $E[r_M] - r_f$ 为市场组合收益， β_s 是 S 对系统性风险的敏感系数。

所以此时当某一股票带来的风险、溢价和市场风险所带来的溢价构成斜率为 β ，截距 r_f 的线性关系。但在实际情况下，股票和市场中股票收益不完全相同，所以需要重新考量。

以 α 表示的超额收益， $\alpha = (E[r_s] - r_f) - \beta_s \cdot (E[r_M] - r_f)$ 。按照 CAPM 模型，假设某股票价格正常时， α 的值为 0；当 $\alpha < 0$ 时，则股价被低估；当 $\alpha > 0$ 时，股价高估。所以在股票投资中，最佳的投资策略是在 α 低的时候买进，在 $\alpha = 0$ 时卖出。

所以经过 CAPM 推算， S 夏普比率和市场组合 M 的夏普比率关系如下所示：

$$\begin{aligned} \text{Sharpe}(S) &= \frac{E[r_s] - r_f}{\sigma_s} \\ &= \frac{\beta_s \cdot (E[r_M] - r_f)}{\sigma_s} \\ &= \frac{Cov(r_M, r_s)}{\sigma_M \sigma_s} \cdot \frac{E[r_M] - r_f}{\sigma_M} \\ &= \text{Corr}(r_M, r_s) \cdot \text{Sharpe}(M). \end{aligned}$$

S 与 M 的相关性越强， S 的夏普比率越大，而这一比例如大的结果就表示风险与收益是成正比的，给与投资者相应的参考。

2.3 Fama-French 三因素模型

Fama、French 在 1992 年发现市值较小的企业其规模通常较小，但这类公司所面对的市场往往风险更大，因此这类企业难以保证较高的收益，也不适合作为投资目标；而账面市值较高的企业有可能存在市值被高估的情况，也需要对其过高的估值进行弥补，以避免投资人在高位买入。于是 Fama、French 综合考虑企业所有者权益市值因素、企业所有制权益账面价值与市值比等，对 CAPM 模型进行改进，引入了 SMB、HML 这两个因素，组成了三因素模型，其表达式如下：

$$R_i = a_i + b_i R_M + s_i E(SMB) + h_i E(HML) + \varepsilon_i$$

其中 $R_i = E(r_M - r_f)$ ，指的是股票 i 比起无风险投资的期望超额收益率。 $R_M = E(r_M - r_f)$ ，E(SMB) 是小市值公司相对大市值公司的期望超额收益率，E(HML) 则是高账面市值比公司的股票比起低账面市值比公司的股票的期望超额收益率，而 ε_i 是回归残差项。

PEG 是一种对未来股价的期望值，又称为预测值，PEG 的计算公式如下：

$$PEG = \frac{PE}{G \cdot 100}$$

PE 表示市盈率，是当前股价 (P) 相对每股收益 (EPS) 的比值。表达为：

$$PE = \frac{P}{EPS}$$

而 EPS 表示每股收益，其公式如下：

$$EPS = \frac{\text{归属于普通股股东的当期净利润}}{\text{当期实际发行的普通股的加权平均数}}$$

G 则为企业的收益增长率，其公式如下：

$$G = \frac{EPS(\text{this year}) - EPS(\text{last year})}{EPS(\text{last year})}$$

2.4 套利定价理论 (APT)

CAPM 由于在假设环境下针对较多的现实环境并不适用，所以史蒂芬·罗斯 (1976) 基于此提出套利定价理论 (APT) 以替代 CAPM。APT 在理论层面中表示，即便在均衡市场，套利可能性也极低。两种证券的收益风险相同，其定价也相同。并在完全交易市场的理想状态下，不需要考虑交易成本。当市场证券数量 N 远高于影响因素 n 时，则资本市场任何资产的收益都为 N 构成。那么表示为

$$R_i = E[R_i] + b_{i1} F_1 + b_{i2} F_2 + b_{i3} F_3 + \cdots + b_{in} F_n + \varepsilon_i \quad i=1,2,\dots,n$$

此时 F_i 表示资产收益第 i 个系统风险, b_{ij} 代表因子负载, 表示资产 i 收益由第 j 个公共风险因子印象, ER_i 表示所有相关公共风险因子全部为 0 的时, 资产 i 的预期利润; ε_i 作为资产 i 的偶发误差, 即是指非系统风险。所以在高度自由竞争的资本市场中, 套利时机的出现会引发众多投资者跟随。当套利时机消失, 市场再度回复平衡。因此结合线性回归, APT 标准表达式为:

$$ER_i = \lambda_0 + \lambda_1 b_{i1} + \lambda_2 b_{i2} + \lambda_3 b_{i3} + \cdots + \lambda_n b_{in}。$$

其中 λ_0 表示不受到所有公共风险因子影响的资产收益率 ER_z 。假设一种无风险资产的话出现, λ_0 就等同于该资产的收益率 R_f ; λ_j 是关于第 j 个风险因子的相关风险溢价。

同样假设当资产无风险时, $\delta_j = 0$, 则代表资产组合不受到任何风险影响。但当面对关键 j 风险, 且敏感程度为 1 时, 则投资者期望回收率为 $\lambda_j = \delta_j - R_f$ 。因此代入标准式, APT 也可表达为:

$$ER_i - R_f = (\delta_1 - R_f) b_{i1} + (\delta_2 - R_f) b_{i2} + \cdots + (\delta_n - R_f) b_{in}$$

当公共风险因子不相关时, APT 便为:

$$ER_i - R_f = (\delta_1 - R_f) b_{i1} + (\delta_2 - R_f) b_{i2} + \cdots + (\delta_n - R_f) b_{in}$$

由此可见, 在 b_{in} 的形式与 CAPM 中的 β_i 的定义完全符合。

2.5 本章小结

本章主要对量化投资的理论基础进行了分析, 包括现代资产配置理论、资本资产定价模型、量化分析三因素模型、套利定价理论等, 为本文的研究提供了相应的理论基础。根据上述的分析可以看出, 在众多的量化投资, 以及相应的理论模型中, 大部分的模型主要是考察市场中股票的基本信息情况及市场的变动情况, 而对市场中的投资主体“投资者”的投资情绪和投资者本身的考察较少, 这并不符合当前的投资环境, 也使得相应的理论体系存在局限性。而在上述的模型之中, Fama-French 三因素模型是应用最为广泛且有众多的研究成果的模型, 选择这一模型有较为丰富的研究成果可供参考。因此, 本文最终选择在 Fama-French 三因素模型之中加入新浪大数据 100 因子作为补充, 对三因素模型进行优化, 能够弥补这一模型对投资者情绪观察以及对市场短期变化情况的考察, 因此本文针对 Fama-French 三因素模型引入了大数据因子。

第三章 因子组合量化投资模型建模过程

3.1 因子组合量化投资策略

量化投资策略是借助计算机程序计算，依靠数量化统计分析工具模型，对投资行为、投资收益进行分析，进而形成投资策略的投资方式。量化投资不仅仅依靠纯粹的数据及投资信息而形成，量化投资需要包括多数的市场因素、分析要素及分析工具才能够完成。所以量化投资是一种主动式投资行为，从前文可知量化投资行为更适用于主动投资空间大、发展时间较长、市场理念较为成熟的环境。丁鹏在《量化投资-策略与技术》中将量化投资策略分为判断趋势与判断波动率两种。前者适用于二级市场，能够对大盘走势及单股发展情况进行较好分析，通常这种策略运用风险程度较大，但考虑到资本的逐利性，因此对应收益率也随之增加。因此也可在下分为量化选股与量化择时两种策略。而后者的本质在于降低消除整体性大风险，因此策略相对保守，收益率相对较低，但能够稳健获得投资收益。这一种策略在衍生品套利中运用较多。不过我国衍生品市场发展当前并不成熟，量化投资更多集中在逐利性投资中。

投资行为中包括三个循环过程，即投资开发、执行与评估。投资行为无论是公募还是私募基金，都需要严格按照这个流程执行。所以投资开发中需要投资人根据个人价值偏好选取投资方向，对投资策略进行创建，本文所阐述的多因子量化投资策略也符合这一流程。在投资执行阶段（投资交易阶段），量化投资是在投资交易阶段的创新，属于高频交易，在该阶段内投资人可以通过快速对证券买入卖出获得短期套利而不用付出明显代价。最后投资评估中，就是将投资过程与投资收益综合研判，为下一阶段投资开发进行准备。所以这也就是投资行为的循环过程，该过程的良性循环，是投资策略核心竞争力的重要构成原因，同样也能在多个层面反馈出投资者良性的投资回报。

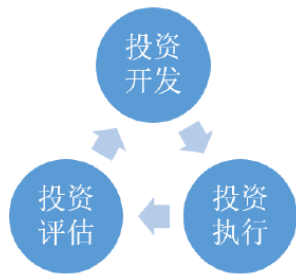


图 3-1 投资的三个过程

前文所提及的主动式投资行为的量化策略，该策略在特别是当下的智能信息化时代，更需要注入足够的计算机技术、策略创意等必须要素，才能产生足够的市场竞争力，也就有在众多投资公司中保持优势的资本，这同样就对投资回报做出了前情保证。例如在多因子量化策略下，各个企业可以用自身的数据因子代入市场进行投产分析，在保证分析结果合理的情况下，也就一定程度对企业在市场中的投资预期回报提供了保证。所以量化投资策略，也成为了当下投资公司的核心竞争策略。

并且因为量化投资先天性所具备的规模效应，伴随着投资规模的加大，投资硬件设施技术使用的增加，较大降低了企业人工的边际成本，边际成本下降，企业进一步较少资金压力，可以有更多的精力专注于投资规模增加。如下图 3-2 美国智能化量化平台管理规模所示，在 2017 年中投资平台的智能化量化程度更高。2017 年同期对比 2016 年，能够发现规模增长 8%，整体平台管理规模达到 183 亿美元，全球市场占比 23%。

以国内为例，国内前十大公募基金的管理规模不超过 50%。本文认为，投资行业本身是一个规模效应很强的行业，但最后的集中度仍然保持较低的市场格局，是因为投资行业天生是较为分散的行业。原因主要是规模过高虽然人力边际成本会减小，但是交易成本可能过大，对市场会造成较大的波动，对自身产品的收益率也会造成影响；另一方面规模过大可能导致政策监管的更加关注，造成更多的限制。所以本文认为量化投资行业本身虽然具有一定的规模效应，但是天生是一个较为分散的行业。

所以本文分析认为，投行是一个规模效应极其强烈的行业，因为投行投资领域众多且分散，所以在高度集中化管理下，投资集中程度便可以获得较大程度提升。但仍然小于市场整体。尽管人力边际成本下降着实会带来好处，但过高的资本涌入市场，会对市场产生波动，受到市场波动，任一产品的收益率都将呈不同变化状态。

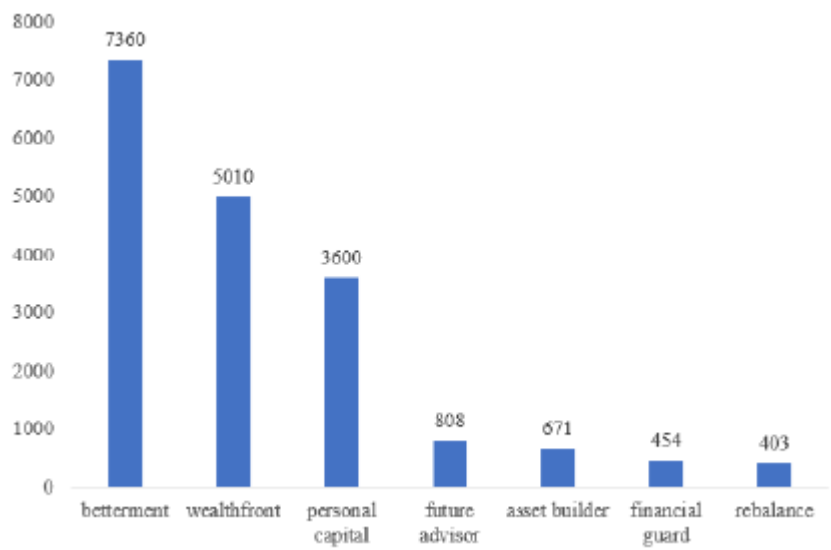


图 3-2 美国智能化量化平台管理规模

所以经过上述分析，本文研究认为量化投资行为所带来的差异化结果，完全集中在其核心竞争力的构建。尽管投行先天性的分散，但基于国内外情况，依旧规模效应强烈。所以本文在此基础上，也将提出未来量化投资行业的发展预估。

3.2 大数据因子

3.2.1 大数据因子基本情况

加入模型的大数据 100 指数因子是新浪财经、南方基金与证券信息公司三家公司结合中国国情联合编制，中国首个大数据系列指标，是对样本空间股票，根据市场驱动因素、财务因素及大数据因素得分进行模型评分，取综合评分最高的前 100 只股票构成大数据 100 指标的初始样本股。市场驱动因素涵盖了最新一个月股票的换手率、价格波动率、流动性等因素，通过量化因子模型计算得出市场驱动因素得分；财务因素涵盖最新的净资产收益率、年度营业收入、市盈率的同比和环比增长率等因素，根据模型计算主营业务收入以及净利润同比和环比预测结果相对前一期该指标幅度的变化情况作为业绩得分等，最终得出财务因素综合得分；大数据因素涵盖了单个股票搜索热度得分、由财经频道里正负面新闻影响股价得分、由微博上正负面文章影响股价得分等，综合优化得出大数据因素得分。最终进行权重计算与调整，再根据历史回测优化结果来确定指标成分股，最终得出相应的大数据指数。

3.2.2 大数据指数的特点

(1) 计算大数据指数的股票选择

本文加入到 Fama-French 模型内的“大数据因子”是指“新浪大数据 100”，这是由新浪财经与南方基金、证券公司等共同联合编制的因子。这一因子也是国内首个大数据指标、是基于市场的具体情况、财务因素以及筛选到大数据指标内的股票表现情况等公共决定并评分的。根据大数据指数的具体可以看出大数据 100 是综合对市场驱动因素、财务因素以及传统的技术基本面考察后，综合市场的信息因素等最终筛选并进行计算的。与传统的 K 线、基本面考察不同，大数据 100 不注重单一指标、而是对所有纳入选择的 100 只股票同等重视，而大数据 100 每月会进行重新筛选与计算，这一点充分利用了大数据技术海量数据计算的优点、同时能够更为有效的捕捉市场的波动情况。

(2) 纳入大数据指数计算的股票选择

需要注意的事，大数据指数并非是对单一股票进行研究，也并非对单一股票进行考察，大数据指数是在筛选了一段时间内表现较好的股票数据后、对选中的股票数据重新展开分析，并判断投资者的投资行为、投资者情绪等等，并最终计算出大数据指数。因此，最终获取到的大数据指数实际上是不包含任何一只股票的，只是在每个交易日内对市场预期的一种判断，是作为投资者投资情绪的判断依据。但“新浪大数据 100”指数每期发布其实关注的是在 15 天-30 天周期内在计分模型下表现较好的股票，因此可以看出其观察的周期较短、主要是对短期内投资市场、投资者情绪的预测，而并非对整体未来股票市场发展趋势的预测。因此在 FAMA-French 三因素模型之中引入这一因素是对 FAMA-French 三因素模型过于关注市场本身、而忽视投资者情绪、以及忽略了短期市场的快速变化的重要补充。

(3) 个股纳入到大数据指数中对最终的模型计算并无影响

此外，大数据因子计算过程之中选择的股票是在每一期不断发生变化的，纳入到大数据因子的计算过程中最终在 FAMA-French 三因子模型之中的 SMB、HML 等因素的分析无任何影响。从定义上来看 SMB_t 表示小规模股票构成的投资组合之中与同期大规模股票的投资组合在收益率上的差值。 HML_t 表示高低账面市值比股票构成的相应投资组合在同一时期的收益率差值。而这两者与大数据因子没有任何联系，也不会因为股票被纳入到大数据因子之中作为计算成分，就会提高 SMB 或者 HML 的得分。而在本文优化后的 FAMA-French 三因子模型中，新增的大数据 100 因子仅是一个指数结果，结合本文研究的 80 期观测期来看，在理想状态下一共会选中 8000 只股票（100 只股票*80 期=8000 只股票），而当前沪深 A 股仅

有 4000 多只股票，从这一点也可以看出虽然“新浪大数据 100”的最终指数结果可能与单个股票存在联系，但这一联系分布在众多股票之中，只要股票在短期内表现较好就会被纳入到大数据指数之中，这一影响广泛存在、并非个例，因此可以排除个股本身已经存在于大数据成分对最终计算结果的影响。

（4）截止 2022 年 1 月的最新大数据指数成分

综合从大数据 100 的历史数据中来看，其自 2017 年出现以来的历史收益率都是比较高的，投资回报率较高。根据 2022 年 1 月 13 日最新的数据显示，其中的大数据指数的主要成分，包括 000035、000039.....002101、002109 等在内的 100 只股票，具体如下表 3-1 所示：

表 3-1 大数据 100 指数的最新指数成分

| 品种代码 | 品种名称 | 纳入日期 | 品种代码 | 品种名称 | 纳入日期 |
|--------|------|------------|--------|------|------------|
| 000035 | 中国天楹 | 2022-01-04 | 002101 | 广东鸿图 | 2022-01-04 |
| 000039 | 中集集团 | 2022-01-04 | 002109 | 兴化股份 | 2022-01-04 |
| 000055 | 方大集团 | 2022-01-04 | 002110 | 三钢闽光 | 2022-01-04 |
| 000061 | 农产品 | 2022-01-04 | 002131 | 利欧股份 | 2022-01-04 |
| 000065 | 北方国际 | 2022-01-04 | 002135 | 东南网架 | 2022-01-04 |
| 000157 | 中联重科 | 2022-01-04 | 002146 | 荣盛发展 | 2022-01-04 |
| 000338 | 潍柴动力 | 2022-01-04 | 002160 | 常铝股份 | 2022-01-04 |
| 000401 | 冀东水泥 | 2022-01-04 | 002169 | 智光电气 | 2022-01-04 |
| 000402 | 金融街 | 2022-01-04 | 002202 | 金风科技 | 2022-01-04 |
| 000411 | 英特集团 | 2022-01-04 | 002255 | 海陆重工 | 2022-01-04 |
| 000425 | 徐工机械 | 2022-01-04 | 002262 | 恩华药业 | 2022-01-04 |
| 000488 | 晨鸣纸业 | 2022-01-04 | 002345 | 潮宏基 | 2022-01-04 |
| 000507 | 珠海港 | 2022-01-04 | 002357 | 富临运业 | 2022-01-04 |
| 000513 | 丽珠集团 | 2022-01-04 | 002365 | 永安药业 | 2022-01-04 |
| 000528 | 柳工 | 2022-01-04 | 002373 | 千方科技 | 2022-01-04 |
| 000546 | 金圆股份 | 2022-01-04 | 002425 | 凯撒文化 | 2022-01-04 |
| 000552 | 靖远煤电 | 2022-01-04 | 002443 | 金洲管道 | 2022-01-04 |
| 000557 | 西部创业 | 2022-01-04 | 002448 | 中原内配 | 2022-01-04 |
| 000581 | 威孚高科 | 2022-01-04 | 002454 | 松芝股份 | 2022-01-04 |
| 000623 | 吉林敖东 | 2022-01-04 | 002461 | 珠江啤酒 | 2022-01-04 |
| 000628 | 高新发展 | 2022-01-04 | 002558 | 巨人网络 | 2022-01-04 |
| 000635 | 英力特 | 2022-01-04 | 002563 | 森马服饰 | 2022-01-04 |
| 000655 | 金岭矿业 | 2022-01-04 | 002602 | 世纪华通 | 2022-01-04 |
| 000656 | 金科股份 | 2022-01-04 | 002422 | 科伦药业 | 2021-12-01 |
| 000671 | 阳光城 | 2022-01-04 | 002081 | 金螳螂 | 2021-12-01 |

表 3-1 大数据 100 指数的最新指数成分（续）

| 品种代码 | 品种名称 | 纳入日期 | 品种代码 | 品种名称 | 纳入日期 |
|--------|------|------------|--------|------|------------|
| 000685 | 中山公用 | 2022-01-04 | 002022 | 科华生物 | 2021-12-01 |
| 000703 | 恒逸石化 | 2022-01-04 | 002042 | 华孚时尚 | 2021-12-01 |
| 000709 | 河钢股份 | 2022-01-04 | 002010 | 传化智联 | 2021-12-01 |
| 000717 | 韶钢松山 | 2022-01-04 | 002421 | 达实智能 | 2021-12-01 |
| 000753 | 漳州发展 | 2022-01-04 | 002077 | 大港股份 | 2021-12-01 |
| 000756 | 新华制药 | 2022-01-04 | 000652 | 泰达股份 | 2021-12-01 |
| 000761 | 本钢板材 | 2022-01-04 | 002449 | 国星光电 | 2021-12-01 |
| 000825 | 太钢不锈 | 2022-01-04 | 000789 | 万年青 | 2021-12-01 |
| 000830 | 鲁西化工 | 2022-01-04 | 000498 | 山东路桥 | 2021-12-01 |
| 000898 | 鞍钢股份 | 2022-01-04 | 002441 | 众业达 | 2021-12-01 |
| 000928 | 中钢国际 | 2022-01-04 | 000925 | 众合科技 | 2021-12-01 |
| 000937 | 冀中能源 | 2022-01-04 | 000650 | 仁和药业 | 2021-11-01 |
| 000951 | 中国重汽 | 2022-01-04 | 000156 | 华数传媒 | 2021-11-01 |
| 000959 | 首钢股份 | 2022-01-04 | 002545 | 东方铁塔 | 2021-11-01 |
| 000961 | 中南建设 | 2022-01-04 | 000680 | 山推股份 | 2021-11-01 |
| 000976 | 华铁股份 | 2022-01-04 | 002573 | 清新环境 | 2021-11-01 |
| 000989 | 九芝堂 | 2022-01-04 | 002375 | 亚厦股份 | 2021-10-08 |
| 000990 | 诚志股份 | 2022-01-04 | 002038 | 双鹭药业 | 2021-10-08 |
| 000999 | 华润三九 | 2022-01-04 | 000028 | 国药一致 | 2021-10-08 |
| 002035 | 华帝股份 | 2022-01-04 | 002020 | 京新药业 | 2021-09-01 |
| 002048 | 宁波华翔 | 2022-01-04 | 000153 | 丰原药业 | 2021-08-02 |
| 002060 | 粤水电 | 2022-01-04 | 002252 | 上海莱士 | 2021-08-02 |
| 002078 | 太阳纸业 | 2022-01-04 | 002469 | 三维化学 | 2021-08-02 |
| 002091 | 江苏国泰 | 2022-01-04 | 002106 | 莱宝高科 | 2021-06-01 |
| 002092 | 中泰化学 | 2022-01-04 | 002101 | 广东鸿图 | 2022-01-04 |
| 002097 | 山河智能 | 2022-01-04 | 002109 | 兴化股份 | 2022-01-04 |

从上述指数的结果之中可以看出，当前的新浪大数据 100 包含了医药行业、科技行业、钢铁制造行业、造纸、建设行业、水电发电、化工、矿业、能源领域等等，涵盖的面非常广，基于此计算出的大数据指数能够给予投资者较为准确的参考。

3.2.3 大数据指数的构造情况

大数据指数是对样本空间股票，根据市场驱动因素、财务因素及大数据因素得分进行模型评分，取综合评分最高的前 100 只股票构成大数据 100 指标的初始

样本股。其评分标准包括：股票的热度、股票的换手率、价格波动率、流动性等，并最终通过量化因子模型计算得出市场驱动因素得分。由于其考察了散户投资者的情绪反应，因此在实际的量化分析之中引入这一因子能够补充传统量化分析对投资者情绪分析不足的问题。根据上述计算方法，最终能够确定各股票在每个交易日的具体得分，并按照得分顺序从高到低来列出 100 只股票，并结合这 100 只股票再综合计算大数据指数，根据大数据指数能够对市场内投资者的情绪进行合理预测，以最终为投资者的投资决策提供依据。大数据 100 模型之中还涵盖对股票主营收入、利润率、业绩、历史股价、市场评价等方面的因素，最终结合历史回测的相关数据来确定对哪些股票纳入到大数据 100 之中，最终通过被纳入的 100 只股票计算出当期的大数据指数。从上述分析也可以看出大数据指数并不是一成不变的，而是在每期会随着被纳入股票的不同有一定的变化，其尊重的市场的变动性以及投资的即时性等特点，因此在实际的投资之中具有较好的参考价值。大数据指数的构造情况如下图 3-3 所示。

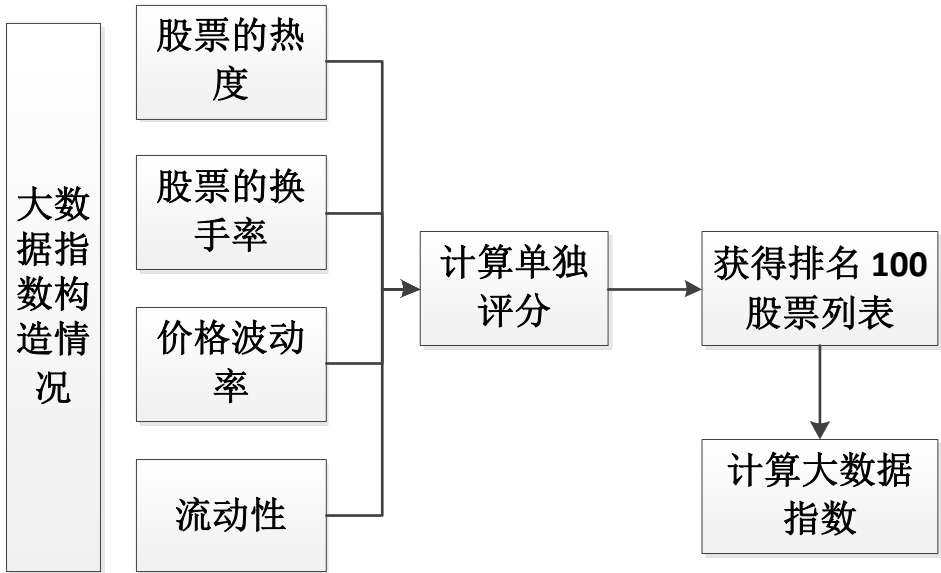


图 3-3 大数据指数的构造情况

数据来源：新浪大数据 100

由于大数据指数、大数据的股票列表都可以直接读取、无需计算，因此本文在后续引入大数据指数中直接引用了新浪大数据 100 公布的大数据指数数据，作为后续研究的基础。

大数据指数一般来说有三个方面的重要意义：第一是能够直接反映出市场中投资者的情绪，通过大数据 100 数据能够直接观察投资者的投资行为，因为大数据因子的计算逻辑之中考察了每个交易日或者月度周期内用户的交易行为，因此

通过大数据 100 的数据变动能够分析投资者的情绪变化情况；第二是能够给与投资者一定的参考建议，自 2017 年上线以来大数据 100 的指数变化情况与市场的直接波动关联较大，在实际投资中大数据 100 能够给予投资者一定的参考，以避免投资风险；第三是大数据 100 本身的收益率表现就比较好，投资大数据 100 能够跑赢市场的指数，在投资市场上是一个很好的参照值。

3.2.4 2021 年大数据指数变动情况

大数据指数实际上是处于一直不断变化的过程之中，自发布开始大数据指数也经历了上涨、下调、缓步上升等多个阶段。由于其计算原理与股市内表现最好的 100 只股票有关，因此其指数也处于不断变化的过程之中，如下图所示为 2021 年年度大数据指数的变化情况。



图 3-4 2021 年大数据指数 100 变动情况

如上图 3-3 所示 2021 年全年的大数据指数的变化情况，可以看出在 1 月份左右指数在不断下降，并最终在 1 个月后达到拐点上涨，而在之后的 8 个月时间内一直到 10 月又在上涨通道中。而结合沪深 A 股的指数变化情况进行对比可以看出，2021 年大数据指数的整体变化趋势是与沪深 A 股的指数趋同的，即在稳步上升的通道中。但在具体月份的具体走势上，大数据指数的波动，或者说上涨的幅度要大于沪深 A 股，且整体上来看大数据指数的下降阶段要少的多。尤其是在 2021 年的 2 月以及 10 月相关的指数差距能够达到 1.4-1.2 倍左右，具体情况如下图 3-5 所示：



图 3-5 2021 年沪深 A 股数据变化情况

3.3 基于大数据因子的 French 模型构建

未引入大数据因子的 Fama-French 理论模型如下所示：

$$R_{it} - R_{ft} = \alpha_i + \beta_i(R_{mt} - R_{ft}) + s_iSMB_t + h_iHML_t + \varepsilon_{it}$$

而在引入了新浪大数据 100 因子后，相关的模型调整如下：

$$R_{it} - R_{ft} = \alpha_i + \beta_i(R_{mt} - R_{ft}) + s_iSMB_t + h_iHML_t + b_iBD_t + \varepsilon_{it}$$

与原始模型相比，即在最后的表达式中加入了 b_iBD_t ，即大数据因子。

对这一模型的解释如下：

R_{it} 是目标投资组合 i （包括 1 只或者多只股票）在 t 时间段内的收益率。

R_{ft} 是在 t 时期内的无风险利润率（包括 1 只或者多只股票）

而上述两者之差是投资组合 i 的超额收益率（超过无风险利润率的额外收益率）

α_i 是截距项

α_i 、 β_i 、 s_i 、 h_i 、 b_i 、分别表示不同的系统，如 s_i 是 SMB 的系数，大数据因子 BD 的系数是 b_i 。

最后 SMB_t 表示小规模股票构成的投资组合之中与同期大规模股票的投资组合在收益率上的差值。

HML_t 表示高低账面市值比股票构成的相应投资组合在同一时期的收益率差值。

BD_t 表示新浪大数据 100 构成的投资组合在特定时期 t 内的加权收益率。

3.4 自变量因子构成

在上述因子之中 BD 因子是大数据因子可以直接获取，无风险收益率可以通

过对标 5 年期国债利率来进行读取,而 SMB 以及 HML 需要进行计算模型的构建:

一是 SMB: 第一步 SMB 的计算需要首先将股票与账面市值联系起来,对其市值规模进行排列,主要是使用上一年年末的数据来进行分组,并计算单个股流通市值、进行大小排列、剔除掉账面市值为负的股票、以及没有收益率数据的股票。在计算中将其分为三类,市值在前 50%的记为 B 组、在后 50%的为 S 组;第二步再对上述按账面市值排列的股票进行分组,前 30%为 H 组、而排名最后的 30%为 L 组、在这两者中间的记为 M 组;第三步对这两步进行组合能够获得 6 个投资方案,即 SH、SM、SL、BH、BM、BL 等组合,其具体涵义如下表 3-2 所示;

表 3-2 不同组合的具体含义

| 组合简写 | 按市值排列(顺序) |
|------|------------------|
| SH | 后 50%股票 +前 30%股票 |
| SM | 后 50%股票+30-70%股票 |
| SL | 后 50%股票+后 30%股票 |
| BH | 前 50%股票+前 30%股票 |
| BM | 前 50%股票+30-70%股票 |
| BL | 前 50%股票+后 30%股票 |

第四步也是最后一步在上述不同的投资组合方案进行加权平均收益率后构建出 SMB 以及 HML:

$$SMB = \frac{1}{3} \times (SH + SM + SL) - \frac{1}{3} \times (BH + BM + BL)$$

$$HML = \frac{1}{2} \times (SH + BH) - \frac{1}{2} \times (SL + BL)$$

这一公式表达的意思是各个投资类型的收益情况,其中 SMB_t 表示在 t 时期内公司规模大小差异对投资组合收益率的影响, HML_t 表示在 t 时期内账面市值比的高低之分对最终投资组合收益率的影响情况。

3.5 因变量的选取

由于我国股票市场不同板块的特点以及收益率均有很大差别,因此对上述沪深 A 股的股票按照其经营内容、主营业务的不同进行板块划分,结合 Wind 以及 Choice 客户端对板块数据进行筛选,并对 2021 年最新的股票数据进行整理,最终筛选出来十六个不同的板块类别,其中包括:矿业、电力燃气、建筑领域、邮政仓储、医疗卫生、农林渔牧业、餐饮酒店、制造业、综合板块、软件与技术信息、融资租赁、交通运输、科学技术研发、公共设施建设、以及文创文化板块、金融

业等，如下图 3-6 所示：

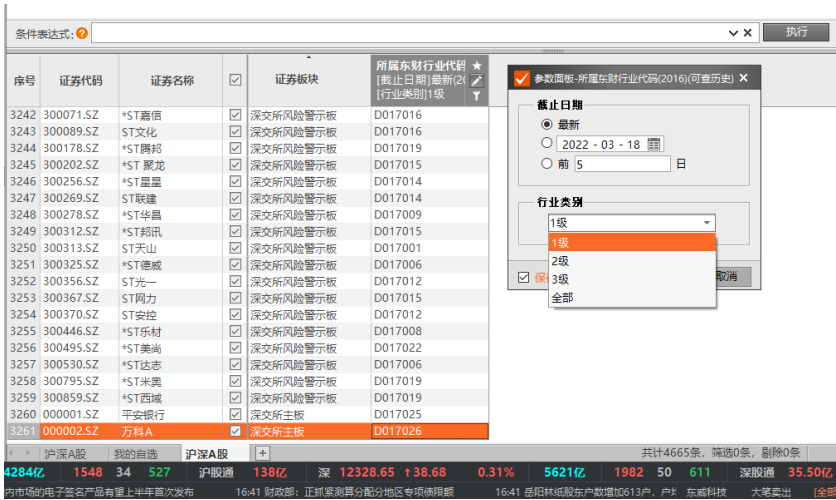


图 3-6 对沪深 A 股中股票的证券板块、所属行业板块进行筛选

如上图所示对证券板块、所属行业板块进行筛选后获得原始数据。

3.6 技术准备

本文选用的数据是来自沪深 A 股共 5724 个股票，剔除了部分停牌等无法正常交易的股票。同时对数据进行整理对每一个股票的 excel 表格整理为每日指标、财报指标、周线行业、月线行情等，具体如下图 3-7 所示：

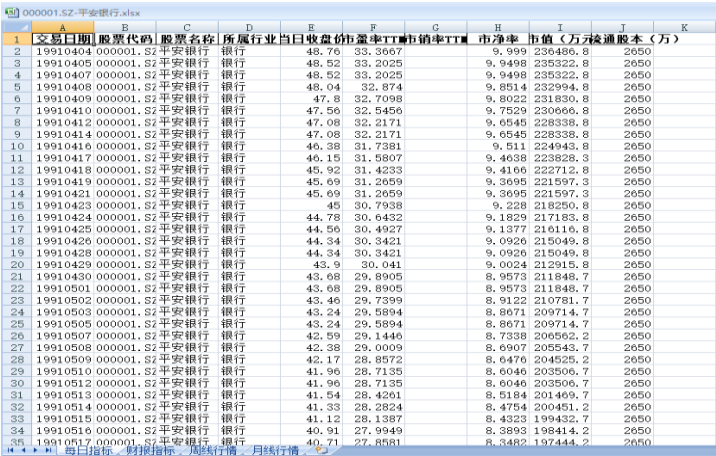


图 3-7 单个股票数据整理

对上述股票数据进行整理，利用了 Python 的 IDE 平台 Pycharm 内置的编译器进行 excel 批量处理功能，仅保留了 2021 年 12 月 31 日为截止日往前的一共 80 期股票数据，对关键数据存在缺失的股票（连续 6 期以上部分数据为空的）予以剔除。对数据进行合并处理，最终按照公司文件、个股回报率、日个股回报率、

三因子+大数据因子、是否 ST、收益率、无风险利率、月个股回报率、资产负债表、综合月市场回报率等进行股票数据的合并，由于最终是利用 Stata 执行数据分析，因此最终对所有数据整理成为 dta 文件。最终将所有数据处理后，通过建模获取 MKT、SMB、HML 等数据，并利用直接获取到的 BD（大数据因子数据）进行处理。利用 Stata 的 Do 文件编写具体分析的过程，以及各指标的具体计算方式，利用 Do 文件能够方便的对各类函数进行调用、同时可以使用中文进行编程，能够便捷的实现本文针对大数据因子加入后量化模型的有效性的研究。

3.7 本章小结

本章的主要内容是首先对因子组合量化的投资策略进行分析，并在此基础上介绍了大数据因子、大数据指数的特点、大数据指数的意义等内容、以及对大数据指数的成分、以及指数变动与沪深 A 股变动之间的关系做了初步的分析；接着构建了基于大数据因子的 French 模型，对其自变量因子的构成、因变量的选取、技术准备等进行了较为详细的介绍，也分析了利用 Python、以及 stata 的 Do 文件进行程序编写的过程。

第四章 实证分析

4.1 描述性统计分析

4.1.1 行业分析

表 4-1 沪深 A 股市场板块数据描述

| 名称 | 分析期数 | 均值(%) | 样本差(%) | 含有股票数 |
|-------------|------|-------|--------|-------|
| 无风险利率 R_f | 80 | 0.20 | 0.05 | |
| 市场收益 | 80 | 0.09 | 7.36 | 1025 |
| 矿业板块 | 80 | 0.00 | 6.98 | 55 |
| 燃气能源板块 | 80 | 1.02 | 7.67 | 68 |
| 建筑业 | 80 | 1.55 | 13.32 | 45 |
| 交通运输物流业 | 80 | 1.06 | 10.03 | 72 |
| 金融板块 | 80 | 0.65 | 7.66 | 51 |
| 科研板块 | 80 | 3.53 | 14.19 | 8 |
| 医疗卫生板块 | 80 | 2.96 | 13.39 | 2 |
| 农林渔牧业 | 80 | 0.75 | 11.33 | 26 |
| 批发与零售板块 | 80 | 1.68 | 9.79 | 97 |
| 公共设施建设 | 80 | 2.47 | 10.11 | 21 |
| 文创文化板块 | 80 | 1.56 | 10.80 | 27 |
| 信息技术板块 | 80 | 3.01 | 14.84 | 44 |
| 制造板块 | 80 | 1.59 | 8.67 | 586 |
| 餐饮酒店板块 | 80 | 1.09 | 9.98 | 14 |
| 综合板块 | 80 | 2.49 | 11.15 | 28 |
| 融资租赁板块 | 80 | 2.37 | 10.70 | 25 |

从上表的数据之中可以看出，在分析的 80 期内收益率波动比较小的有金融板块、制造业等，以及矿业、燃气能源板块等，且与沪深 A 股的整体变化波动比较接近，这说明这些板块受市场波动影响较小。而在上述板块之中收益率最高的是科研板块，其次是医疗卫生板块，以及公共设施建设板块等，这些板块都有较为强有力的业务支撑，如医疗卫生板块的药品研发与销售、科研板块的技术研发授权与出让、公共设施建设等等发展势头都非常强劲，这类板块内的股票盈利率比较高，但同时也面临较大的市场风险，受到市场因素影响后的波动比较大。此外，由于创业板大多数股票上市的时间比较短，在投资组合之中能够支撑研究 80 期的股票只数比较少，最终在研究里对创业板的相关股票进行剔除。

4.1.2 因子相关性检验

表 4-2 pearson 相关系数

| | Rm-Rf | SMB | HML | BD |
|-------|---------|--------|---------|--------|
| Rm-Rf | 1.0000 | | | |
| SMB | 0.0794 | 1.0000 | | |
| HML | -0.1413 | 0.2251 | 1.0000 | |
| BD | 0.1984 | 0.2939 | -0.0279 | 1.0000 |

从上面的相关系数来看，可以看到各因子之间的相关性系数都比较小，全部低于 0.3，而相关性系数比较小则表明各因子之中不存在线性关系。从其具体涵义上来看，大数据因子是排名前 100 的股票综合计算后获得的指数，而 SMB、HML 等都是通过对市值比进行计算组合前 50%、后 50%、前 30%、后 30%以及中间 30-70%的不同组合确定其高低市值比之间收益率的差值，这一数据本身就与大数据 100 之间的关联性不大，只与市场的无风险收益率等有关。

4.1.3 BM 效应、规模效应分析

表 4-3 三因素模型按市值比的 6 组合收益率

| 按市值比 (后 50%/前 50%) | 账面市值比(BE/ME) | | |
|-----------------------|--------------|-------|-------|
| | L | M | H |
| S | 0.82 | 1.322 | 1.743 |
| B | 0.55 | 1.167 | 1.626 |

表 4-4 三因素模型按市值比的 6 组合方差

| 按市值比 (后 50%/前 50%) | 账面市值比(BE/ME) | | |
|-----------------------|--------------|------|------|
| | L | M | H |
| S | 10.94 | 9.11 | 9.06 |
| B | 9.32 | 6.75 | 6.30 |

从上述表 4-3 的数据之中可以看出，在按市值比的 6 组合投资方案之中所得投资组合的超额收益率均大于 0，且其规模与收益变化走势相反，账面市值比变化则与收益变化幅度一致。这就表明在公司的规模市值固定时，其超额收益率会跟随市值比的提升而逐步扩大。

4.2 Fama-Macbeth 回归分析

为了进一步研究在加入大数据因子后期模型与我国股票市场的匹配情况，在本节之中应用 Fama-Fama-Macbeth 的回归分析，对不同因子与超额收益率的关系进行研究，最终对其方程确定如下：

$$R_{it} = \alpha_t + \beta_t \ln(X_{it}) + \xi_{it}$$

在上述方程之中 i 表示沪深 A 股中被筛选后剩下的单只股票，t 表示计算期具体的时间范围，本次为 80 期。X 表示影响股票超额收益率的具体因素。同时上述公式可以看出，在第 t 时期对其计算，可以算得其在 80 期内不同期数有关 β_t 的按时间顺序的排列。

随后，需要对 β_t 时间序列的均值、标准差进行计算，并进行 T 检验，具体过程如下：

$$\begin{aligned}\bar{\beta} &= \frac{1}{\tau} \sum_{t=1}^{\tau} \beta_t \\ \sigma(\beta_t) &= \sqrt{\frac{1}{\tau} \sum_{t=1}^{\tau} (\beta_t - \bar{\beta})^2} \\ T(\beta_t) &= \sqrt{t} \frac{\bar{\beta}_t}{\sigma(\beta_t)}\end{aligned}$$

然后在 Fama-French 的应用假设下，可以通过上述公式计算出被测股票对股票收益率的影响是否显著。在随后针对规模因子、账面市值比因子以及 BD 新浪大数据 100 进行分析后，展开综合分析，构建分析的方程式，如下：

$$R_{it} = \alpha_t + \beta_{1t} \ln(X_{1it}) + \beta_{2t} \ln(X_{2it}) + \beta_{3t} \ln(X_{3it}) + \xi_{it}$$

针对本文研究的四因子模型，最终的因子检验结果如下表所示：

表 4-5 四因子模型的检验结果

| ln(ME) | ln(BE/ME) | ln(BD) |
|-----------------------|----------------------|----------------------|
| 单因子检验 | | |
| -0.59% ** (-1.82) | 0.65% ** (-1.68) | 1.47% *** (-2.17) |
| 多因子检验 | | |
| -0.53% *** (-2.08) | 0.71% *** (-1.72) | |
| 0.69% **** | | 1.11% *** |

表 4-5 四因子模型的检验结果（续）

| ln(ME) | ln(BE/ME) | ln(BD) |
|----------|-----------|----------|
| (-1.75) | | (-2.05) |
| 0.35%*** | 0.43%** | 0.88%*** |
| (-1.69) | (-1.54) | (-2.25) |

通过上述回归结果可以看出，规模因子、账面市值比因子都有较强的显著性，这两者对股票收益率都有影响，也符合国内外针对三因子模型的相关研究。尤其是本文新加入的新浪大数据 100，在上述因子检验之中数据结果也较为显著，这说明新浪大数据 100 对股票的收益具有可信的解释能力。

4.3 引入大数据因子改进的 Fama-French 模型适应性分析

对沪深 A 股按照上述的主营业务进行区分最终划分为 16 个板块，并用 stata 对时间范围在 2010 年 12 月 31 日—2021 年 12 月 31 日为样本展开 Fama-Macbeth 的回归，相关结果如下：

表 4-6 四因子模型中各板块回归系数分析

| 投资组合 | α_i | β_i | S_i | h_i | b_i | R^2 |
|--------|---------------|-----------|----------|--------------------|---------------|-------|
| 矿业板块 | 0.12%*** | 0.914*** | 0.509*** | -0.497*** | -0.244** * | 0.932 |
| | (4.65) | (24.71) | (8.01) | (-7.72) | (-4.45) | |
| 燃气能源板块 | 0.09%*** | 0.894*** | 0.359*** | -0.282*** | 0.0249 | 0.925 |
| | (4.83) | (20.47) | (7.03) | (-5.25) | (0.55) | |
| 建筑业 | 0.129%** | 0.916*** | 0.263*** | . | -0.0287 | 0.924 |
| | (3.81) | (20.48) | (3.59) | 0.113** (-2.16) | (-0.449) | |
| 交通运输物流 | 0.169%** * | 0.957*** | 0.402*** | -0.182*** | -0.097** | 0.902 |
| | (3.88) | (24.00) | (7.07) | (-3.05) | (-2.19) | |
| 金融板块 | 0.139%** * | 0.986*** | -0.052 | 0.601 | -0.166** * | 0.871 |
| | (4.16) | (21.62) | (-0.68) | (1.06) | (-2.51) | |
| 科研板块 | 0.179% | 0.959 | 0.533 | -0.642 | 0.510** | 0.886 |
| | (1.11) | (1.45) | (1.42) | (-1.41) | (2.32) | |
| 农林渔牧业 | 0.67%** | 0.893*** | 0.731*** | -0.400*** | 0.088 | 0.921 |
| | -1.86 | -11.62 | -7.68 | (-4.00) | -1.11 | |

表 4-6 四因子模型中各板块回归系数分析（续）

| 投资组合 | α_i | β_i | S_i | h_i | b_i | R^2 |
|---------|----------------|-----------|----------|-----------|----------|-------|
| 批发与零售板块 | 0.276%** * | 0.785*** | 0.538*** | -0.278*** | 0.212*** | 0.942 |
| | -3.88 | -20.85 | -11.55 | (-5.71) | -5.47 | |
| 公共设施建设 | 0.19%** | 0.279 | 0.096 | -0.196 | 0.753*** | 0.866 |
| | (1.56) | -1.41 | -0.469 | (-0.83) | -3.88 | |
| 医疗卫生板块 | -0.88% | -0.248 | -0.395 | 0.44 | 1.252*** | 0.896 |
| | (-0.48) | (-0.88) | (-1.02) | -1.19 | -4.03 | |
| 文创文化板块 | -0.129%** * | 0.303*** | 0.205** | -0.024 | 0.642*** | 0.930 |
| | -2.05 | -3.21 | -1.93 | (-0.279) | -6.37 | |
| 信息技术板块 | 0.139%** | 0.415*** | 0.314*** | 0.0231 | 0.721*** | 0.891 |
| | -1.87 | -4.01 | -2.48 | -0.179 | -6.64 | |
| 制造板块 | 0.139%** * | 0.739*** | 0.575*** | -0.322*** | 0.256*** | 0.917 |
| | -5.44 | -44.35 | (2.00) | (-15.22) | -14.97 | |
| 餐饮酒店板块 | 0.169%** | 0.728*** | 0.646*** | -0.19 | 0.048 | 0.901 |
| | -2.29 | -4.96 | -3.59 | (-1.01) | -0.379 | |
| 综合板块 | 0.198%** * | 0.902*** | 0.948*** | -0.435*** | 0.009** | 0.926 |
| | (3.78) | -8.72 | (7.38) | (-3.28) | (2.32) | |
| 融资租赁板块 | 0.139%** * | 0.837*** | 0.604*** | -0.180** | 0.101** | 0.952 |
| | -2.62 | -4.48 | -2.58 | -2.27 | -2.05 | |

注：*表示显著性程度，*、**、***分别表示 $P < 0.1$ 、 0.05 、 0.01 等。而上述 P 值下方括号内的数值为 t 值。

4.3.1 股票各板块回归系数分析

(1) 采矿板块

从上述回归结果之中可以看出，针对采矿板块的具体情况如下：

rm-rf 的回归系数 β_i 为 0.914，接近数值 1 且大于 0，同时其 $P < 0.01$ 通过显著性检验。因此，这一系数能够较好的解释市场因素是影响股票收益率变化的最主要因素之一，也符合国内外学者针对 FAMA-French 模型的检验结果。

SMB 的系数 S_i 为 0.509 > 0，同时其 P 值 < 0.01 通过显著性检验。这说明 SMB 可以解释在超额收益率能够解释的因素以外的方面，能够显著影响矿业板块股票的收益率。

HML 回归系数 h_i 为 -0.497 < 0，同时其 P 值 < 0.01 通过显著性检验。这表明通

过账面市值比因子能够较好的解释矿业板块下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为 $-0.244 < 0$ ，同时其 P 值 < 0.01 通过显著性检验。这表明通过应用新浪大数据 100 因子与矿业板块的收益率是负相关的，也能够解释矿业板块的具体收益情况。

（2）燃气能源板块

从上述回归结果之中可以看出，针对燃气能源板块的具体情况如下：

rm-rf 的回归系数 β_i 为 0.894，接近数值 1 且大于 0，同时其 $P < 0.01$ 通过显著性检验。因此，这一系数能够较好的解释市场因素是影响股票收益率变化的最主要因素之一，也符合国内外学者针对 FAMA-French 模型的检验结果。

SMB 的回归系数 S_i 为 $0.359 > 0$ ，同时其 P 值 < 0.01 通过显著性检验。这说明 SMB 可以解释在超额收益率能够解释的因素以外的方面，能够显著影响燃气能源板块股票的收益率。

HML 的回归系数 h_i 为 $-0.282 < 0$ ，同时其 P 值 < 0.01 通过显著性检验。这表明通过账面市值比因子能够较好的解释燃气能源板块下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为 $0.0249 > 0$ ，同时其 P 值 < 0.01 通过显著性检验。这表明通过应用新浪大数据 100 因子与燃气能源板块的收益率是正相关的，也能够解释燃气能源板块的具体收益情况。

（3）建筑业

从上述回归结果之中可以看出，针对建筑业的具体情况如下：

rm-rf 回归系数 β_i 为 0.916，接近数值 1 且大于 0，同时其 $P < 0.01$ 通过显著性检验。因此，这一系数能够较好的解释市场因素是影响股票收益率变化的最主要因素之一，也符合国内外学者针对 FAMA-French 模型的检验结果。

SMB 回归系数 S_i 为 $0.263 > 0$ ，同时其 P 值 < 0.01 通过显著性检验。这说明 SMB 可以解释在超额收益率能够解释的因素以外的方面，能够显著影响建筑业股票的收益率。

HML 回归系数 h_i 为 $0.113 > 0$ ，同时其 P 值 < 0.01 通过显著性检验。这表明通过账面市值比因子能够较好的解释建筑业下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为 $-0.0287 < 0$ ，同时其 P 值 > 0.1 未通过显著性检验。这表明通过应用新浪大数据 100 因子与建筑业收益率的相关性不高，不能够解释建筑业的具体收益情况。

（4）交通运输物流业

从上述回归结果之中可以看出，针对交通运输物流业的具体情况如下：

rm-rf 回归系数 β_i 为 0.957，接近数值 1 且大于 0，同时其 $P < 0.01$ 通过显著性

检验。因此，这一系数能够较好的解释市场因素是影响股票收益率变化的最主要因素之一，也符合国内外学者针对 FAMA-French 模型的检验结果。

SMB 回归系数 S_i 为 0.402>0, 同时其 P 值<0.01 通过显著性检验。这说明 SMB 可以解释在超额收益率能够解释的因素以外的方面，能够显著影响交通运输物流业股票的收益率。

HML 回归系数 h_i 为-0.182<0, 同时其 P 值<0.01 通过显著性检验。这表明通过账面市值比因子能够较好的解释交通运输物流业下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为-0.097<0, 同时其 P<0.05 通过显著性检验。这表明通过应用新浪大数据 100 因子与交通运输物流业的收益率的相关性高，能够解释交通运输物流业的具体收益情况。

(5) 金融板块

从上述回归结果之中可以看出，针对金融板块的具体情况如下：

rm-rf 回归系数 β_i 为 0.986, 接近数值 1 且大于 0, 同时其 P<0.01 通过显著性检验。因此，这一系数能够较好的解释市场因素是影响股票收益率变化的最主要因素之一，也符合国内外学者针对 FAMA-French 模型的检验结果。

SMB 回归系数 S_i 为-0.052<0, 同时其 P>0.1 未通过显著性检验。这说明 SMB 不能解释在超额收益率能够解释的因素以外的方面，不能显著影响金融板块股票的收益率。

HML 回归系数 h_i 为 0.601>0, 同时其 P>0.1 未通过显著性检验。这表明通过账面市值比因子不能够较好的解释金融板块下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为-0.166<0, 同时其 P<0.01 通过显著性检验。这表明通过应用新浪大数据 100 因子与金融板块的收益率的相关性高，能够解释金融板块的具体收益情况。

(6) 科研板块

从上述回归结果之中可以看出，针对科研板块的具体情况如下：

rm-rf 回归系数 β_i 为 0.959, 接近数值 1 且大于 0, 同时其 P>0.1 未通过显著性检验。因此，这一系数不能够较好的解释市场因素是影响股票收益率变化的最主要因素之一。

SMB 回归系数 S_i 为 0.533>0, 同时其 P>0.1 未通过显著性检验。这说明 SMB 不能解释在超额收益率能够解释的因素以外的方面，不能显著影响科研板块股票的收益率。

HML 回归系数 h_i 为-0.642<0, 同时其 P>0.1 未通过显著性检验。这表明通过账面市值比因子不能够较好的解释科研板块下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为 0.510>0, 同时其 $P<0.01$ 通过显著性检验。这表明通过应用新浪大数据 100 因子与科研板块的收益率的相关性高, 能够解释科研板块的具体收益情况。

(7) 农林渔牧业

从上述回归结果之中可以看出, 针对农林渔牧业的具体情况如下:

rm-rf 回归系数 β_i 为 0.893, 接近数值 1 且大于 0, 同时其 $P<0.01$ 通过显著性检验。因此, 这一系数能够较好的解释市场因素是影响股票收益率变化的最主要因素之一, 也符合国内外学者针对 FAMA-French 模型的检验结果。

SMB 回归系数 S_i 为 0.731>0, 同时其 $P<0.01$ 通过显著性检验。这说明 SMB 可以解释在超额收益率能够解释的因素以外的方面, 能够显著影响农林渔牧业股票的收益率。

HML 回归系数 h_i 为 -0.400<0, 同时其 $P<0.01$ 通过显著性检验。这表明通过账面市值比因子能够较好的解释农林渔牧业下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为 0.088>0, 同时其 $P>0.1$ 未通过显著性检验。这表明通过应用新浪大数据 100 因子与农林渔牧业的收益率的相关性不高, 不能够解释农林渔牧业的具体收益情况。

(8) 批发零售板块

从上述回归结果之中可以看出, 针对批发零售板块的具体情况如下:

rm-rf 回归系数 β_i 为 0.785, 接近数值 1 且大于 0, 同时其 $P<0.01$ 通过显著性检验。因此, 这一系数能够较好的解释市场因素是影响股票收益率变化的最主要因素之一, 也符合国内外学者针对 FAMA-French 模型的检验结果。

SMB 回归系数 S_i 为 0.538>0, 同时其 $P<0.01$ 通过显著性检验。这说明 SMB 可以解释在超额收益率能够解释的因素以外的方面, 能够显著影响批发零售板块股票的收益率。

HML 回归系数 h_i 为 -0.278<0, 同时其 $P<0.01$ 通过显著性检验。这表明通过账面市值比因子能够较好的解释批发零售板块下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为 0.212>0, 同时其 $P<0.01$ 通过显著性检验。这表明通过应用新浪大数据 100 因子与批发零售板块的收益率的相关性高, 能够解释批发零售板块的具体收益情况。

(9) 公共设施建设

从上述回归结果之中可以看出, 针对公共设施建设板块的具体情况如下:

rm-rf 回归系数 β_i 为 0.279, 数值较小, 且其 $P>0.1$ 未通过显著性检验。因此, 这一系数能够不能较好的解释市场因素是影响股票收益率变化的最主要因素之

一。

SMB 回归系数 S_i 为 $0.096 > 0$ ，同时其 $P > 0.1$ 未通过显著性检验。这说明 SMB 不能解释在超额收益率能够解释的因素以外的方面，不能够显著影响公共设施建设板块股票的收益率。

HML 回归系数 h_i 为 $-0.196 < 0$ ，同时其 $P > 0.1$ 未通过显著性检验。这表明通过账面市值比因子不能够较好的解释公共设施建设板块下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为 $0.753 > 0$ ，同时其 $P < 0.01$ 通过显著性检验。这表明通过应用新浪大数据 100 因子与公共设施建设板块的收益率的相关性高，能够解释公共设施建设板块的具体收益情况。

（10）医疗卫生板块

从上述回归结果之中可以看出，针对公共设施建设板块的具体情况如下：

rm-rf 回归系数 β_i 为 -0.248 ，数值较小且小于 0，同时且其 $P > 0.1$ 未通过显著性检验。因此，这一系数能够不能较好的解释市场因素是影响股票收益率变化的最主要因素之一。

SMB 回归系数 S_i 为 $-0.395 < 0$ ，同时其 $P > 0.1$ 未通过显著性检验。这说明 SMB 不能解释在超额收益率能够解释的因素以外的方面，不能够显著影响公共设施建设板块股票的收益率。

HML 回归系数 h_i 为 $0.44 > 0$ ，同时其 $P > 0.1$ 未通过显著性检验。这表明通过账面市值比因子不能够较好的解释公共设施建设板块下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为 $1.252 > 0$ ，同时其 $P < 0.01$ 通过显著性检验。这表明通过应用新浪大数据 100 因子与公共设施建设板块的收益率的相关性高，能够解释公共设施建设板块的具体收益情况。

（11）文创文化板块

从上述回归结果之中可以看出，针对文创文化板块的具体情况如下：

rm-rf 回归系数 β_i 为 $0.303 > 0$ ，同时其 $P < 0.01$ 通过显著性检验。因此，这一系数能够较好的解释市场因素是影响股票收益率变化的最主要因素之一，也符合国内外学者针对 FAMA-French 模型的检验结果。

SMB 回归系数 S_i 为 $0.205 > 0$ ，同时其 $P < 0.01$ 通过显著性检验。这说明 SMB 可以解释在超额收益率能够解释的因素以外的方面，能够显著影响文创文化板块股票的收益率。

HML 回归系数 h_i 为 $-0.024 < 0$ ，同时其 $P > 0.01$ 未通过显著性检验。这表明通过账面市值比因子不能够较好的解释文创文化板块下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为 $0.642 > 0$ ，同时其 $P < 0.01$ 通过显著性

检验。这表明通过应用新浪大数据 100 因子与文创文化板块的收益率的相关性高，能够解释文创文化板块的具体收益情况。

（12）信息技术板块

从上述回归结果之中可以看出，针对信息技术板块的具体情况如下：

rm-rf 回归系数 β_i 为 0.415>0，同时其 $P<0.01$ 通过显著性检验。因此，这一系数能够较好的解释市场因素是影响股票收益率变化的最主要因素之一，也符合国内外学者针对 FAMA-French 模型的检验结果。

SMB 回归系数 S_i 为 0.314>0，同时其 $P<0.01$ 通过显著性检验。这说明 SMB 可以解释在超额收益率能够解释的因素以外的方面，能够显著影响信息技术板块股票的收益率。

HML 回归系数 h_i 为 0.0231>0，同时其 $P>0.01$ 未通过显著性检验。这表明通过账面市值比因子不能够较好的解释信息技术板块下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为 0.721，>0，同时其 $P<0.01$ 通过显著性检验。这表明通过应用新浪大数据 100 因子与信息技术板块的收益率的相关性高，能够解释信息技术板块的具体收益情况。

（13）制造板块

从上述回归结果之中可以看出，针对制造板块的具体情况如下：

rm-rf 回归系数 β_i 为 0.739>0，同时其 $P<0.01$ 通过显著性检验。因此，这一系数能够较好的解释市场因素是影响股票收益率变化的最主要因素之一，也符合国内外学者针对 FAMA-French 模型的检验结果。

SMB 回归系数 S_i 为 0.575>0，同时其 $P<0.01$ 通过显著性检验。这说明 SMB 可以解释在超额收益率能够解释的因素以外的方面，能够显著影响制造板块股票的收益率。

HML 回归系数 h_i 为 -0.322<0，同时其 $P<0.01$ 通过显著性检验。这表明通过账面市值比因子能够较好的解释制造板块下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为 0.256>0，同时其 $P<0.01$ 通过显著性检验。这表明通过应用新浪大数据 100 因子与制造板块的收益率的相关性高，能够解释制造板块的具体收益情况。

（14）餐饮酒店板块

从上述回归结果之中可以看出，针对餐饮酒店板块的具体情况如下：

rm-rf 回归系数 β_i 为 0.728>0，同时其 $P<0.01$ 通过显著性检验。因此，这一系数能够较好的解释市场因素是影响股票收益率变化的最主要因素之一，也符合国内外学者针对 FAMA-French 模型的检验结果。

SMB 回归系数 S_i 为 $0.646 > 0$ ，同时其 $P < 0.01$ 通过显著性检验。这说明 SMB 可以解释在超额收益率能够解释的因素以外的方面，能够显著影响餐饮酒店板块股票的收益率。

HML 回归系数 h_i 为 $-0.19 < 0$ ，同时其 $P > 0.1$ 未通过显著性检验。这表明通过账面市值比因子不能够较好的解释餐饮酒店板块下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为 $0.048 > 0$ ，同时其 $P > 0.1$ 通过显著性检验。这表明通过应用新浪大数据 100 因子与餐饮酒店板块的收益率的相关性低，不能够解释餐饮酒店板块的具体收益情况。

（15）综合板块

从上述回归结果之中可以看出，针对综合板块的具体情况如下：

rm-rf 回归系数 β_i 为 $0.902 > 0$ 且接近 1，同时其 $P < 0.01$ 通过显著性检验。因此，这一系数能够较好的解释市场因素是影响股票收益率变化的最主要因素之一，也符合国内外学者针对 FAMA-French 模型的检验结果。

SMB 回归系数 S_i 为 $0.948 > 0$ ，同时其 $P < 0.01$ 通过显著性检验。这说明 SMB 可以解释在超额收益率能够解释的因素以外的方面，能够显著影响综合板块股票的收益率。

HML 回归系数 h_i 为 $-0.435 < 0$ ，同时其 $P < 0.1$ 通过显著性检验。这表明通过账面市值比因子能够较好的解释综合板块下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为 $0.009 > 0$ ，同时其 $P < 0.05$ 通过显著性检验。这表明通过应用新浪大数据 100 因子与综合板块的收益率的相关性高，能够解释综合板块的具体收益情况。

（16）融资租赁板块

从上述回归结果之中可以看出，针对融资租赁板块的具体情况如下：

rm-rf 回归系数 β_i 为 $0.837 > 0$ 且接近 1，同时其 $P < 0.01$ 通过显著性检验。因此，这一系数能够较好的解释市场因素是影响股票收益率变化的最主要因素之一，也符合国内外学者针对 FAMA-French 模型的检验结果。

SMB 回归系数 S_i 为 $0.604 > 0$ ，同时其 $P < 0.01$ 通过显著性检验。这说明 SMB 可以解释在超额收益率能够解释的因素以外的方面，能够显著影响融资租赁板块股票的收益率。

HML 回归系数 h_i 为 $-0.180 < 0$ ，同时其 $P < 0.1$ 通过显著性检验。这表明通过账面市值比因子能够较好的解释融资租赁板块下股票的收益情况。

新浪大数据 100 因子 BD 回归系数 b_i 为 $0.101 > 0$ ，同时其 $P < 0.05$ 通过显著性检验。这表明通过应用新浪大数据 100 因子与融资租赁板块的收益率的相关性高，能够解释融资租赁板块的具体收益情况。

4.3.2 Fama-French 模型中引入大数据因子更适合中国投资市场

通过上述实证研究可以得出下列结论：在上述由三因子模型和大数据因子构成的四因子模型之中，各因子之间不存在显著的线性关系。同时也验证了 FAMA-French 的重要结论即当公司的市场规模固定的时候，其月均收益率的变动会随着账面市值比的变动同向变化；而这一点在沪深 A 股之中也有验证，即当公司账面市值比确定的时候市值排名靠前的公司，其实际由经营产生的收益对股票收益的影响将会越低。

在上述的分析之中 16 个具体的板块分析内容上只有 3 个板块的 α_i 未达到显著标准，而有三个具体板块的市场因子未通过 t 检验、其它有 4 个板块大数据因子的回归分析未通过 t 检验、7 个 HML 未通过 t 检验、4 个 SMB 因子未通过 t 检验。这也说明了在中国股票市场之中除了上述因子之外还有很多潜在因素会影响市场行情，但最终分析这四个因子对绝大部分投资组合都能够很好解释、同时也能通过 t 检验，因此适用于中国沪深 A 股市场展开相关研究。

4.4 本章小结

本章对引入大数据因子后的模型解释情况进行研究，发现针对绝大部分投资组合都能够有很好的解释。而出现不符合的板块，如建筑业、制造板块、餐饮酒店板块等，这三个板块的经营内容实际上是与短期内的经营情况关联不大的，如餐饮酒店业、制造板块、建筑业都存在淡旺季且淡旺季差距明显，但其股票市值情况并不一定是与大数据指数的变动情况相符。如出现较大的利好，如国家对制造业的利好政策等，都会极大提升其股票账面价值，而与大数据因子关联不大。同时也存在部分板块，如科研板块、金融板块、公共设施建设板块、医疗卫生板块等，大数据因子对这部分板块不能很好地解释其收益情况，主要是因为这类板块可能存在短期内变动较大，且受到当前后疫情时代影响，相关账面市值短期内有较大变动。如科研板块内的疫苗研发、公共设施建设的各类公共设施的投产等，实际情况需要具体分析，因此对其收益的解释可能存在偏差。

此外，本章通过研究发现在应用新浪大数据 100 因子后发现，这一因子对于绝大多数板块的收益情况都能够作出较好的解释，这也说明应用大数据 100 加入到 Fama-French 模型之中是非常有必要的，也能够对投资者的情绪、市场波动情况进行很好的解释。

第五章 量化策略的设计及优化

本文在第四章的实证分析之中加入了新浪大数据 100 因子，并对加入了大数据因子后的四因子模型进行了验证，结果发现能过对 A 股数据进行较好的解释，同时也能够给予投资者以有效的建议。因此在本章之中对加入了新浪大数据 100 的模型基础上展开量化选股策略的分析，以给予投资者相应的股票投资建议。

5.1 Choice 量化策略平台

Choice 平台东方财富开发的集合股票历史数据查询、股票数据分析、量化分析等的集成平台。在 Choice 之中也有丰富的数据量化接口，能够应用中文直接进行 Choice 内的量化分析建模，同时利用 Choice 的大型服务器能够对数据模型进行快速验证。同时 Choice 还提供了从交易到投资分析以及最终的投资效果检验的闭环服务，也能够为投资者提供各类定制化的量化分析接口，其量化平台的界面具体如下：

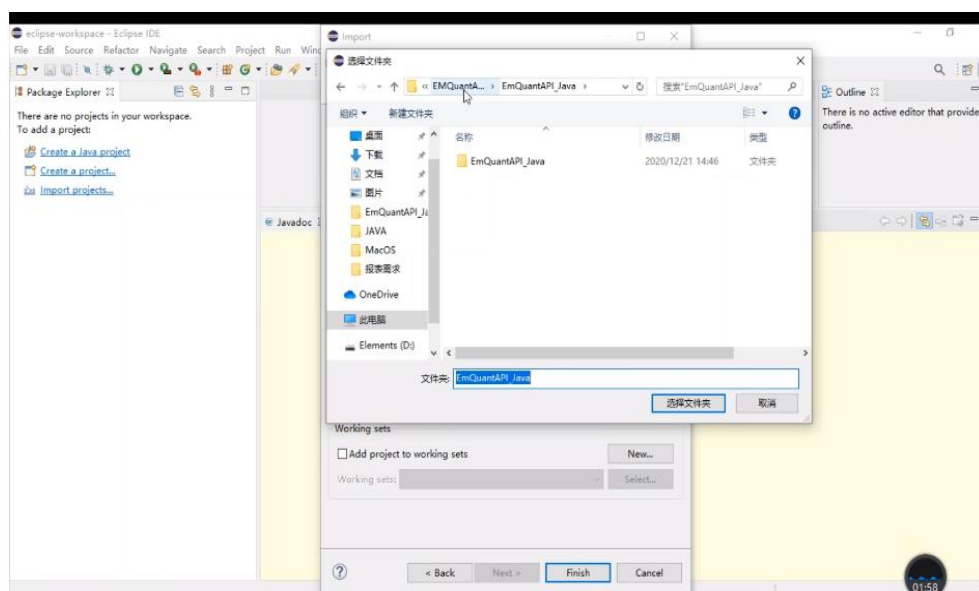


图 5-1 Choice 量化平台界面

从上图之中可以看出量化平台主要分为三个部分组成，最右侧是程序的便携区可以用来选择各类语言来编写量化投资的建模策略；左侧表示代码的层级以及代码之间的关系等；上部是量化操作的菜单栏可以打开录入数据、打开执行文件、运行代码等。同时在右侧编程去的上面还有日志区域，能够显示在具体策略运行

之中的相关时间参数、操作步骤、操作结果等，以及提示在操作中存在的错误等。对投资策略进行编写之后需要安装编译器进行运行，并通过运行来判断程序编写是否成功，如果有错误还需要按照语法提示进行调整给予修改，如果通过编译器的检查就可以运行程序。在实际对程序运行的过程之中能够显示各类投资策略的具体信息，这些信息主要包括对无风险收益率、投资组合的收益率、最大回撤率等进行计算，能够提供给投资者相应的参考。利用 python 编写增加了新浪大数据 100 的三因子模型的测算模型，其中对各类限制条件、数值的计算方法、以及数据的运算等都进行了限定，而在程序编写好了之后只需要录入数据就可以获得相应的策略结果。

5.2 量化投资策略设计和优化

本章将对量化投资选股的策略的设计思路进行阐述，主要是利用上述构建的四因子模型进行量化选股，回测投资组合的收益。本次分析共选择了 2016 年 1 月 1 日—2021 年 12 月 31 日，共 6 年的数据进行分析，对最终的总收益率、年化收益等进行分析，并在此基础上对初步设计出的模型进一步进行优化，最终让投资组合能够获得较大的超额收益，并能够运行相对稳定，为实际投资股票组合提供相应支持。

5.2.1 量化投资策略设计

本节运用多因子策略进行量化程序的设计，借用 Choice 金融终端自带的编程功能，设计策略的模块规则包括买卖策略规则、买卖时间规则、买卖频率规则、买卖仓位规则。

1、对股票交易的基本规则是按照上述验证后的四因子模型，每个月月开始对市场个股计算因子得分，并按照得分的具体分值选择分值较高的前 20 只股票进行投资，同时在每个月都会对股票展开新一轮筛选，对已经不再符合模型、或者得分、或者排名的股票出售。由于选定的股票在每个月都会变化，因此并没有固定的股票投资组合，其股票的得分公式如下：

$$R_{it} = \alpha_i + \beta_{1i} \ln(X_{1it}) + \beta_{2i} \ln(X_{2it}) + \beta_{3i} \ln(X_{3it}) + \xi_{it}$$

同时结合 2016 年 1 月 1 日—2021 年 12 月 31 日的不同时间，在 6 年时间的每个月初都需要对各板块的回归系数进行重测，以保证量化策略的预测功能。代入到上式中即计算单个股票的超额收益率，以矿业板块为例代入其各项值，如 $\alpha_i=0.0012$ 、 $\beta_i=0.914$ 、 $S_i=0.509$ 、 $h_i=-0.497$ 、 $b_i=-0.244$ 。为了继续定义股票的影

响因子，对 X 进行定义，具体如下：

表 5-1 股票收益率影响因子

| 估值因子 | 成长因子 | 资本结构因子 |
|-----------|---------|------------|
| 市净率 P/BV | 净利率 P/R | 固定资产比率 FAP |
| 市盈增长比 PEG | | 流通市值 CMV |

最终单个股票的超额收益率，就可以通过股票收益率的影响因子的加权平均取对数，代入到 R_u 的整体公式中，最终预测股票未来的超额收益率，并在数据库中获得下表：

表 5-2 计算超额收益率及排名

| 股票板块 | 板块系数 | 股票收益率影响因子 (加权平均取对数) | 超额收益率 % |
|-------|---|--|------------|
| 矿业板块 | $\alpha_i=0.0012$ 、 $\beta_i=0.914$ 、 $S_i=0.509$ 、 $h_i=-0.497$ 、 $b_i=-0.244$ | 市净率 P/BV、市盈增 长比 PEG、净利率 P/R、 固定资产比率 FAP、流 通市值 CMV | R1 |
| 燃气能源板 | α_i 、 β_i 、 S_i 、 h_i 、 b_i | P/BV、PEG、P/R、FAP、 CMV | R2 |
| 建筑业 | | | R3 |
| 交通运输物 | | | R4 |
| 流 | | | |
| 金融板块 | | | R5 |
| | | | |

按其超额收益率的大小进行排列，选择其中排名前 20 的股票作为投资对象。同时每个月都需要重新结合模型进行计算，对不符合的股票要进行剔除，并获得新的投资模型，在具体研究的 2016 年 1 月 1 日—2021 年 12 月 31 日，共 6 年的时间内，一共有 72 个研究期，每一期都选出 20 个股票纳入到投资组合中。

2、如果股票的指标符合上述测算标准，则应当在每个月的月初进行购买，并在下一月重新进行测算在月初开盘时予以购买。

3、购买的频率一般来说原则上是每个月都需要调整，即每个月都购入一次，同时还需要根据当天市场的反应来进行调整。

4、对初始的资金为方便计算设定为 100 万元，而仓位则为所有扣除已经投资

的资金后的剩余资金，而对前 20 只股票全部等额权重购入，也即购入的金额是相当的，每个股票都为 5 万元。若对单只股票卖出后，则应当按照相应的金额进行换股。

结合上面设计好的量化投资策略进行计算，得到的相应投资效果如下图 5-3 所示，相关测算的时间区间为 2016 年 1 月 1 日—2021 年 12 月 31 日，共 6 年。选择的为沪深 A 股，基准收益率为长期国债收益率的年化收益率为 4.02%。



图 5-2 2016 年 1 月 1 日—2021 年 12 月 31 日测算结果

根据上面测算结果，6 年获得的收益率数据（扣除交易成本）如下表 5-1 所示：

表 5-3 四因子模型的选股测算收益率结果（扣除交易成本）

| 策略总收益率 | 年化收益率 | 基准年化收益率 | 最大回撤 |
|--------|--------|--------------|-------------------|
| 91.69% | 11.83% | 4.02% | 42.02% |
| Alpha | Beta | Sharpe Ratio | Information Ratio |
| 0.085 | 0.873 | 0.431 | 0.662 |

5、测算结果分析

结合上述测算结果以及相应的数据来看，主要有以下几个方面的结论：

（1）这一四因子模型经测算在 2016-2021 年间的总收益率为 91.69%，而同期的年基准收益率为 4.02%，6 年收益率为 24.12%。投资组合的复合年化收益率为 11.83%，总体来说效果较好。从数据分析来看在熊市以及市场震荡期间的 2018 到 2019 年这一投资策略的回撤结果在能接受的范围内，同时也跑赢了沪深指数，在总体收益上的趋势也趋于稳定。

（2）这一四因子模型的最大回撤率达到了 42.02%，具体时间是在 2019 年的 1 月 31 日到 2020 年的 5 月 29 日之间，这期间沪深 A 股的最大回撤率是 47.52% 左右，但是回撤幅度大则意味着在这期间购入的股票可能面临最大 42.02% 的损失，如果依据这个模型组建基金投资组合的话，还可能会有清盘的风险（基金的止损线一般设置为 70%）。因此最大回撤率需要保持在一定可控的范围内（<

70%)，避免类似清盘这样的风险出现。

(3) Alpha 的值为 0.085，这说明该策略的超额收益率仅为 0.085、其超额收益较弱；而其 Beta 数值为 0.873，说明这一投资策略面临的波动风险与沪深 A 股的指数风险变动方向非常相似；而 Sharpe Ratio 为 $0.431 < 1$ 表示此项投资的风险大于收益，表明其风险大于收益投资价值不大。

(4) 结合测算的具体运算结果来看，应用上述四因子模型选出来的股票组合有部分处于停牌、ST 或者涨停的股票，为了避免其他市场因素不能买入影响投资资金的利用效率，在后续优化的过程之中应当对这些股票数据删除后再进行测算。

5.2.2 量化投资策略的优化

在上述 5.2.1 设计的量化投资模型之中存在收益率一般，同时回撤的幅度也比较大，同时在投资组合之中还存在部分股票停牌、涨停、或者 ST 无法买入的情况，为了避免其他市场因素而影响投资组合的效果，且由于选股是结合本文构建的四因子模型展开，因此本部分主要是针对交易策略进行调整，并结合相关原则进行重新设计。

1、优化原则

具体其交易策略的优化应当遵循以下两个原则：

一是风险控制原则。结合图 5-3 的量化投资测算收益率测算图来看，该投资组合主要存在 Sharpe Ratio 值较小（风险大）、回撤幅度较大等问题。因此在量化投资策略的重新优化设计之中需要考虑到股票投资的风险原则，对最终的投资组合要展开风险控制。

二是程序简单易用原则。由于少数股票存在停牌、涨停、或者 ST 无法买入的情况，因此在本文针对量化投资策略的设计之中对这些股票进行了删除，同时对实际运行之中出现的一些问题进行收集与重新优化。如初始编程思路较为混乱，形成的投资组合的建议重复较多、条理性不强等，对这些内容也进行了优化。

2、优化内容

在程序中重新考虑了停牌、涨停、无法购买等情况，对可能有停牌风险的股票进行剔除，同时在重新的优化设计之中参考大数据指数的波动重新设计股票的交易规则。具体如下：1、结合大数据指数的波动，达到预计的收益率后可以在大数据出现下降趋势的时点进行抛售，在下降区间内减少回撤；2、结合大数据指数的上涨波动，在连续观察数个观测期后予以买入，并在涨幅期间分阶段买入，提升总体收益；3、在整体的股票交易策略设计之中，充分参考大数据指数的波动，

因为这是 A 股市场反应投资者情绪较好的观测指标，后续可以结合大数据指数波动与市场波动的数据进行回归分析，以确定大体上的大数据指数波动与市场波动之间的关系，并结合不同时期的具体情况不断调整模型，以调整交易策略。

5.2.3 量化投资策略的优化结果

1、优化后测算结果

经过上述风险控制原则以及程序简单易用原则，并引入大数据指数作为交易规则的参考指标进行优化后，最终的测算结果如下图 5-4 所示：

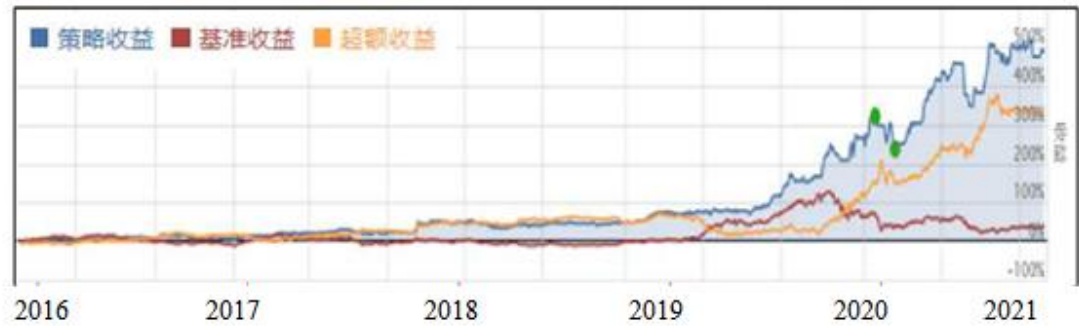


图 5-3 重新设计后的四因子量化选股模型测算结果

经过重新设计优化的四因子量化投资模型的收益率测算结果如下所示：

表 5-4 四因子模型的投资测算收益率结果

| 策略总收益率 | 年化收益率 | 基准年化收益率 | 最大回撤 |
|---------|--------|---------|--------|
| 263.01% | 44.27% | 4.02% | 19.22% |
| Alpha | Beta | 夏普比率 | 信息比率 |
| 0.476 | 0.470 | 1.913 | 1.410 |

2、优化后收益率测算结果解读

结合上述优化后的四因子模型投资策略的收益率测算结果以及相应的数据来看，主要有以下几个方面的结论：

（1）这一四因子模型经测算在 2016-2021 年间的总收益率为 287.13%，而同期的年基准收益率为 4.02%，6 年综合收益率为 24.12%。投资组合的复合年化收益率为 44.27%，总体来说效果较好。从数据分析来看在熊市以及市场震荡期间的 2018 到 2019 年这一投资策略的回撤结果较好远低于同时期沪深 A 股的数据变化，在总体收益上的趋势增长幅度较大。

（2）这一优化后的四因子模型的最大回撤率为 19.22%，远低于同时期沪深 A 股最大回撤率（47.52%）。这意味着在这期间投资组合避免了高达 28.3%左右的

投资损失。同时如果按照这个模型构建投资组合的话，即使购买基金也不会面临清盘，因为其最大回撤为 $19.22\% < 30\%$ ，响应了上述构建的风险控制原则。

(3) Alpha 的值为 0.476，这说明该策略的超额收益率为 0.476，其超额收益较高；而其 Beta 数值为 0.470，说明这一投资策略面临的波动风险与沪深 A 股的指数风险相比要较低；而 Sharpe Ratio 为 1.913 则表明这一投资组合的收益大于风险，是非常值得投资的，应当按这一策略构建投资组合；最终 Information Ratio 的数值为 1.410，也表明在面临主动风险之中仍然获得了 1.410 的超额收益率。

优化后的四因子量化投资模型获得的收益率较高，同时在沪深 A 股有最大回撤的时候也有较好的表现，同时无论是在被动风险还是主动风险情况下，相应的投资选股模型都能够获得丰富的超额收益，这说明本文构建的量化投资模型是值得信赖的。

5.3 投资建议

如果应用上述构建的四因子量化模型来进行实际的投资组合操作，对投资者来说需要注意以下几个方面的内容：

一是，本文构建的量化投资模型由于其收益率整体较大，回撤的幅度也能够控制在一定范围内（低于同期 A 股最大回撤率 47.5%、低于基金止损线 70%），如果是用于投资类似于沪深 300ETF（510300）这类股票的话需要较大的资金量才能够实现可观的收益，如果投入的资金量非常小则有可能最终实现的收益率较弱。此外，这一投资策略其实在很多情况下并非满仓，在量化投资模型之中对股票的选取是遵循一定规律的（一个月内排名最好的前 20 只股票以及其它选股策略如市值等），而购入时间为每个月的月初，这就导致在月中的部分时间内可能一直保持非满仓的装填，闲置的资金其实利用率并不高，浪费了一定的资金时间价值。同时，投资者如果是利用期货来进行交易对冲风险的话，本文构建的四因子模型也能够起到很好的参考作用。例如在仓位未满的时候，如果上述投资策略测算后的行情结果较好，可以对股票做多、对期货做空，两者最终能够实现风险对冲，最大程度降低资金的风险。但同时由于沪深 A 股不能做空、逆向的风险对策无法实现。

二是，本文构建的四因子量化选股模型最终是按照测算的标准公式计算回归系数、表 5-1、5-2 综合市场因子加权平均取对数计算影响因子的水平，最终得到每个股票在当期的未来超额收益率预测，结合收益率数值的大小进行股票的得分排列，并从中筛选出超额收益率最高的前 20 只股票（这一股票组合会在每个月的月初结合四因子模型以及市场表现因子的具体情况进行替换）。而如果投资者

自身的资金量比较小分散到 20 只之中可能资金量更小，在具体操作的时候可能会出现不匹配的情况。因此，对于投资金额在 100 万以下 50 万以上的投资者可以选择投资 10 只左右的股票，而资金量更小的投资者则不推荐使用本文构建的模型进行投资，因为股票数量的减少意味着风险较为集中，在面对市场回撤的时候投资损失的比例可能会更大。

三是，本文构建的四因子量化选股模型虽然是每个月的月初进行换仓，替换表现较差的股票，但是实际上很多公司的数据是有一定延迟的，可能选择的换仓时间刚好不能操作，或者操作的时候刚好涨停出现各类情况。因为在股票市场之中各类信息、投资机会都在不断变化，而各类股票随行情的波动也有巨大差异，在部分行业行情利好的时候另外一些行业则可能出现较大回撤，有些股票可能突然出现市场外的信息导致的涨停等。在这类情况下，对投资者来说如果是已经持仓的情况下遇到较大的涨停，虽然投资策略给出了卖出的决策建议，但投资者也可以再根据市场的实际变化进行观察，在增加一定超额收益后再进行卖出操作，以提升投资收益。

四是，由于本文构建的四因子量化选股模型虽然在历史数据内进行了验证，但是没有在实际的大盘数据中进行预测，也没有投入资金进行检验，对相关历史数据进行测算后，其数据结果非常优秀，但也不能保证在实际的投资过程之中能够做到万无一失，这需要投资者结合选股模型的建议以及实际市场的变化情况进行综合性的分析。为了弥补量化投资策略在实际运行中出现不匹配的情况，在运行之中应当对各类参数展开监控，如果发现异常应当及时的予以调整，对程序进行修改等等。此外，在交易过程之中还需要对各类股票的实际交易情况进行监控，对选取的 20 只股票要实时监测相应数据，发现显著的变化时要及时的调整策略。

5.4 本章小结

本章主要介绍了 Choice 的量化策略平台，以及在具体的量化选股之中的策略，并通过实际数据的测算分析了量化投资策略的效果，在此基础上提出了相应的优化原则、以及最后优化后的量化投资效果。总之本文设计的四因子量化选股模型能够匹配大部分资金量较小的投资者，也能够匹配大型的投资机构等等，但在实际的运行之中需要尽量参考量化选股策略，但同时也需要投资者结合自身的资金情况以及市场的反馈及时的进行策略调整。

第六章 结论及展望

在 2018-2019 年期间,我国散户投资收益率与散户投资亏损程度均较高。在投资市场中,能获得收益并将收益率保持在 10-30%、30%以上的散户,属于极少人群。大部分散户在应收平衡与较大亏损之间。因此由 2018-2019 年散户投资盈亏情况来看,缺乏专业的投资分析策略加持,投资风险较大。而在此背景下,很多投资者开始专项量化投资策略进行选股,当前量化投资策略下各个平均收益率的表现,也表明我国市场金融投资环境愈加清晰并保持良好态势发展。通过量化分析组成的投资产品带来利好的同时,也将吸引更多资金进入市场。据新浪财经报道,在量化策略下的投资行为在投资占比中呈领先态势,但中位数仅有 12.44%,这也表明未来量化产品的发展空间仍然较大。本文的研究目的在于经由我国股市实际发展现状下,结合现实因素构建定价因子模型;通过引入大数据因子对 Fama-French 三因子模型改进,分析其在具体 A 股市场中的表现,并最终结合这一四因子模型展开量化选股构建相应的选股策略,本文通过研究得出以下几个方面的结论:

一、加入了大数据因子后的 Fama-French 三因子模型能够较好的解释 A 股数据,因此在投资决策之中应当参考这一模型的测算结果。

二、在应用新浪大数据 100 因子后发现,这一因子对于绝大多数板块的收益情况都能够作出较好的解释,这也说明应用大数据 100 加入到 Fama-French 模型之中是非常有必要的,也能够对投资者的情绪、市场波动情况进行很好的解释。

三、量化投资策略模型的构建需要考虑风险控制原则,对最终的投资组合要展开风险控制。同时还需要遵循程序简单易用原则,需要对实际运行之中出现的一些问题进行收集与重新优化,优化程序系统的条理性,提升其应用效果。

四、给出了相应的投资建议:选股策略购入时间为每个月的月初,闲置的资金其实利用率并不高,浪费了一定的资金时间价值,可以选择利用期货来进行交易对冲风险;资金量过小(低于 50 万)的投资者则不推荐使用本文构建的模型进行投资,因为股票数量的减少意味着风险较为集中,在面对市场回撤的时候投资损失的比例可能会更大;投资者不要盲目遵守量化投资策略,投资者也可以再根据市场的实际变化进行观察,在增加一定超额收益(10%-20%)后再进行卖出操作,以提升投资收益;此外,在交易过程之中还需要对各类股票的实际交易情况进行监控,对选取的 20 只股票要实时监测相应数据,发现显著的变化时要及时的调整策略。

在未来的应用之中，可以结合实际地操盘来进行量化投资策略的分析，因为只有结合实际的市场变化情况，并且投入资金才能够真正发现量化投资策略以及其相应模型在实际运行之中可能会遇到的问题，以及遇到相应问题的解决方案等。此外，还可以加入对港股、美股等其它国家股票的数据进行测算，以验算本文优化后的模型在其它市场上的具体情况，为投资者的投资决策予以全面的参考。

致 谢

时光荏苒，三年的研究生之旅即将结束。

首先要郑重感谢我尊敬的指导老师，在论文写作方向和专业层面的谆谆教诲，从文献资料到选题，确定论文提纲，把控中期论文方向，解决实验验证问题，以及指导后期格式等各个环节，都以极大的耐心给予了我细致地指导，使我能顺利完成毕业论文，在此，谨向赵武老师致以最诚挚的感谢和最崇高的敬意。

其次，还要感谢求真求实、大气大为的亲爱的母校，各位学术高超、灵魂发光、激情洋溢的老师，各位有趣纷繁、积极向上、知域广阔的可爱同学，清水河畔悠然自得的校宠天鹅，银杏大道飘飘洒洒的纷纷落叶，庄严厚重的主楼，一切的一切，感恩遇见。

十分荣幸，能成为母校千千万万分之一。桃李不言，弦歌百年，愿母校在清水河畔，永续南强之音，再绽至善芳华。

参考文献

- [1] Sharpe W F .Capital asset prices: A theory of market equilibrium under conditions of risk[J].Journal of Finance,1964,19(3):425-442.
- [2] Andrés Ramírez Hassan,Rodriguez M S.Empirical validation of CAPM[J].Ecos De Econom á,2012,16(34):49-74.
- [3] Banz R W,Breen W J.Sample - Dependent Results Using Accounting and Market Data: Some Evidence[J].Journal of Finance,1986,41(4):779-793.
- [4] Fama E F,French K R.Common risk factors in the returns on stocks and bonds[J].Journal of Financial Economics,1993,33(1):3-56.
- [5] Lakonishok J ,Vishny S R W.Contrarian Investment,Extrapolation,and Risk[J].Journal of Finance,1994,49(5):1541-1578.
- [6] Davis H.Davis,H.(1994).Pavlov's bell-less status may still be intact. Psycholoquy ,5 (76), December[J].Psycholoquy,1994,5(2):76.
- [7] Jegadeesh N,Titman S>Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency[J].The Journal of Finance,1993,48(1):65-91.
- [8] Porras D M.Anomalies and testing biases: The CAPM vs.the Fama and French Three-Factor Pricing Model.[D].1998.23-24.
- [9] Andy C.W.Chui and K.C.John Wei.Book-to-market,firm size,and the turn-of-the-year effect:Evidence from Pacific-Basin emerging markets[J].Pacific-Basin Finance Journal, 1998,16(2):6-7.
- [10] Vernon J A,Golec J H,Dimasi J A.Drug development costs when financial risk is measured using the Fama–French three - factor model[J].Health Economics,2010,19(8): 1002-1005.
- [11] Gregory A,Tharyan R,Christidis A.Constructing and Testing Alternative Versions of the Fama–French and Carhart Models in the UK[J].Journal of Business Finance & Accounting,2013,40(1-2): 172–214.
- [12] Brennan M J,Wang A W,Xia Y,et al.A simple model of intertemporal capital asset pricing and its implications for the Fama-French three-factor model[J].Comparative Education Review,2015,9(3):45-46.
- [13] Mahnoor S,Jannatunnesa.CAPM Vs Fama-French Three-Factor Model:An Evaluation of Effectiveness in Explaining Excess Return in Dhaka Stock Exchange[J].International Journal of Business & Management,2017,12(5): 119-121.

- [14] Michael,E,Drew,et al.Firm Size,Book-to-Market Equity and Security Returns:Evidence from the Shanghai Stock Exchange[J].Australian Journal of Management,2003,28(2): 119-139.
- [15] Jegadeesh N,Titman S.Profitability of Momentum Strategies:An Evaluation of Alternative Explanations[J].The Journal of Finance,2001,56(2):46-47.
- [16] Griffin J M,Ji X,Martin J S.Momentum Investing and Business Cycle Risk:Evidence from Pole to Pole[J].Journal of Finance,2003,58(6): 2515-2547.
- [17] Fung K W,Lam K,Lam K M.Do the prices of stock index futures in Asia overreact to U.S.market returns?[J].Journal of Empirical Finance,2010,17(3): 428-440.
- [18] Gregory A,Tharyan R,Christidis A.Constructing and Testing Alternative Versions of the Fama–French and Carhart Models in the UK[J].Journal of Business Finance & Accounting, 2013,40(1-2): 172–214.
- [19] Novy-Marx R.The other side of value:The gross profitability premium[J].Journal of Financial Economics,2013,108(1): 1-28.
- [20] Aharoni G,Grundy B,Zeng Q.Stock returns and the Miller Modigliani valuation formula:Revisiting the Fama French analysis[J].Journal of FinancialEconomics,2013,110(2): 347-357.
- [21] Fama E F,French K R.A Four-Factor Model for the Size,Value,and Profitability Patterns in Stock Returns[J].Ssrn Electronic Journal,2013..
- [22] Ji Z , Chang V , Lan H , et al. Empirical Research on the Fama-French Three-Factor Model and a Sentiment-Related Four-Factor Model in the Chinese Blockchain Industry[J]. Sustainability, 2020, 12.
- [23] Datta S , Chakraborty A . Tests of the Fama and French Three Factor Model with reference to Industry Cost of Equity : Evidence from India. 2020.
- [24] Shaabani J , Jafari A A . A New Look to Three-Factor Fama-French Regression Model using Sample Innovations[J]. 2020.
- [25] 齐岳,周艺丹,张雨.公司治理水平对股票资产定价的影响研究——基于扩展的 Fama-French 三因子模型实证分析[J]. 工业技术经济, 2020, 39(4):10.
- [26] 施昊天. Fama-French 三因子模型适用性及风险警示对组合超额收益的影响探究——基于上证主板数据的实证研究[J]. 现代商贸工业, 2022, 43(1):5.
- [27] 付巍巍,张立卫.基于异质信念的 Fama-French 三因子模型实证与改进[J]. 统计学与应用, 2020, 9(2):11.
- [28] 陈浪南,屈文洲.资本资产定价模型的实证研究[J].经济研究,2000(04): 26-34.
- [29] 黄兴旺,胡四修,郭军.中国股票市场的二因素模型[J].当代经济科学,2002,024(005): 50-57.

- [30] 范龙振,余世典.中国股票市场的三因子模型[J].系统工程学报,2002.
- [31] 张祥建,谷伟,郭岚.上海股票市场"规模效应"的实证研究及原因探析[J].大连理工大学学报(社会科学版),2003,024(004): 24-28.
- [32] 刘蕾,马栋.上海股市规模效应和价值效应实证分析[J].统计与信息论坛,2004(03): 81-86.
- [33] 刘沙.引入住房消费的改进型资产定价模型研究——基于股票收益率横截面的实证比较[J].广东金融学院学报,2012(03): 23-34.
- [34] 李泉,陈雪花.股票市场 CAPM 和三因子模型分析——兼议中国股市的未来发展[J].发展研究,2015,1(6): 43-48.
- [35] 王源昌,汪来喜,罗小明.F-F 三因子资产定价模型的扩展及其实证研究[J].金融理论与实践,2010,1(6): 45-50.
- [36] 李体委.Fama-Fama-French 三因子模型的改进和对中国股市收益率的检验[D].山东大学.2011
- [37] 陈信元,张田余,陈冬华.预期股票收益的横截面多因素分析: 来自中国证券市场的经验证据[J].金融研究,2001,1(6): 22-35.
- [38] 范龙振,余世典.中国股票市场的三因子模型[J].系统工程学报,2002.
- [39] 杨炘,陈展辉.中国股市三因子资产定价模型实证研究[J].数量经济技术经济研究,2003(12): 137-141.
- [40] 陈守东,韩广哲,荆伟.主要股票市场指数与我国股票市场指数间的协整分析[J].数量经济技术经济研究,2003(05): 124-129.
- [41] 吴世农,许年行.资产的理性定价模型和非理性定价模型比较研究——基于中国股市的实证分析[J].经济研究,2004(06): 105-116.
- [42] 田利辉,王冠英.西方资产定价学说和我国股票市场发展[J].南开学报(哲学社会科学版),2014,3(2):9.
- [43] 周骆,张德礼.中国股票市场的有效性实证研究——基于 Fama-French 五因子模型[J].经营管理者,2016(5): 22-22.
- [44] 高春亭.五因子资产定价模型及其在我国证券市场的应用研究 [D].重庆大学,2016.
- [45] 欧阳志刚,李飞.四因子资产定价模型在中国股市的适用性研究[J].金融经济学研究,2016(2): 84-96.
- [46] 赵胜民,闫红蕾.内地与香港股票市场的一体化进程研究[J].中国经济问题,2016,294(1): 124-135.
- [47] 宋光辉,孙影.基于行业市盈率的股票估值 APT 及有效性分析[J].财会月刊,2017, 5(29):61-63.