



主被动融合：基于 XGBoost 算法的策略型债券指数

罗东原 刘义伟 曹轶岚 杨启凡 宋智鑫 王馨平

摘要 从国际债券指数的编制经验来看，主动与被动投资融合是大势所趋。本文基于全市场最活跃的国债及政策性金融债编制了 XGBoost 多因子活跃券久期择时多头策略指数。首先通过流动性加权方式构建五个关键期限的活跃券底层指数，解决了因成分券流动性不足指数可复制性和可交易性较差的问题；然后利用集成学习中的 XGBoost 算法构建多因子模型进行久期择时，实现了指数整体回撤可控、超额收益可观的效果。策略型债券指数在实务中可以收益凭证、收益互换、场外期权等交易模式为载体，满足不同类型投资者的债券投资需求。未来可从引入票息作为成分券考量因素之一、由多头向多空策略转变等方向进一步研究增厚在不同市场环境下的收益。

关键词 策略型债券指数 XGBoost 算法 多因子模型

一、引言

从市场组合、Fama-French 三因子到因子动物园，主动管理获得的超额收益越来越多可以被因子暴露解释^①，指数化投资方兴未艾。但关于债券是否比其他大类资产更容易获得 alpha 的问题始终存在争论。

Barlow & Shu (2021) 认为债券指数化投资发展相对缓慢的原因在于债券基金主动管理表现优于股票基金。仅 20% 的美国股票基金能够超越基准，而在投资级短债基金等部分板块的债券基金中，超越基准的比例达到 53%，股债之间指数化投资规模的分化是投资者“用脚投票”的结果。

而 Brooks et al. (2020) 则认为过去 20 年间债券主动管理人能获取 alpha 只是一种幻觉，久期风险、信用风险、新兴市场风险和波动率风险四种因子暴露能够解释 68%~90% 的主动管理收益，而 alpha 不显著。此外，主动型债券投资对公司信用风险的暴露高意味着其中高收益债券的占比一般高于基准指数，这提高了债券基金与股票资产的相关性，降低了大类资产配置中加入债券资产能够带来的分散化效果。

指数化投资的优势不言而喻。主动型债券基金的年周转率通常在 100% 以上，而指数基金仅 40% 左右，大大降低了交易成本 (Fabozzi & Mann, 2005)，传统指数

罗东原，国泰君安副总裁、机构与交易业务委员会总裁；刘义伟，国泰君安固定收益外汇商品部总经理；曹轶岚，国泰君安固定收益外汇商品部投资业务部总经理；杨启凡，国泰君安固定收益外汇商品部交易经理；宋智鑫，国泰君安固定收益外汇商品部交易经理助理；王馨平，国泰君安固定收益外汇商品部助理。

① Bender et al. (2014) 称管理人获得的超额收益中 80% 可以被因子暴露解释，只有 20% 来自获得 alpha 的能力。



型 ETF 中甚至已经出现了零费率产品。^①

债券总收益指数 20 世纪 70 年代从美国发端,当时标的还仅限于投资级债券,80 年代开始编制高收益债券指数和国际政府债券指数。而后历经利率和信用风险升降的周期,债券指数化已经被证明能提供更具有竞争力的收益并且风险也较低 (Fabozzi & Mann, 2005)。新冠疫情后出现了宏观经济波动加大、发达国家通胀高企、货币政策紧缩、低碳转型加速等新变化,主流指数商积极响应推出了通胀挂钩债券指数、气候风险政府债券指数、ESG 投资级债券指数等。

近年来,主动与被动的界限逐渐模糊 (胡玉婕, 2018),主动与被动融合是未来债券指数创新的重要方向之一。主动型固定收益 ETF^② 帮助主动管理人向更透明、更低成本的投资方式转型,本质上与被动跟踪策略型债券指数的 ETF 类似。Smart Beta 债券指数介于主动和被动投资之间,是海外指数创新的重要方面,已经推出的因子策略指数包括基本面多因子、动量因子、ESG 因子等。债券指数通过与量化策略融合,正在从业绩基准转变为内涵更丰富、底层资产更多元的投资工具。

国内从 21 世纪初开始发布债券指数。2002 年,中国外汇交易中心推出银行间债券指数,中央结算公司推出 12 只债券指数。目前中央结算公司、中证指数公司、中国外汇交易中心等机构已经推出了较为完善的指数族,覆盖不同期限、品种和信

用评级,工具型债券指数较为完备。在指数创新方面,国内涌现出了众多绿色债券指数和 ESG 指数,例如中债-碳中和绿色债券指数、中证-沪深 300 ESG 信用债指数、CFETS ESG 高等级信用债指数。策略型指数的存量相对较少,Smart Beta 指数比如中债-信用债价值因子精选策略指数, alpha 策略指数比如中债-国泰君安久期配置策略指数和久期稳健策略指数。

在梳理国际成熟金融市场债券指数发展脉络的基础上,本文旨在探索策略型债券指数编制新的切入点,利用机器学习中的集成学习算法赋能利率择时,形成一条标的为国债及政策性金融债的 alpha 策略指数,以丰富多品类、多层次、多策略的中国债券市场产品供给,服务债券市场投资者多元的投资需求。传统债券指数中包含的样本券众多,在实际的指数化投资中只能在考虑个券流动性的基础上采用抽样复制的方法进行跟踪,放大了跟踪误差。本文将动态地基于活跃券篮子编制债券指数,减小样本量的同时增强可复制性,以规避可能阻碍债券 ETF 套利机制的因素,完善市场定价,促进底层资产二级市场流动性。

二、国际债券指数与国内债券投资策略发展

(一) 国际债券指数发展脉络

资本资产定价模型中,均衡条件下最优风险资产组合是市场组合;在金融市

① 比如 BNY Mellon Core Bond ETF (跟踪 Bloomberg US Aggregate Total Return Index) 和 Goldman Sachs Access Emerging Markets USD Bond ETF (跟踪 FTSE Goldman Sachs Emerging Markets USD Bond Index)。

② 根据晨星数据,主动型固定收益 ETF 总资产规模从 2012 年底的 96 亿美元快速攀升到 2023 年近 1 500 亿美元。



场上，投资者用市值加权指数代表市场组合。因此在债券指数推出之初，市值加权自然而然成为 20 世纪 70—80 年代主流的指数编制方法。随着学术研究的推进，20 世纪 90 年代出现了风格因子，包括价值、成长和规模，但应用仍集中在权益资产。2000—2010 年，债券指数编制中开始出现另类加权方法，以基本面加权为主。2010 年以后，固定收益领域关于因子投资的研究增加，Smart Beta 开始萌芽。其实不论等权、基本面加权、风险加权还是因子策略，只要不同于市值加权的另类加权方式都可以归类为 Smart Beta。

策略型债券指数也属于另类加权，策略思路较为多元，包括基本面策略、利率对冲、恒定久期、阶梯策略、收益增强、ESG、动量策略等。Smart Beta 是创新发展的重点，且该类指数的策略思路透明度高，能够直接指导指数编制过程中的权重确定。2021 年安硕发布了 iShares USD Bond Factor ETF，底层指数同时融入了宏观择时与微观择券技术，综合多种 Smart Beta 指数常见的因子策略和优化手段，已经是一种更接近主观投资的 Smart Beta 产品。以下简要介绍另类加权债券指数的几种主要类型（附表 1）。^①

①恒定久期指数。恒定久期指数旨在保持投资组合久期稳定，使利率风险可控。根据国际债券指数编制经验，该类指数久期多以中短期为目标期限。

②阶梯策略指数。市场上常见的三种债券投资组合为哑铃形、子弹形和阶梯形策略，阶梯策略指数正对应其中第三种，将成分券期限均匀地分布在各个期限。

③利率对冲指数。利率对冲指数使用卖空现券、做空国债期货、做空利率互换等方式降低组合久期，甚至将组合久期调整为负数，用于应对市场大幅调整对营收的冲击。

④基本面策略指数。多因子策略在 Smart Beta 指数中也得到了推广应用，策略出发点是通过多因子识别发行人的基本面价值、市场估值水平及流动性状况，认为价值因子应该能够带来长期回报。

⑤ ESG 指数。ESG 投资策略无疑是债券 Smart Beta 指数中成长最迅速的一支。ESG 投资策略一方面通过挖掘可持续发展企业提升组合的长期收益，另一方面规避黑天鹅风险、降低组合回撤。

⑥动量策略指数。Asness et al. (2013) 在全球金融市场范围内检验各类资产中的动量效应，发现动量无处不在。策略型债券指数中也出现了利用动量因子分配成分券权重的思路，比如历史收益越高的债券品种权重越大。

⑦收益增强指数。另外还有一些以收益最大化为目标的收益增强债券指数。比如 FTSE Nomura Carry and Roll Down World Government Bond Index Series 以最大化底层债券的持有收益和骑乘收益为目标。

（二）国内债券自营投资策略演进

①杠杆投资。2014 年后金融机构加杠杆进程加速。2017 年人民银行、原银监会、证监会、原保监会发布了《关于规范债券市场参与者债券交易业务的通知》（银发〔2017〕302 号）明确了各类机构负债端的杠杆上限，杠杆投资得到规范。

②套利策略。2013 年国债期货恢复上

① 附表 1~2、附图 1~3、图 1~5 见增强出版，中国知网—《金融市场研究》。



市,衍生出了丰富的套利策略,包括基差策略、跨期套利、跨品种套利等。但随着策略拥挤度上升,近年来套利空间有所收窄(图1)。

③信用下沉。金融危机后,国内地产平台企业大幅扩表,行业回报率高,抵押担保充足,债券投资者能够通过持有大量“高收益低风险”资产获利。但近年来信用违约事件不断浮现,信用下沉策略有所收紧。

④波段交易。随着经济增长模式由高速增长向高质量增长转变,宏观政策由逆周期调节向跨周期调节转变,国内债券或将进入低波动时代(图2)。波段交易的操作空间缩小,亟须向日内高频量化策略转型要收益。

通过梳理国内债券自营投资策略的演进不难发现,杠杆、套利策略、信用下沉、波段交易等一度有效的业务模式都在发展过程中面临新的挑战。证券公司、农商行等交易型机构换手活跃,负债匹配、恒定久期、利率对冲等国际 Smart Beta 指数仍显被动。本土策略型债券指数的编制应该适应国内债券投资策略的新需求,在借鉴海外债券指数化投资经验的基础上,顺应主动与被动投资相结合的发展趋势,以创新方法寻找 alpha。

机器学习算法给投资领域带来巨大变革,国内使用新技术手段赋能债券投资决策的研究也不在少数。周凯文和黄泽松(2020)对比了 AEMLP、LSTM 和 GBDT 算法在国债期货交易中的回测结果,发现树模型表现优于神经网络,并建议在期货交易策略设计中使用遗传算法进行因子挖掘。陆兵和赵辉(2023)使用固定收益分析师情感分析结果作为模型输入数据的一

部分,发现 XGBoost 方法对中债指数的方向性预测效果较好,二分类预测的平均准确率能够达到 75%。施文和高珂(2020)使用随机森林算法预测地产债超涨概率,发现准确率达到 77%,能够切实提升预测效率。现有研究成果普遍反馈树模型具有更优的预测效果,因此本文尝试把基于 XGBoost 算法的债券择时策略编制为一只新的 alpha 策略指数。

三、XGBoost 多因子活跃券久期择时策略指数

(一) 国债及政策性金融债活跃券指数编制

1. 市值加权指数存在的问题

采用市值加权的传统债券指数不论是作为业绩基准还是被动投资工具均存在一些明显的问题,主要有以下两点。一是债券发行量对指数样本结构的影响。Tang & Xu (2017) 指出美国过去 30 年间国债、市政债、公司债、MBS 和 ABS 的市值分布结构发生了巨大变化,比如国债在美国债券市场中的市值占比从 2007 年的 16% 上升到 2014 年的 34%,于是被动跟踪指数的基金也将近 1/3 的资产配置到当时利率处于历史低位的国债当中,使得基金对利率风险的暴露加大。本文编制债券指数的底层资产为国债和政策性金融债。图 3 展示了 2015 年以来国债及政策性金融债在两者余额之和中的相对占比,政策性金融债的相对比例逐年下降,如果直接采用市值加权,对于一些具备税收优势或者追求高流动性的投资者而言可能并不是理想的配置结构。二是传统债券指数样本数量众多,甚至包含大量无流动性个券,对直接复制和跟踪指数造成困难。无流动性个券



长时间无成交，取价存在难度，不同估值机构给出的个券定价差异偏高、做市商报价的买卖价差偏大，使得指数虚拟收益率与实际市场认可的价格存在差距。此外在实际复制指数组合时，股票 ETF 的样本数量占比通常为 100%，但跟踪综合债券指数的债券 ETF 样本最低占比仅 3% (Todorov, 2021)。抽样复制会增加 ETF 的跟踪误差和风险结构偏离，使债券指数实际应用于投资更加困难。

2. 国债及政策性金融债活跃券指数

本文同时考虑债券上市日期和区间成交量，定期调整活跃券篮子，在各个关键期限上编制国债及政策性金融债活跃券指数。活跃券指数关键期限^①包括 1 年、3 年、5 年、7 年、10 年，底层资产包括国债、国开债、口行债及农发债。活跃券指数样本调整频率为月度，考虑到各期限各品种活跃券的特点，规定最近四年内上市、过去一个月成交量排名在各品种个券中排名前 2 位的代码进入指数样本。本文考虑了两种加权方式：等权及流动性加权，等权即在每个调样日等面额配置样本中的 8 只个券；流动性加权选择换手率作为流动性指标，经标准化处理后计算个券权重。

活跃券指数为全价指数，计算公式如下：

$$\text{总市值} = \sum_i [(\text{净价}_i + \text{应计利息}_i) \times \text{权重}_i] + \text{月中派息再投资}$$

月中收到的派息及再投资收益将在月底从指数市值中剔除，在此之前以指数本身的回报率考虑再投资。每月月底为数据提取日，以当天能够获得的指标值重新确定活跃券指数样本及个券权重。次月第一

个交易日为调样生效日，每次调样时需重新确定除数用于调样生效后的指数点位计算。活跃券指数基期设定为 2016 年 7 月 29 日，基期除数为当天确定的活跃券指数的市值。之后每月底重新修正除数的计算公式为：

$$\frac{\text{修正前市值}}{\text{原除数}} = \frac{\text{修正后市值}}{\text{新除数}}$$

当等权配置时，个券 i 面额所占权重为：

$$\omega_i = \frac{1}{\text{活跃券指数样本券只数}}$$

当使用流动性加权时，首先将所有样本券的月换手率进行标准化得到 z_i ，然后通过下式得到个券的流动性分数，将流动性分数在全样本中所占比重作为每一期个券的流动性权重。

$$\text{liquidity_score}_i = \begin{cases} (1+z_i), & z_i \geq 0 \\ (1-z_i)^{-1}, & z_i < 0 \end{cases}$$

由于全价指数计算方式参考中证指数公司编制方法，为对比活跃券指数与市场上现有利率债工具型指数表现，选择中证国债及政策性金融债指数族作为比较对象。以 2023 年 9 月底活跃券指数选择的成分券为例，与中证指数底层个券的可交易性进行对比。分别统计不同关键期限上两指数底层债券个数、2023 年 9 月的区间成交量和月换手率，如附图 1 所示。

在各个关键期限上，中证指数的成分券个数众多且流动性与活跃券相差较大，对复制指数表现造成困难。不过在近年利率水平整体下行的大背景下，新券发行票面利率随之走低，中证指数包含的老券能够提供更厚的票息收益，使得用活跃券编

① 关键期限 1 年、3 年、5 年、7 年、10 年对应剩余期限分别为 (0,5,1)、(2,3)、(4,5)、(6,7)、(9,10)。



制全价指数存在天然劣势。但对于纯多头债券策略指数而言,在牛市环境中活跃券能够受益于更高的流动性溢价从而在资本利得上弥补绝对收益的劣势。

从整体回报表现来看,1年、3年、7年期活跃券指数总回报率超过中证对应期限指数;10年期活跃券指数与中证国债及政策性金融债指数7-10差距较为明显,可能源于中证7-10指数中剩余期限在6.5~7年区间的个券在收益率曲线较为陡峭的区间获得的骑乘收益,而9~10年区间曲线较为平坦(附图2)。

(二) 多因子久期择时策略指数

本文在确定国债及政策性金融债活跃券久期择时策略时,首先针对市场主要关注的各类宏观数据与远期收益率数据分别建立单因子久期择时模型。同时,伴随着机器学习技术的不断发展进步,本文还采取了XGBoost算法构建多因子久期择时模型。通过对比,可以发现XGBoost多因子久期择时模型相较于单因子久期择时模型具有非常显著的择时效果优势,平均年化收益率可达7.48%,夏普比1.538。

1. 策略概述

利率为货币政策、经济景气度、市场流动性、融资需求、通胀等要素综合影响下的定价结果。理论情况下,决定期限利差的主要因素为对未来利率变化的预测,即远期利率。故本文将拟定市场主要关注的各类宏观数据与远期收益率数据作为待筛选指标池,宏观数据包括M2、黑色冶炼行业产能利用率、超额准备金同比、PMI及其分项、PPI及其分项、CPI及其分项,以及社融与信贷数据原值、同比多增与存量同比等。其中超额准备金同比数据需结合信贷收支表等相关数据进行推算方可得

到月度数据;远期收益率数据则直接取用中证指数公司提供的国债远期收益率。

传统多因子模型多以对宏观数据和市场数据进行降维作为主要建模手段,旨在压缩原有数据规模,降低数据矩阵特征向量的维数,以求用最少的维数来概括最重要的数据指标信息。主流建模方法为主成分分析法(PCA)。PCA算法虽然能够有效地减少数据维度,提高数据处理的效率,但是也存在一些缺点,主成分特征维度的含义具有模糊性。PCA算法也一般只用于处理线性关系,而不能处理非线性关系,且各个指标和收益率曲线的线性关系并不显著或具有时变性。伴随着机器学习技术的不断发展进步,机器学习也开始活跃于金融市场,金融机构纷纷开始利用机器学习算法处理数据、预测市场动态和管理风险,各类量化金融产品不穷。量化投资作为机器学习最重要的应用领域,在金融市场中展示出了巨大潜力。本文尝试采取机器学习方法构建多因子模型。

目前应用范围最广的两种基于梯度提升框架的机器学习算法分别是LightGBM(Light Gradient Boosting Machine)和XGBoost(eXtreme Gradient Boosting)。

LightGBM在处理大规模数据集时具有明显的优势,采用直方图算法能够高效地利用多核CPU进行并行计算,加快训练速度。通过GOSS技术,LightGBM能够在保持较高预测性能的同时,显著减少内存消耗。如果面临处理大规模数据集和追求更快的训练速度的需求,选择LightGBM可能是更明智的选择。而对于参数调节和防止过拟合有更高的要求,XGBoost则是一个不错的选择。

XGBoost属于集成学习中的boosting



框架算法，针对分类或回归问题具有非常好的预测效果。该算法本质思想是通过构建多个基学习器使用加法模型，不断学习之前基学习器的数值偏差，通过多个学习器的学习，不断降低模型预测值和实际数值的误差。模型最终的预测值为之前所有基学习器预测值的加总。XGBoost 算法并非通过对残差进行拟合实现，而是通过求解目标函数直接得到的树结构。因为 XGBoost 算法是对损失函数进行二阶泰勒展开，不仅提高了模型准确度，还可自定义损失函数。同时在目标函数中加入了正则项，使 XGBoost 可控制模型的复杂度。正则项降低了模型的方差，使学习出来的模型更加简单，有助于防止过拟合。

基于此，本文将 XGBoost 应用于债券市场活跃券久期择时指数编制中，并基于债券市场远期利率曲线数据和各类宏观数据的组合预测各个期限国债及政策性金融债活跃券指数走势，所取得模型效果远优于单因子久期择时模型的效果。

2. 数据处理

为避免由基本面本身的剧烈变化或者统计原因造成的部分因子数值可能出现的极端数值，本文将因子按正态分布的累积分布函数进行逆变换，将变换后的新数据作为因子的调整值。

$$factor_{new,i} = NorDist^{-1}(SorthIndex(factor_i)/M)$$

其中， $SorthIndex$ 为将因子从小到大升序排列后的序号； M 为因子滚动排序的数据长度； $NorDist^{-1}$ 是均值为 0、方差为 1 的累积正态分布函数的逆变换。

因为夏普比率是普遍用于衡量策略表现的量化指标，本文将采取三分位因子检验法来指定看多或看空的策略执行信号。

三分位因子检验法即依据指标的观察样本数据确定上下两个三分位点，结合该指标的方向逻辑，确定下一期的头寸方向（分别对应看多、看平和看空三个策略方向），进而分析在不同分位点情景下，市场下一期收益率的统计差异，其差异程度可以通过以下统计量衡量：

$$t = \frac{\bar{F}_1 - \bar{F}_3}{\sqrt{\frac{(n_1-1)S_1^2 + (n_3-1)S_3^2}{n_1+n_3-2} \left(\frac{1}{n_1} + \frac{1}{n_3} \right)}}$$

其中， $\bar{F}_1$ 为因子指示看多时，未来一期交易标的收益率的均值； $\bar{F}_3$ 为因子指示看空时，未来一期交易标的收益率的均值； S_1^2 为因子指示看多时，未来一期交易标的收益率的方差； S_3^2 为因子指示看空时，未来一期交易标的收益率的方差； n_1 为看多信号的样本容量； n_3 为看空信号的样本容量。

不难看出 t 统计量越大，该因子发出看多和看空信号未来收益的差异越明显，其预测效果越好。即可以通过 t 统计量来进行因子筛选，将筛选出的优秀单因子再进行多因子建模。

3. 单因子模型久期择时信号产生机制及效果

单因子模型久期择时信号生成机制如下，对因子按正态分布的累积分布函数进行逆变换，对新产生的因子调整值进行差分，当差分大于预设阈值 d ，且 t 值大于 0，则产生看多信号（若 t 值小于 0，则产生看空信号）；当差分小于预设阈值 $-d$ ，且 t 值大于 0，则产生看空信号（若 t 值小于 0，则产生看多信号）；当差分介于预设阈值 $-d$ 和 d 之间，则产生振荡看平信号。由于本



文讨论为期限选择策略, 根据市场观察, 大部分机构久期管理基本在 3 年左右, 故我们设定看平信号表示持有 3 年期活跃券; 看多信号则为增加久期, 即设定看多信号表示持有 10 年期活跃券; 看空信号则为减少久期, 即设定看空信号表示持有 1 年期活跃券。本策略不考虑增加杠杆, 始终保持满仓状态, 只是以月频对持仓活跃券进行期限调整。其中几个单因子策略表现效果如表 1 和附图 3 所示。

4. 多因子模型久期择时信号产生机制及效果

多因子模型久期择时信号生成机制如下, 基于 XGBoost 算法分别预测各个期限国债及政策性金融债活跃券指数预期持有收益率, 在五个期限中选择预测持有收益最高的期限作为债券久期择时的配置期限。同单因子模型一样, 本策略同样不考虑增加杠杆的策略, 始终保持满仓状态, 只是以月频对持仓成分券进行期限调整。

使用 XGBoost 建立回归模型过程中, 需对参数进行调优。具体参数包括 learning_rate (学习率: 控制每个回归树对最终预测结果的贡献程度, 较小的学习率可以使模型更加稳定)、colsample_bytree (列采样比例: 指定每棵树在训练时使用的特征列的子样本比例, 较小的值可以减少过拟合风险)、n_estimators (树的数量: 过多的树可能导致过拟合, 而过少的树可能无法捕捉到数据中的复杂关系, 使用交叉验证来确定最佳的树的数量)、max_depth (树的最大深度: 较深的树可以学习到更复杂的特征, 但也容易过拟合, 将 max_depth 设置为较小的值可以避免过拟合) 等。调参过程中进行交叉验证和评估指标选择 RMSE 和 R_Squared Score。

图 4 上半部分表示各个期限的国债及政策性金融债活跃券指数和加入 XGBoost 多因子久期择时策略后的指数走势对比。在回测区间, XGBoost 多因子期限选择策略表现为夏普比 1.53, 最终全价净值 1.718, 平均年化收益率 7.48%。

5. 策略表现总结

策略表现总结方面, 相比单因子策略, 多因子策略的主要优势是资金曲线更加光滑。光滑的资金曲线意味着可以放更大的杠杆, 也就意味着可以实现更大的资金利用率和绝对收益。加入机器学习的多因子模型相对于单因子模型具有显著优势, 机器学习模型在实现经济数据对利率非线性影响关系计量与预测方面有更优的效果。基于 XGBoost 模型对国债及政策性金融债活跃券的五个关键期限进行久期择时 (图 5), 在回测区间 2016 年 8 月—2024 年 2 月, 该策略实现了最终净值 1.718, 平均年化收益率 7.48%, 夏普比率 1.53。在相同回测区间, 中证国债及政策性金融债指数最终净值为 1.343。在 65.28% 的回测时间区间内, 该久期择时策略月度环比收益率高于中证国债及政策性金融债指数。具体到月度数据, 可以发现 XGBoost 多因子久

表 1 各个单因子久期择时策略效果

因子名称	夏普比率	回测期间最终净值	平均年化收益率
BCI：企业利润前瞻指数	0.616	1.2540	3.29%
社融存量同比 -M2 同比	0.597	1.3358	4.23%
超额准备金同比	0.470	1.2790	3.57%
PMI：主要原材料购进价格	0.617	1.3359	4.22%
M2 同比 -M1 同比	0.720	1.3500	4.38%
SHIBOR_1M	0.576	1.3250	4.10%



择时模型很好地实现了在利率上行周期(2016年8月—2017年12月,2020年4月—2020年11月,2022年11月—2023年1月)大部分月份选择降低久期,即持有1年期活跃券,依靠票息抗住收益率上行带来的净价亏损。在利率下行周期(2017年12月—2018年12月,2019年11月—2020年4月,2021年3月—2022年8月,2023年11月—2024年2月)大部分月份选择增加久期,即持有10年期活跃券博取资本利得、增厚整体收益。且模型在上行或下行周期内产生的信号也较为稳定,对于债券收益率在中枢利率位置振荡时,模型可能会选择中等期限活跃券作为持仓标的券(附表2)。

四、策略型债券指数的应用场景

从实务应用的角度来看,策略型债券指数的应用场景相较于工具型指数更为丰富。传统意义上,工具型指数的主要应用场景集中在业绩分析基准和被动型投资两个方面。而且,受制于国内债券市场单券流动性差异化分布的特征,被动型债券指数投资普遍存在指数复制跟踪误差较大、成分券买卖困难等问题,导致国内被动型债券指数投资产品存量和增速与海外市场相比显著偏低的现状。因此,国内工具型指数的应用被局限在业绩分析基准这一范围。

本文研究编制的策略型债券指数解决了传统工具型指数在应用范畴的两个痛点。第一,活跃券指数的底层编制思路,保证了策略指数较强的可复制性和可交易性,为基于该指数的产品研发提供了极大便利;第二,择时策略的加入为指数走势贡献了较高的 α 收益,对投资者而言,择时策略在超额收益和回撤控制方面都具有更强的吸引力。

近年来,伴随着无风险利率的长期下行趋势,以及大量配置资金的不断涌入,国内债券市场的投资收益显著下行,投资者传统习惯使用的杠杆策略和久期策略受到挑战,息差水平的不断压缩叠加资产负债表限制,使得杠杆策略的收益下降;久期策略则在收益率曲线不断平坦化及宏观政策缩减原则的限制下捉襟见肘。择时型策略指数可以通过以下几种方式为投资者提供更为便捷优质的投资工具。

(一) 收益凭证

在资管新规实施后,证券公司收益凭证凭借其保本特性成为市场上较为稀缺的投资工具。浮动收益型收益凭证的收益组成主要来源于两个部分,一是保底收益,二是内嵌期权收益。其内嵌期权结构可以挂钩的标的品种十分丰富,只要是发行方与认购方能够认可的品种都能成为收益凭证的参考标的。指数产品由于其第三方编制发布的公允性,能够与收益凭证业务契合。认购策略型债券指数收益凭证的投资者除了享受保底收益,还能获取指数上涨所带来的浮动收益,策略型债券指数也同时提供了额外的超额收益。

根据收益凭证的特点,其可以适用的投资者需求类型主要包括:上市公司、国有企业,以及其他机构客户具有的资金理财需求;商业银行、信托、保险、基金等金融机构发行的理财产品;证券公司柜台市场的合格个人投资者。

(二) 收益互换

收益互换作为重要的场外交易模式在股票市场中发展较为成熟,近些年国内固定收益市场的收益互换交易发展也较为迅速,主要以融资性收益互换为主。非银类机构由于资产负债表的种种限制,通过收



益互换资产出表的方式与银行类机构开展收益互换交易,实现投资收益的最大化。挂钩策略型债券指数的收益互换由于其保证金交易的模式,可以在享受策略型债券指数超额收益的同时,获取一定的杠杆收益加成。

根据收益互换的特点,其可以适用的投资者需求类型主要包括:商业银行、证券公司、保险等金融机构的自营资金交易需求;私募基金等机构客户的资金交易需求。

(三) 场外期权

场外期权的标的和条款完全由交易参与双方协商约定,其灵活度决定了投资者在使用场外期权挂钩策略型债券指数的过程中,可以设计更为丰富的结构,以此优化投资者的收益结构。通过期权的非线性结构可以降低潜在投资风险,通过提升参与率等方式可以获取更高的杠杆收益。对于具备衍生品交易资质的机构投资者来说,场外期权是参与策略型债券指数投资最高效的方式之一。

根据场外期权的特点,其可以适用的投资者需求类型主要包括:商业银行结构化存款的风险平盘需求;商业银行、证券公司等金融机构的自营资金交易需求;私募基金等机构客户的资金交易需求。

五、结论与研究方向

本文以探索策略型债券指数编制为切入点,首先通过流动性加权方式构建活跃券底层指数,解决了指数成分券流动性不足,指数可复制性和可交易性较差的问题。构建出的活跃券底层指数与同期限标准指数相比,在流动性和可复制性层面表现出显著优势。其次,以活跃券底层指数为基础,利用XGBoost算法构建多因子模型

对利率债活跃券五个关键期限进行久期择时,有效解决了单因子模型资金曲线波动较大,以及利率的基本面线性模型预测效果较差等问题,构建出一条以高流动性国债及政策性金融债为成分券的alpha策略指数,实现了指数整体收益平稳,回撤可控,超额收益可观的效果。最后,讨论策略型债券指数的部分应用场景,以收益凭证、收益互换、场外期权等交易模式为载体,为投资者提供多品类、多层次的债券指数投资模式。

针对本文中底层指数构建和择时策略研发等过程,我们认为还有以下几点可以作为未来研究的方向。

一是本文中的流动性加权方案,通过将成分券集中在新发行活跃券之中的方式来提升指数本身的可复制性和可交易性,但其缺点在于长期利率下行环境中,活跃券票面利率显著低于同久期老券,对指数的票息收益造成了一定程度的削弱。后续考虑使用替代方案综合考量成分券的流动性和票息收益,增厚指数的静态收益。

二是本文研发的择时策略以多头策略为主,在收益率上行环境中通过降低久期的方式,减少指数的债券多头持仓亏损。该方式的缺点在于,难以有效应对收益率曲线平坦化的熊市环境。后续考虑研发以多空策略为主的策略型指数,以保证指数在收益率上行环境中依然有可能获得正向收益,为投资者提供更为多元化的投资选择。^[N]

学术编辑: 张毓



参考文献

- [1] 胡玉婕.2018年境外债券策略指数及产品研究[Z].2018.
- [2] 陆兵,赵辉.利用情感分析预测中债价格指标[J].债券,2023(03):61-67.
- [3] 施文,高珂.如何使用机器学习预测债券价格? [J].金融市场研究,2020(03):105-113.
- [4] 周凯文,黄泽松.机器学习在国债期货中的应用[J].债券,2022(11):30-33.
- [5] Cliff Asness,Tobias J.Moskowitz,Lasse H.Pedersen.Value and momentum everywhere[J].The Journal of Finance, 2013,68(3):929-985.
- [6] Jennifer Bender,Paul Breff Hammond,William Mok.Can alpha be captured by risk premia?[J].The Journal of Portfolio Management,2014,40(2):18-29.
- [7] Brooks J,Gould T,Richardson S A.Active fixed income illusions[J].Journal of Fixed Income,2020,29(4):5-19.
- [8] Fabozzi F J, Mann S V.The handbook of fixed income securities[M].New York: McGraw Hill,2005.
- [9] Michele Barlow,Yichan Shu.Fixed Income Indexing:Additive in a Global Multi-Asset Portfolio[R/OL]. <https://www.ssga.com/library-content/pdfs/insights/adding-indexing-to-active-fixed-income.pdf>.
- [10] Tang H,Xu X E.A Review of Fixed-Income Indexing and Index Investing[J].The Journal of Wealth Management,2017,19(4):85.
- [11] Todorov K.The anatomy of bond ETF arbitrage[R].BIS Quarterly Review,2021.

Development of a Strategic Bond Index Based on the XGBoost Algorithm

LUO Dongyuan¹ LIU Yiwei² CAO Yilan² YANG Qifan² SONG Zhixin² WANG Xinping²

(1.Institutional and Trading Business Committee of Guotai Junan;

2.Fixed Income, Currency and Commodities Department, Guotai Junan)

Abstract Judging from the experience of compiling international bond indices, the integration of active and passive investment is a general trend. In this paper, a multi-factor active bond duration-timing long strategy index is compiled based on the most active treasury bonds and policy financial bonds across the entire market. Firstly, the liquidity-weighted method is employed to construct underlying indices for securities of five key maturities,, thereby solving the problem of poor replicability and tradability of the index due to the lack of liquidity of the constituent securities. The XGBoost algorithm in integrated learning is subsequently used to construct a multi-factor model for duration timing. This achieves the effect of controllable overall retraction of the index and results in considerable excess returns. Strategic bond indices can be used in practice as a vehicle for trading modes such as income certificates, income swaps and over-the-counter (OTC) options to meet the bond investment needs of different types of investors. In the future, further research can be conducted to enhance returns under different market environments in the direction of introducing bond coupons as one factor to be considered for constituent securities , and shifting from long to long-short strategies.

Keywords Strategic Bond Index, XGBoost Algorithm, Multi-Factor Modeling

JEL Classification G12 C22 C53