

上证新兴产业（000067）指数增强量化策略设计——基于 Fama-French 三因子模型

葛栩滢

上海大学，上海，201899；

摘要：随着金融工程学的进一步发展，金融数据已成为金融产品研发与组合的关键。专业的发展有筛选股，择时，资产配置等。本文以此为基础，将现代投资理论中的多因子模型引入到股票市场中，构建股票市场中以收益率为盈利指标的量化模型，并给出相应的股票选择策略。该策略通过对指数跟踪的标的股价在复权之前的收盘价来计算其收益，并设定交易参数等指标，以主动管理的方式，以 50 支成份股为候选，每季度动态买卖个股，不多于 10 支。在 2013 年 12 月 31 日至 2023 年 12 月 31 日的回测区间内，本交易策略 10 年间累计获得的收益率为 111.94%。

关键词：Fama-French 三因子；指数增强；量化投资；仓位控制

DOI：10.69979/3029-2700.25.05.004

1 引言

1.1 研究背景与意义

1.1.1 研究背景

定量投资因其可以解释“主动管理悖论”而越来越受到人们的关注。它的实质就是与基本面分析相结合，通过对资产种类的选择（如资产配置）、交易进场时机的把握以及特定目标下进行的风险控制等，建立一个能够合理地分散风险，获取超越基准收益率的投资策略。从社会层面来看，ETF 指数基金便是其中之一，其能够按照一定的理念对基金进行合理配置，从而为基础的投资人提供一个更好的投资途径。在运用数学和计量经济学等定量工具时，通常还可以进行投资、风控、监管等方面的细分。

目前运用定量方法进行投资的方向以套利和高频为主，今后应该在基本面分析等更为广泛的领域展开应用研究。对于国内投资者来说，他们对于被收取的管理费用率并不像国外成熟市场上的投资人那样敏感。换言之，他们的管理费用受限于国内基金不稳定的收入来源以及行业内的竞争等，这些都造成了资源的使用效率低下，体现在当下则表现为基金缺乏规模效应、行业竞争以及资本偏爱等市场机制。在今后的研究中，机构投资者将更加重视与市场发展阶段相关的产品结构以及投资资本的流向。

指数追踪是指在特定的金融市场中，通过对某个特

定的市场指数进行模拟，从而实现与指数相近似的投资组合，如完全复制指数，便是一种典型的被动投资策略。在此基础上，建立了相应策略的模型。EMH 理论认为，市场指数往往具有更高的回报，历史数据的回测也表明，大部分主动管理基金并不能取得优于市场整体的长期回报，因此，以指数跟踪为基础的基金投资策略在国际上得到了广泛的响应。

指数追踪存在的问题是，如要尽量接近基准指标，就需要进行高频的交易，与之伴随的则是高额的交易资产，这也就意味着建仓成本就会很高。此外，当股市下滑时，消极的投资策略必然会面临亏损。因此，基于指数追踪模型，本文提出了一种增强型指数追踪模型，以实现超额回报。虽然以“自适应”的有效性为代表的指数追踪模型的理论基础并不多，但是从本世纪开始，特别是最近十年，已经涌现出了许多关于增强指数跟踪的模型，并在多国金融市场上得到了实证应用。

1.1.2 研究意义

从理论角度来说，增强型指数追踪策略及相关模型仍未得到广泛应用，也即市场上的公开交易，所需成本也较高。因此，本文结合 FF 三因子模型，拓宽指数增强策略应用领域。

从现实角度来说，对上证新兴产业指数进行全周期的追踪能够为量化投资以及相关的机构交易者和监管者提供决策支持，为完善与改进相关策略在本国市场的应用提供参考。

1.2 研究方法 with 内容

1.2.1 研究方法

本文采用的是文献研究法和实证研究分析法。通过对已有的相关文献进行梳理，就本文的研究问题，使用 FF 三因子模型构建增强型指数并制定交易策略，为下文的实证回测部分奠定理论基础。

1.2.2 研究内容

本文旨在验证三因子量化投资策略对上证新兴产业指数的拟合力度，总体可分为引言、文献综述、理论分析及策略构建、实证分析、结论与建议以及参考文献、附录七大部分。其中理论分析与策略构建部分会阐述因子模型构建、因子选择以及交易逻辑，随后会在实证分析部分进行检验。最后进行总结与建议，希望有助于量化投资在中国市场的未来发展。

2 文献综述

孟叶等人（2018）实现了结合基本面和技术分析的研究，以 A 股上所有上市公司的股票作为样本数据集，采用 AdaBoost 算法结合近邻传播聚类模块对算法进行集成优化，再对数据集进行训练。该模型的目的是为了在 A 股市场上挑选出若干只在未来会大幅上涨的股票。最后研究得出基于基本面指标和技术指标的量化模型在 A 股市场获取超额收益有效；并认为 KDJ 指标在该研究中的择时效果最佳。张若凡（2019）将 World Quant 提出的 101 个有效因子放入 A 股市场中进行实证研究，结果发现其中的 16 个因子在国内依然生效。该文作者通过降低交易方法成本进一步提高了模型胜率和策略收益，主要通过对换手率进行优化目标函数等方法。Li u J、Stambaugh 和 Yuan Y（2019）设计了一个以 FF 三因子模型为模型基础的 A 股三因子模型。传统的三因子模型对于国内市场的解释力度较低，改良后的模型很好的解释了 A 股收益率截面的有关异象。王春丽等人（2018）选用了上证 180 指数的成分股，将择时策略运用于多因子的选股模型，用于解决因子贡献收益率的不稳定性。确保收益的同时，进一步控制风险。该研究结果发现，在多因子模型中加入择时后。仍然能取得超额收益。其次，在技术因子层面上，股票的收益率因子几乎无效。最后也是最为关键的，量化择时的加入在更大程度上会起到降低回撤、控制风险的效果，尤其是当市场处于震荡阶段时，其控制风险、稳定收益率的效果更明显。

3 理论分析与策略构建

3.1 收益模型

近代经济学家提出的资本资产定价模型（CAPM），在一定的假设条件下可以被量化投资工程师进行对收益的解释，为后人的投资实践具有重要的理论意义。

$$R_i = \beta_{1,i}X_{1,i} + \beta_{2,i}X_{2,i} + \dots + \beta_{n,i}X_{n,i} + \varepsilon_i \quad (1)$$

其中， $X_{j,i}$ 是第 i 支股票在第 j 个因子上的收益；同理， $\beta_{j,i}$ 是股票在因子上的暴露。而 CAPM 模型是多因子模型的特例，表现为市场收益的单一因子；结合主动管理投资模型对因子暴露度定义。

$$r_n(t) = \sum_k x_{n,k}(t) \cdot b_k(t) + u_n(t) \quad (2)$$

其中， $r_n(t)$ 为资产 n 从时刻 t 到 $t+1$ 的超额收益率， $x_{n,k}(t)$ 为时刻 t 时刻资产 n 对因子 k 的暴露值， $b_k(t)$ 为因子 k 从时刻 t 到 $t+1$ 的因子收益率， $u_n(t)$ 为资产 n 从时刻 t 到 $t+1$ 的特异收益率。以上，等式右边建立了个股的因子值和相对应的暴露度，与左边组合投资的超额收益。该模型在现代投资理论框架为延伸，使用的套利定价理论引入了无套利条件，从数学上解释了非系统性风险被分散的原理，还在满足的一定条件下，获得了能够定价的价值因子关系式。

3.2 选股策略

本文沿用了市场经验发现平均收益率较高的选股因子、FF 三因子：P/E 市盈率、PB 市净率和无风险收益率，得到公式（3）：

$$R_i = a_i + b_i R_M + s_i E(SMB) + h_i E(HMI) + \varepsilon_i \quad (3)$$

其中， $R_i = E(r_i - r_f)$ 为股票 i 针对无风险投资下的、所能够获得的数学期望超额回报率； $R_M = E(r_M - r_f)$ 为市场环境下就无风险投资的数学期望超额回报率， $E(SMB)$ 是相对大市值公司股票下，小市值公司的数学期望超额收益率， $E(HMI)$ 则是账面市值比高的公司股票比起账面市值比低的公司股票的期望超额收益率， ε_i 则是残差回归项。

3.3 交易逻辑

调仓月选取上月末的相关因子数据和前复权的收益率，按三因子的大小对股票池内分成：“大中小”3 个层次（分组），并分别计算本月同层内股票的平均收益率，之后用不同层股票的平均收益率做差。获得的层间差值大小实质上可以等价于：衡量了选取的因子是否

具有收益效应（作回归）。通过以上过程，完成了因子收益率的计算；且通过分层的方法，使得因子间收益率的差异具有可比性。

实验假设的 FF 三因子模型中的风险暴露度均能解释上证新兴产业指数下追踪标的的超额收益，那么模型指标中的 α 数学期望为 0。如果观测周期内的股票，出现模型的截距项小于 0，说明这段时间里面收益率偏低且根据均值回归，交易系统则按事先设定的参数获取 α 最小并且小于 0 的 10 支股票进行买入，若少于 10 支则全部买入。

表 3.1 抽样方式

大市值股票（50%）	小市值股票（50%）
大账面市值比从大到小前 30%	大账面市值比从大到小前 30%
中账面市值比从小到大 30%~70%	中账面市值比从小到大 30%~70%
小账面市值比从小到大后 70%	小账面市值比从小到大后 70%

4 实证分析

4.1 样本选择与数据来源

本文的被拟合样本为 2013 年 12 月 31 日至 2023 年 12 月 31 日的上证新兴产业指数，相关数据来自 CSMAR 数据库。该指数成立的时间为 2010 年 4 月 30 日，但是

前期的数据在回测平台的数据库中缺失。因此，为了便于检验分析，本文将回测起始时间修改为 2013 年 12 月 31 日，保证了回测期内市场波动的全面性和每年度回测分析的完整性。

各季度进行拟合的 10 支股票为当期的上证新兴产业指数成份股，每季度初进行移仓换股。期初资金为 100 万元。手续费率为 0.01%。

4.2 回测结果与分析

本文实证分析采用的是掘金量化回测平台和 Python 3.11.5（Anaconda 3），主要通过年化收益、最大回撤、夏普比率、卡玛比率、相关系数、跟踪误差等指标来衡量策略结果。具体的拟合走势图详见附录。

4.2.1 年化收益

对比各年度的年化收益率和基准收益率可以发现，2014 年和 2017 年的拟合收益率显著低于基准收益率，2016 年、2018 年、2019 年和 2023 年的拟合收益较为相近，其中 2016 年仅相差 0.07%，其余年份的拟合收益率均显著高于基准水平，在 2020 年达到最高点 55.73%。这也使得十年期的投资策略优势从 2020 年开始逐步显现，最后拟合的累计收益率达到 111.94%，显著高于 26.89% 的基准水平，年化收益率止步 11.20%。

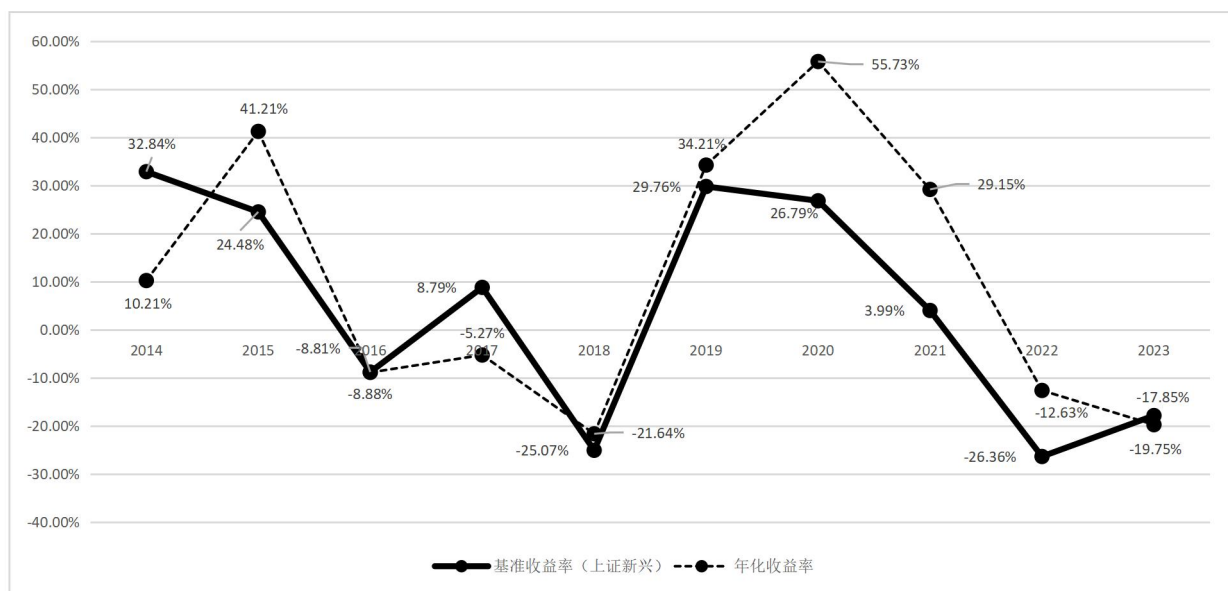


图 4.1 2014~2023 年年化收益率对比图

4.2.2 最大回撤

最大回撤是指在选定周期内的任一时刻点往后推，产品净值走到最低点时的收益率回撤幅度的最大值，用于

衡量策略的抗跌能力。简要说，首先便是选定周期，而后寻找到周期内的最低值和最高值，最大回撤的计算方式之一便是用两者之差去除以最高值，得到的结果越

小越好。从回测结果来看，除了 2015 年高达 50.53%，其余年份均低于 30%，其中 2014 年、2017 年、2019 年至 2021 年更是低于 20%。相较于 2015 年表现出的风险大收益高，2019 年至 2021 年则表现得更为稳健，风险更低。

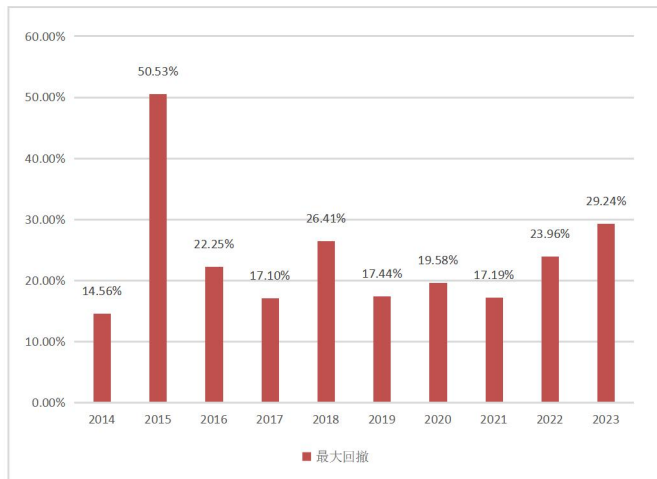


图 4.2 2014~2023 年最大回撤对比图

4.2.3 夏普比率

夏普比率的定义为策略的年化收益率与无风险回报率之差除以策略回报率的标准差，可以理解多承担一单位的风险，该策略所能够获得的超额回报。因此，夏普比率通常期望为正，并且值越大，意味着每付出一单位的风险所获得的风险也越大，也即性价比越高。反之，若是夏普比率极小甚至为负，相应的超额回报也就很小甚至会造成损失。从回测结果来看，2016 年至 2018 年和 2022 年至 2023 年的夏普比率为负。但拟合指数的走势图显示，除了 2017 年度本策略的拟合结果不敌上证新兴产业指数，其余年份的表现均优于原指数。由此结合来看，夏普比率的负值在一定程度上是原指数的表现不佳造成的，这反而验证了本策略的拟合是较为相近的。

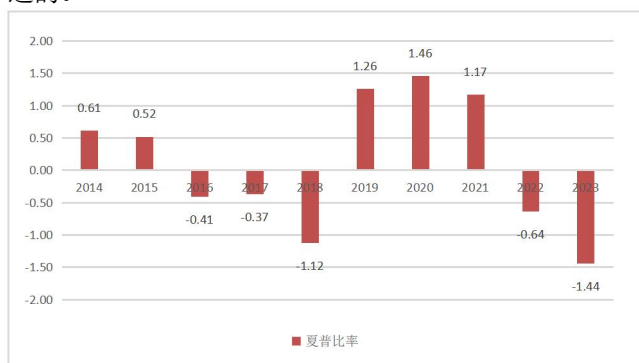


图 4.3 2014~2023 年夏普比率对比图

4.2.4 卡玛比率

卡玛比率的计算方法为超额收益除以最大回撤，其与夏普比率的不同在于分母，但本质上均是衡量投资组合的回报与风险的关系。因此从回测结果中也能够看出，两者并无显著区别，因此不再赘述。

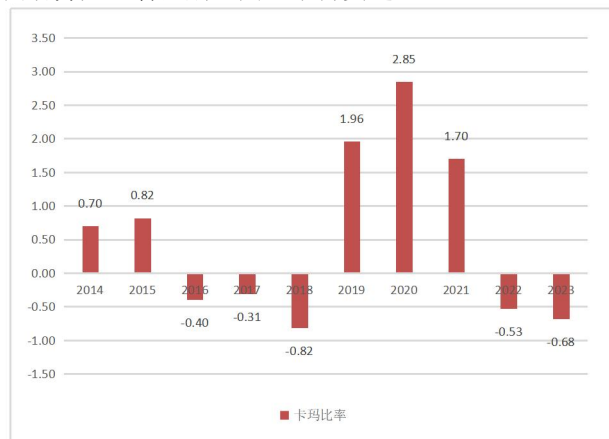


图 4.4 2014~2023 年卡玛比率对比图

4.2.5 相关系数

相关系数通常用来衡量两个变量之间的线性关系，计算的公式为两者的协方差除以两者各自的标准差。从回测的结果来看，仍旧是从 2020 年，更为准确的是 2020 年的下半年开始，拟合的指数与原指数逐渐拉开差距。但依据形状判断，其并不是由于 beta 的偏离造成的拟合度降低，而是主动投资的策略迎合了市场，使得 alpha 逐渐展现。

4.2.6 跟踪误差

跟踪误差，或称为追踪误差，主要观察两个方面，一方面是相较于基准收益率的差值，另一方面是相较于基准收益率的标准差，可以较为直观地衡量基金追踪的拟合程度。从回测的结果来看，2016 年至 2019 年的差值较小，拟合结果较好。但整体的波动较大，这其中有跟踪误差计算方法本身选择的原因，也有投资策略的原因。前者衡量指标站在的时刻点为每年度末，相对静止，相较于需实时比较的追踪表现存在一定的偏差，如 2015 年便容易在此标准下被排除。后者不是单纯的被动投资，而是为增强指数采用了主动投资策略，2019 年 28.94% 和 2021 年 25.16% 的差值便验证了本策略在该时间段的适用性。即便在 2022 年和 2023 年因追踪原指数而表现不佳的情况下，本策略的十年期累计收益率仍能够维持在破百的水平。



图 4.5 2014~2023 年收益率差值对比图

5 结论与不足

5.1 结论

综合以上衡量指标的分析结果来看,即便在每个季度都是按照同一的代码与标准挑选后进行买入和卖出的操作,在实行策略之后指数的拟合表现在不同年度是截然不同的,此处的不同既有拟合方面又有策略自身收益方面。在这其中能够造成差别的有两个原因,一方面是原指数的 50 支成分股在变化,另一方面便是拟合指数的 10 支成分股也在变化。因此,在衡量该指标的优劣之时,需要将其对原指数的拟合度和其回测结果本身两方面全面地进行分析与思考,才能不失偏颇。

5.2 不足

对于本文所采用的策略仍需要改进之处,首先便是需要权衡拟合的效果和追求正收益。如果过度拟合,但年度的收益有可能会因原指数表现不佳而收获负值。这可能涉及到更为根本的层面,便是被跟踪指数的挑选。

若是假设指数已定,则回到策略的制定上来。由于在各个交易软件中未找到公开交易的跟踪上证新兴指数的基金,其他跟踪指数的基金也都为被动跟踪,所以本策略并未持仓公募基金,也没有交易债券,而是单一地选择了指数的成分股。从回测结果也可看出,虽然最终获得了良好的收益,长期持有也能够有不错的年度收益率,但是在指数表现不佳的年份,本策略没有能够有效抗跌或避险的能力。反而是因为持仓股更少、更集中,

年化收益显著低于原指数的基准收益。

由于策略的制定最终是为了未来的交易,但是本策略仅仅对历史数据表现进行观测,无法判断是否能预防黑天鹅事件等。因此,在指数增强策略中,对于实盘跟踪标的,必须对持仓股票的价格区间设置止盈和止损点。同时,量化模型在实盘操作时应谨慎根据回测的历史数据进行模型的参数的估计,以避免模型的收益率在未来时间序列中出现意外回撤。

总而言之,上证新兴产业指数所对标的新兴产业是未来经济发展的主要内容,具有良好的投资价值。但也要注意,相对于大盘,新兴产业股指波动性更大,投机性因素更多,再加上其自身产业特性以及市场内外环境的影响,投资新兴产业面临较高风险。

参考文献

- [1]古志婷,宋泽芳,李元.基于 LASSO 变量选择与多因子模型的增强型指数基金的构造研究[J].数理统计与管理,2020(03):417-428.
- [2]黄金波,吴莉莉,尤亦玲.基于 UPM-LPM 的增强指数投资策略[J].中国管理科学,2019(09):26-35.
- [3]孟叶,于忠清,周强.一种安全的量化选股策略——来自 A 股市场的实证[J].金融理论与实践,2018(08):102-107.
- [4]王春丽,刘光,王齐.多因子量化选股模型与择时策略[J].东北财经大学学报,2018(05):81-87.
- [5]Fama E F, French K R. Common risk factors in the returns on stocks and bonds[J]. Journal of Financial Economics,1993,33: 3-56.
- [6]Liu J, Stambaugh R F, Yuan Y. Size and Value in China[J]. Journal of Financial Economics, 2019,12: 63-77.

作者简介:葛栩滢(出生年月:2000 年 10 月),性别:女,民族:汉,籍贯:浙江省宁波市人,学历:硕士研究生,单位:上海大学,研究方向:证券投资。