

债券量化系列之二：

债券组合量化策略研究



报告日期：2022年3月21日

★主要内容

本文我们探讨目标久期下的收益率曲线策略，探讨哑铃型组合与子弹型组合量化配置方案，基于不同模型进行的蝶式组合收益预测，得到最优配置结论，方法论上继续探究机器学习决策树模型在金融场景中的应用。

蝶式组合分析方面，蝶式组合的主要影响因素包括凸性溢价、利率曲线变化等，蝶式组合平均月度收益率为负，受曲线水平因子影响较小，受益于曲线变平与凸性增加。

模型构建方面，我们基于实证可行性原则选择直接预测活跃债券蝶式组合收益率，解释变量基于相关市场指标。采用线性降维模型（Lasso）和回归树模型（随机森林、AdaBoost、XGBoost），探讨特征筛选、交叉验证以及超参数调优等建模流程。

模型评价方面，从 R2 的角度，测试集预测效果最佳的模型为 XGBoost 模型，测试集与训练集预测效果差异最小的为随机森林模型，测试集预测效果最差的为 Lasso 模型，测试集与训练集预测效果差异最大的为 AdaBoost 模型。此外，胜率（HIT）的角度，测试集随机森林模型表现最好；盈亏比角度，Lasso 线性模型表现最佳。

蝶式策略表现方面，测试集 Lasso、随机森林、AdaBoost、XGBoost 四个模型分别实现年化收益 1.7%、1.4%、1.0%、1.5%，夏普比率 2.08、1.63、1.16、1.84，最大回撤 -0.7%、-0.6%、-0.8%、-0.4%。活跃券组合多头配置策略年化收益分别为 4.6%、4.4%、4.2%、4.5%，最大回撤为 -2.6%、-3.1%、-3.0%、2.7%，收益风险比较等权基准有效提升。

★风险提示

量化模型有效性基于历史数据得出，不排除失效的可能。

★致谢

感谢东方证券金工首席分析师朱剑涛老师的指导与帮助。

王冬黎 高级分析师(金融工程)
从业资格号： F3032817
投资咨询号 Z0014348
Tel: 8621-63325888-3975
15802167045
Email: dongli.wang@orientfutures.com

目录

1. 债券组合策略.....	5
1.1. 债券组合类型介绍与实现.....	5
1.2. 蝶式组合的构建方式.....	5
1.3. 基于活跃券的组合历史表现.....	6
1.4. 构建活跃券组合的补充说明.....	8
2. 蝶式组合影响归因.....	10
2.1. 利率曲线因子暴露与情景分析.....	10
2.2. 利率曲线变化的情景分析.....	11
2.3. 利率曲线 NS 模型拟合情况.....	13
3. 蝶式组合多空策略建模与表现.....	15
3.1. 模型解释变量说明.....	15
3.2. 模型构建.....	16
3.3. 模型效果分析.....	17
3.4. 蝶式策略净值结果.....	19
4. 基于蝶式策略的债券组合多头配置.....	20
5. 风险提示.....	22

图表目录

图表 1: 蝶式组合常用构建方案.....	6
图表 2: 基于活跃券组合构建蝶式组合的主要风险暴露属性.....	7
图表 3: 蝶式组合月度收益.....	7
图表 4: 蝶式组合月度累计收益.....	7
图表 5: 子弹型组合与杠铃型组合到期收益率利差.....	8
图表 6: 子弹型组合与杠铃型组合到期收益率走势.....	8
图表 7: 参考国债期货交割券规则制定活跃券筛选标准.....	8
图表 8: 国债活跃券组合全收益指数 (2Y)	9
图表 9: 国债活跃券组合全收益指数 (5Y)	9
图表 10: 国债活跃券组合全收益指数 (10Y)	9
图表 11: 国债活跃券组合久期 (2Y)	9
图表 12: 国债活跃券组合久期 (5Y)	10
图表 13: 国债活跃券组合久期 (10Y)	10
图表 14: 情景分析之债券基本信息.....	11
图表 15: 水平因子对蝶式组合影响.....	12
图表 16: 水平因子情景.....	12
图表 17: 斜率因子对蝶式组合影响.....	12
图表 18: 斜率因子情景.....	12
图表 19: 曲率因子对蝶式组合影响.....	13
图表 20: 曲率因子情景.....	13
图表 21: 国债收益率 NS 模型曲线拟合与离散点比较 (2022/03/03)	14
图表 22: 国债收益率 NS 模型曲线拟合与离散点比较 (2021/10/18)	14
图表 23: 债券组合、离散收益率与 NS 模型斜率因子.....	14
图表 24: 债券组合、离散收益率与 NS 模型曲率因子.....	14
图表 25: 蝶式组合收益率预测主要指标与变化形式说明.....	15
图表 26: 基于特征筛选的线性降维模型建模流程图.....	16
图表 27: 回归树模型建模流程图.....	16
图表 28: 模型训练集与测试集预测效果评价对比.....	17
图表 29: 模型训练集与测试集预测效果评价对比 (预测值解释能力 R2)	17
图表 30: 模型训练集与测试集预测效果评价对比 (预测方向胜率 HIT)	18
图表 31: 模型训练集与测试集预测效果评价对比 (预测盈亏比 PL)	18
图表 32: 蝶式组合套利策略净值 (测试集)	19

图表 33: 蝶式组合套利策略净值分析 (测试集)	19
图表 34: 蝶式组合套利策略净值 (全样本)	19
图表 35: 蝶式组合套利策略净值分析 (全样本)	19
图表 36: 基于蝶式利差预测的债券哑铃与子弹型组合轮动策略净值 (测试集)	20
图表 37: 基于蝶式利差预测的债券哑铃与子弹型组合轮动策略净值分析 (测试集)	20
图表 38: 基于蝶式利差预测的债券哑铃与子弹型组合轮动策略净值 (全样本)	21
图表 39: 基于蝶式利差预测的债券哑铃与子弹型组合轮动策略净值分析 (全样本)	21

1. 债券组合策略

我们在上一篇专题中研究了债券久期轮动策略，基于债券超额收益预测的框架探讨了通过预测不同久期组合超额收益的方式构建久期轮动策略，得到一个基础的利率债久期量化配置的方案，我们也基于此初步探究了机器学习 XGBoost 模型在时间预测预测上的应用。本文，我们进一步探讨目标久期下的收益率曲线策略，即对于特定久期敞口，选择长债加短债的哑铃型组合，还是选择仓位集中的子弹型组合，策略实现方面，我们通过构建不同的模型进行蝶式组合收益预测，得到对应配置信号进而指导债券多头组合构建。方法论上，我们继续探究机器学习决策树模型在金融场景中的应用，将着重探究随机森林、AdaBoost、XGBoost 三种决策树模型的构建与调参，以及与线性降维模型 (Lasso) 的效果对照。

1.1. 债券组合类型介绍与实现

我们对债券组合的主要类型进行简要回顾，介绍三种构建目标久期债券组合的方式：子弹型 (Bullet)、哑铃型 (Barbell) 与阶梯型 (Ladder)。采用子弹型组合，投资者将仓位集中在收益率曲线上的特定期限点，即目标久期对应的债券组合；采用哑铃型组合，投资者选择买入低久期债券高久期的债券进行组合搭配，形成持仓集中于长债与短债两个极端期限的目标久期组合。阶梯型组合即均衡持仓组合，将长久期、中久期与短久期债券进行等额配置。

1.2. 蝶式组合的构建方式

研究子弹与哑铃组合的相对强弱与研究蝶式组合策略思路较一致，我们先对蝶式组合的不同构建方法进行简要回顾，尽管对于债券组合而言，受制于资金一般采用等现金价值的方法进行，而实际上对蝶式组合的构建方案并非唯一确定、且都具有其特定的风险敞口特征。蝶式组合的构建上，仅考虑久期中性的设定并不能得到唯一的权重向量，尤其是基于具有杠杆交易属性、不受组合现金价值限制的国债期货蝶式策略组合，可更加灵活地基于多种预期情景进行相应策略设计，即下式中 β 的确定方法。

$$q_S D_S + q_L D_L + \alpha D_M = 0$$

$$q_S D_S = \beta q_L D_L$$

具体而言，我们简要介绍四种蝶式组合构建方案以及其相应的风险敞口，不同方案主要在于确定两翼的权重配比。(1) 现金中性 (Cash-Neutral) 赋权方法的约束条件为左右翼组合市值相等，该方法更适合现券市场实证，也与我们的多头组合配置策略思路相一致，其构建得到的组合对水平因子具有小幅正向暴露，斜率因子方面受益于曲线走平。

(2) Fifty-fifty 赋权方法为左右翼组合现金久期相等，即实现对于利率曲线长短端等额变化中性。(3) 回归赋权法基于对长端与中端收益率利差的变化和中端与短端收益率利差的变化进行回归得到，将短端与长端收益率波动性特征差异纳入权重配置，所得

组合斜率因子暴露方面对于长短端常规相对变化幅度中性。(4) 到期日赋权即基于短端与长端债券期限确立两翼的权重配比, 所得组合斜率因子暴露方面对于长短端与期限等比幅度变化中性, 若长短端利率波动的差异主要由期限所解释, 到期日赋权的结果会相对简便与稳定。

本文对于现券蝶式策略研究我们采用现金中性的基础设定, 对于其他赋权方法, 我们后续将在衍生品策略中加以探讨, 后文中的蝶式组合也均默认现金中性配比的蝶式组合。

图表 1: 蝶式组合常用构建方案

	Cash-neutral	Fifty-fifty Weighting	Regression Weighting	Maturity Weighting
权重计算公式	$q_S + q_L = -\alpha$	$q_S D_S = q_L D_L$ $= -\alpha D_M / 2$ $\beta = 1$	$X = \Delta(R_L - R_M)$ $Y = \Delta(R_M - R_S)$ $Y = \beta X + \varepsilon$	$\beta = \frac{M_M - M_S}{M_L - M_M}$
曲线水平因子敞口	中性	中性	中性	中性
曲线斜率因子敞口	具有敞口, 受益于曲线平坦化	对长短端等额变化中性	对长短端常规变化幅度中性	对长短端与期限等比幅度变化中性

资料来源: 东证衍生品研究院

1.3. 基于活跃券的组合历史表现

我们采用等现金价值的方法构建蝶式组合, 并对子弹型组合与哑铃型组合的历史收益率和利差情况进行简要展示。我们选择四年左右的目标久期, 基于 2Y 活跃券组合、5Y 活跃券组合与 10Y 活跃券组合构建蝶式组合, 基于蝶式组合进一步进行收益率预测建模, 进而得到子弹型与哑铃型组合的相对强弱信号。首先, 采用活跃券组合而非直接使用相应的中债指数是因为债券组合套利持仓周期相对短期, 换手率也高于久期轮动策略, 因而保证交易标的流动性是十分必要的, 指数套利看似简洁实际或因交易摩擦而不易实现。其次, 我们仅对哑铃型 (2Y 活跃券与 10Y 活跃券) 与子弹型 (5Y 活跃券) 进行判断, 而仅将阶梯型 (2Y、5Y 与 10Y 活跃券等额) 作为轮动策略的对照基准净值, 而不是三种方案都进行轮动, 也主要是基于控制交易成本考虑。

蝶式组合风险敞口的统计结果方面 (统计区间为 2010 年 2 月至 2022 年 3 月), 躯干与两翼部分平均久期约为 4.13 年, 两翼的权重分布结果为低久期与高久期组合权重平均配比约 1.3:1, 躯干组合平均到期收益率 3.18, 两翼组合平均到期收益率 3.13, 躯干组合凸性约 21.78, 两翼组合凸性约 30.01。基于该统计结果, 我们也可以初步认识到债券的凸性溢价, 基于债券定价原理, 高凸性组合在瞬时利率水平移动的影响下收益优于低凸性组合, 因而在定价上到期收益率相对较低, 因此当收益率短期大幅变动时高凸性组合较优, 而当收益率变化较小时低凸性组合因其固有到期收益率更高而更优。

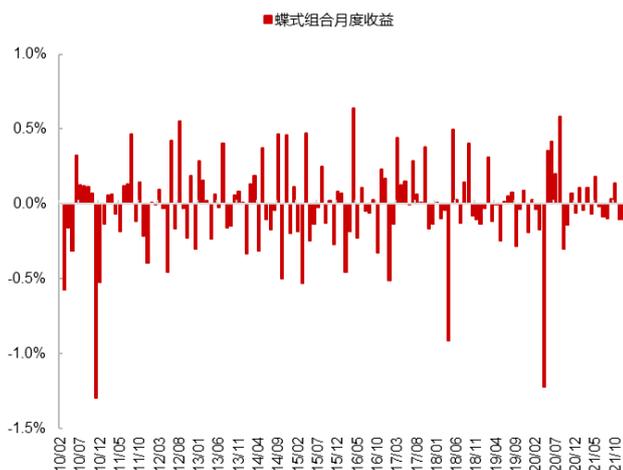
从我们构建的蝶式组合（多两翼空躯干）的月度收益率结果上来看，与凸性溢价分析较为相符，蝶式组合平均月度收益率为负，累计月度收益净值（未考虑月内波动，仅为大致参照）长期具有向下压力，但整体震荡格局较为明显。子弹型组合与哑铃型组合的到期收益率利差也中枢为正，具有一定的均值回复特征（到期收益率利差对未来组合相对收益有显著解释能力）。

图表 2：基于活跃券组合构建蝶式组合的主要风险暴露属性

	蝶式组躯干久期	蝶式组两翼久期	蝶式组合躯干到期收益率	蝶式组合两翼到期收益率	蝶式组合躯干凸性	蝶式组合两翼凸性	低久期组合久期	中久期组合久期
均值	4.13	4.13	3.18	3.13	21.78	30.01	1.75	4.13
	高久期组合久期	低久期组合凸性	中久期组合凸性	高久期组合凸性	低久期组合到期收益率	中久期组合到期收益率	高久期组合到期收益率	两翼低久期与高久期权重比
均值	7.29	4.84	21.8	63.7	2.91	3.17	3.41	1.31

资料来源：东证衍生品研究院；统计区间：2010-02 至 2022-03

图表 3：蝶式组合月度收益



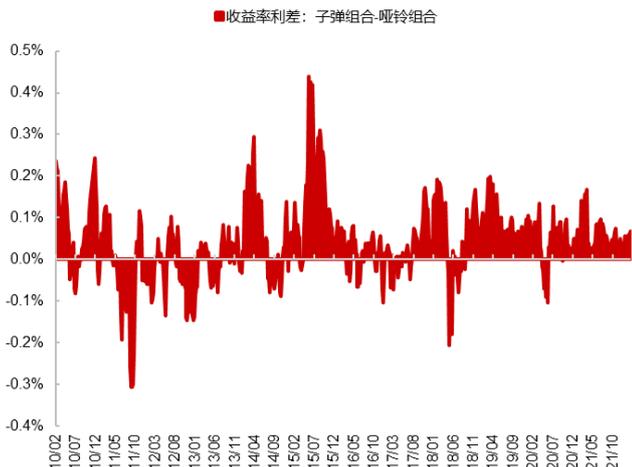
资料来源：东证衍生品研究院

图表 4：蝶式组合月度累计收益



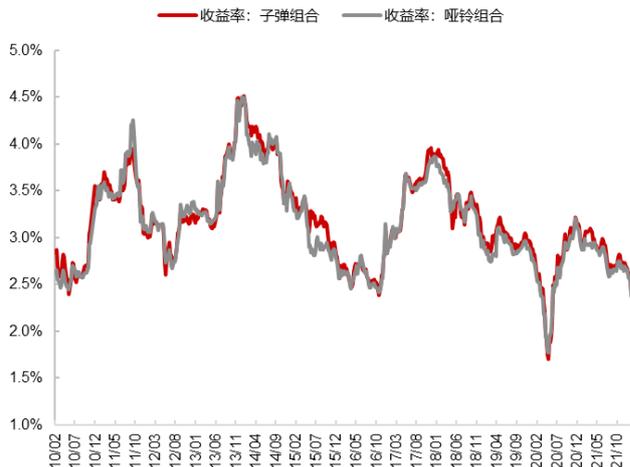
资料来源：东证衍生品研究院

图表 5: 子弹型组合与杠铃型组合到期收益率利差



资料来源: 东证衍生品研究院

图表 6: 子弹型组合与杠铃型组合到期收益率走势



资料来源: 东证衍生品研究院

1.4. 构建活跃券组合的补充说明

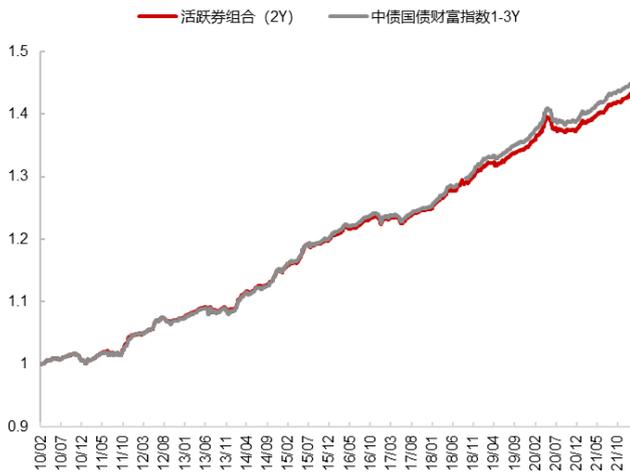
前文提到, 出于流动性考虑我们构建策略基于活跃券组合, 本节我们补充介绍债券组合策略选取的 2Y、5Y 与 10Y 活跃券组合的具体构建方法。首先, 我们基于三个期限的活跃券择券标准均参考了最新的国债期货交割券规则, 2Y 活跃券选择发行期限不高于 5 年, 合约到期月份首日剩余期限为 1.5-2.25 年的记账式付息国债; 5Y 活跃券选择发行期限不高于 7 年, 合约到期月份首日剩余期限为 4-5.25 年的记账式付息国债; 10Y 活跃券选择发行期限不高于 10 年, 合约到期月份首日剩余期限不低于 6.5 年的记账式付息国债。其次, 对于债券组合构建方面, 我们基于基于 20 日平均成交量选择关键期限前三只券作为活跃券, 以月度组合再平衡的频率构建等权组合, 票息收入不计算月度内的利息下月进行再投资。

图表 7: 参考国债期货交割券规则制定活跃券筛选标准

数据处理	详细说明
2Y 活跃券	选择发行期限不高于 5 年, 合约到期月份首日剩余期限为 1.5-2.25 年的记账式付息国债。
5Y 活跃券	选择发行期限不高于 7 年, 合约到期月份首日剩余期限为 4-5.25 年的记账式付息国债。
10Y 活跃券	选择发行期限不高于 10 年, 合约到期月份首日剩余期限不低于 6.5 年的记账式付息国债。
活跃券组合构建	采用上述标准筛选关键期限对应的债券池, 基于 20 日平均成交量选择关键期限前三只券作为活跃券, 以月度组合再平衡的频率构建等权组合, 票息收入不计算月度内的利息下月进行再投资。

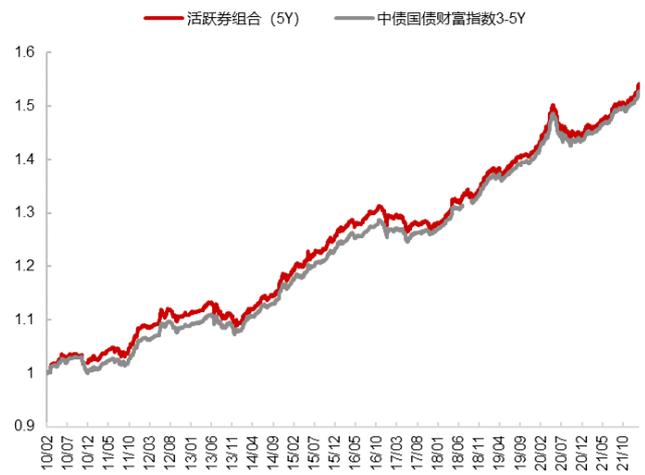
资料来源: 东证衍生品研究院

图表 8: 国债活跃券组合全收益指数 (2Y)



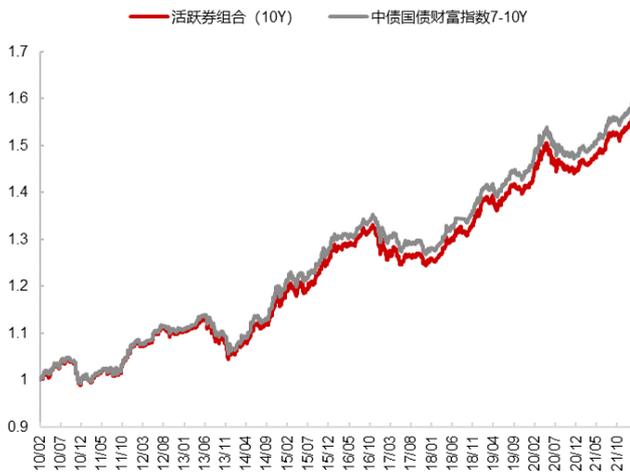
资料来源: 东证衍生品研究院

图表 9: 国债活跃券组合全收益指数 (5Y)



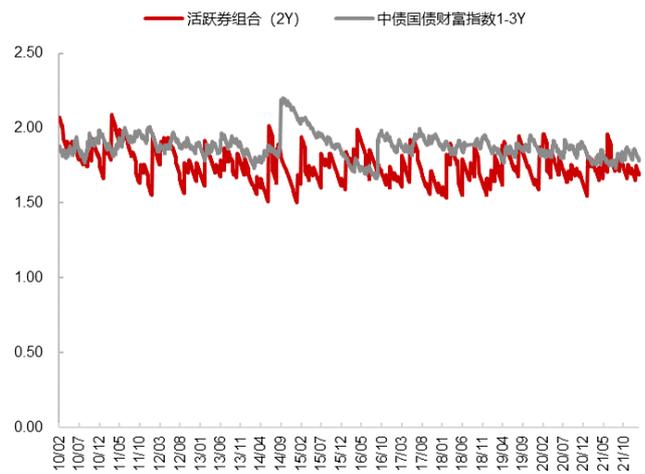
资料来源: 东证衍生品研究院

图表 10: 国债活跃券组合全收益指数 (10Y)



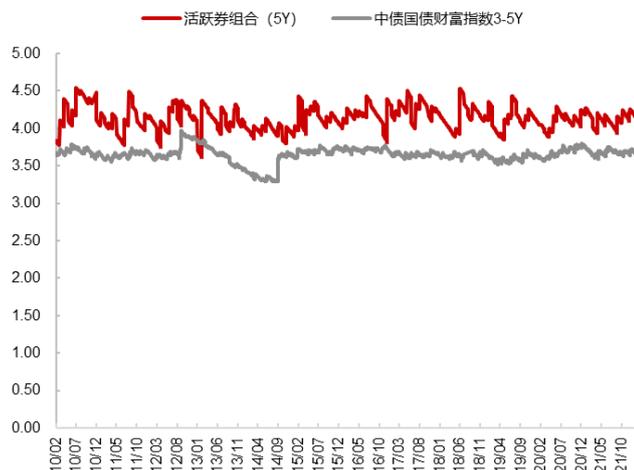
资料来源: 东证衍生品研究院

图表 11: 国债活跃券组合久期 (2Y)



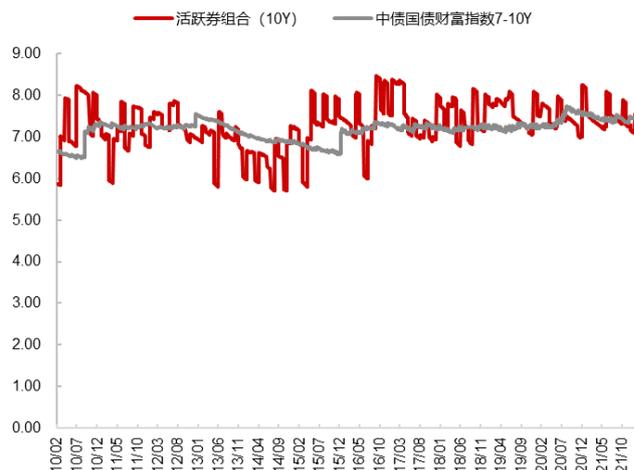
资料来源: 东证衍生品研究院

图表 12: 国债活跃券组合久期 (5Y)



资料来源: 东证衍生品研究院

图表 13: 国债活跃券组合久期 (10Y)



资料来源: 东证衍生品研究院

2. 蝶式组合影响归因

2.1. 利率曲线因子暴露与情景分析

基于收益率曲线相关的经典理论文献研究, 将收益率曲线抽象描述为三个主要因子特征是一种常用做法, 其中 Nelson and Siegel (1987) 提出的 NS 模型将收益率曲线描述为如下方程式, 模型不仅具有较佳的拟合效果, 同时也有具有明确经济涵义的三因子参数。Diebold and Li (2003) 进一步拓展收益率曲线的研究, 基于 AR、VAR、ECM 等不同统计模型研究收益率曲线三因子的预测, 除此之外收益率曲线因子的预测课题有一众丰富文献研究资料。

$$R^c(t, s_i) = \beta_0 + \beta_1 \frac{1 - e^{-\lambda s}}{\lambda s} + \beta_3 \left(\frac{1 - e^{-\lambda s}}{\lambda s} - e^{-\lambda s} \right)$$

我们尝试从收益率曲线的角度对蝶式组合收益进行归因, 因子暴露的公式推导方面, 记 P_t 为蝶式组合 t 时的价格, 基于组合债券定价公式, 其为未来 n 期现金流基于对应久期收益率进行折现的现值乘以其仓位加总计算, 可写为:

$$P_t = \sum_{i=1}^n q_i F_i e^{-s_i R^c(t, s_i)}$$

基于 NS 模型对收益率曲线三因子分别求导可得蝶式组合对水平、斜率、曲率的现金久期暴露:

$$D_{LVL,t} = \frac{\partial P}{\partial \beta_0} = -\sum_i q_i s_i F_i e^{-s_i R^c(0,s_i)}$$

$$D_{SLP,t} = \frac{\partial P}{\partial \beta_1} = -\sum_i q_i s_i \frac{1 - e^{-\lambda s_i}}{\lambda s_i} F_i e^{-s_i R^c(0,s_i)}$$

$$D_{CVT,t} = \frac{\partial P}{\partial \beta_2} = -\sum_i q_i s_i \left(\frac{1 - e^{-\lambda s_i}}{\lambda s_i} - e^{-\lambda s_i} \right) F_i e^{-s_i R^c(0,s_i)}$$

$D_{LVL,t}$ 为蝶式组合对水平因子的暴露，因组合构建基于久期中性，该因子值一般非常小，即利率的小幅平移对蝶式组合影响不大； $D_{SLP,t}$ 为蝶式组合对斜率因子的暴露， $D_{CVT,t}$ 为蝶式组合对曲率因子的风险暴露，通过对特定组合暴露值的计算可对蝶式组合的风险敞口进行监控。

2.2. 利率曲线变化的情景分析

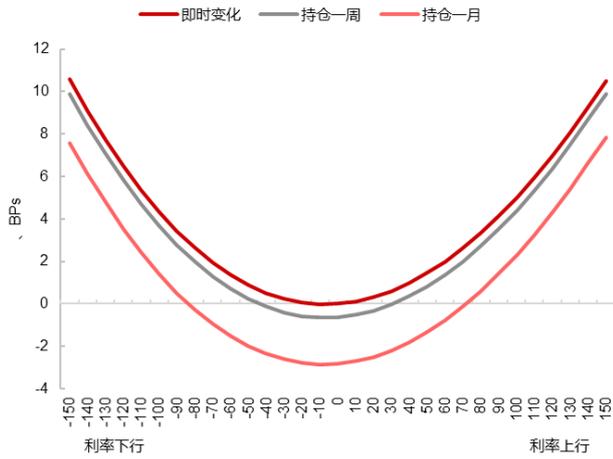
我们基于三个不同久期的债券进行蝶式组合基于利率曲线三因子变化的情景分析展示，使用个券而非债券组合主要是简化计算流程。我们分别选择的三只债券久期分别为 2.15 年、4.32 年和 8.21 年，蝶式组合权重配比为 0.64：-1：0.36，基于该组合进一步分析水平、斜率、曲率因子变化在不同持仓周期下对蝶式组合的影响。水平因子方面，我们尝试将曲线在估值日水平基础上平移，收益率曲线的水平即时变化对蝶式组合有正向影响，但影响程度较为有限，而随着持仓时间的增加，收益率曲线需要更大的水平移动量才能实现蝶式组合的盈亏平衡，反应出债券的凸性溢价。斜率因子方面，蝶式组合显著受益于曲线变平，曲线斜率变化的影响系数足够大，使得不同持仓周期（即时变化到持仓一个月）对最终的收益影响相对较小。曲率因子方面，蝶式组合多两端空中间的做法本来则是直接交易曲线的凸性，因而其影响较为直观，影响系数也是最大的。

图表 14：情景分析之债券基本信息

	债券 ID	起息日	到期日	票息率 (%)	付息频率	估值日久期	构建蝶式组合权重
低久期债券	210004.IB	2021/4/8	2024/4/8	2.84	Y1	2.15	0.64
中久期债券	210011.IB	2021/8/12	2026/8/12	2.69	Y1	4.32	-1.00
高久期债券	210009.IB	2021/5/27	2031/5/27	3.02	M6	8.21	0.36

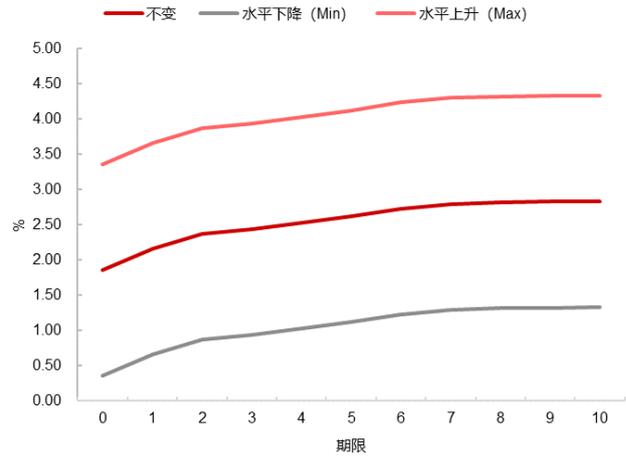
资料来源：东证衍生品研究院；估值日：2022-01-14

图表 15: 水平因子对蝶式组合影响



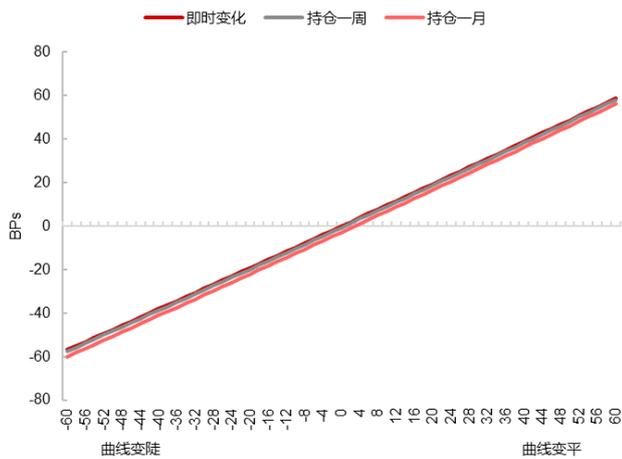
资料来源: 东证衍生品研究院

图表 16: 水平因子情景



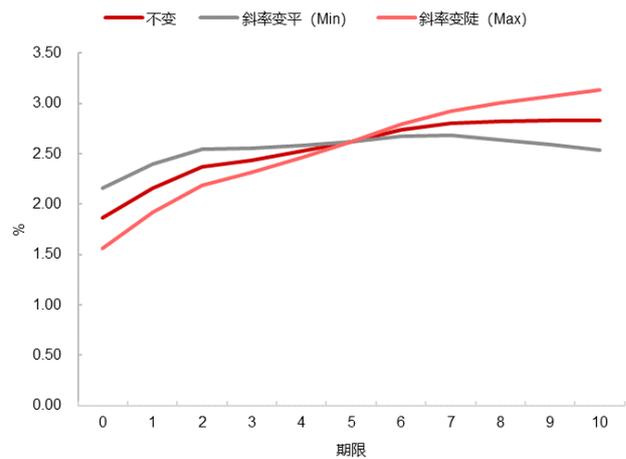
资料来源: 东证衍生品研究院

图表 17: 斜率因子对蝶式组合影响



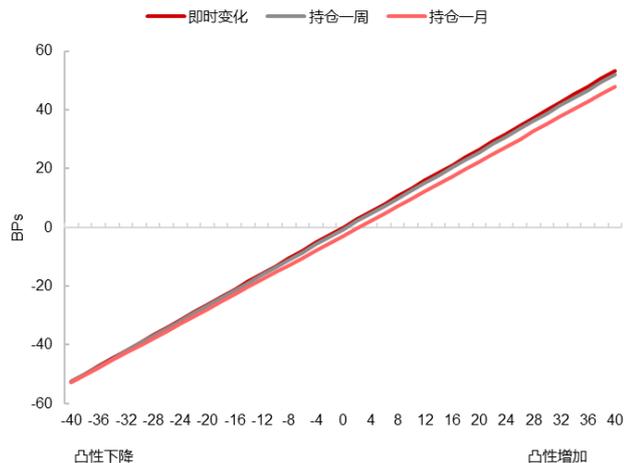
资料来源: 东证衍生品研究院

图表 18: 斜率因子情景



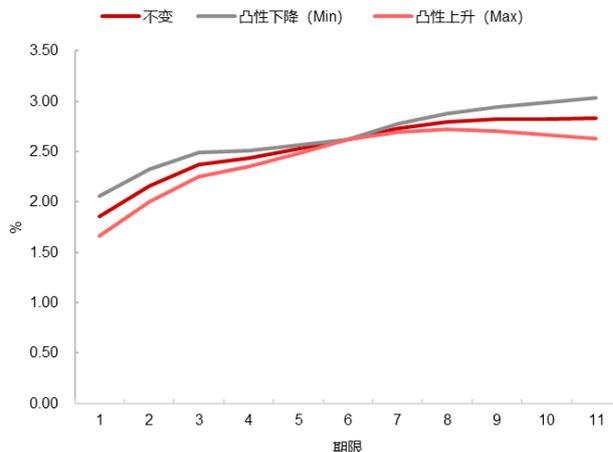
资料来源: 东证衍生品研究院

图表 19: 曲率因子对蝶式组合影响



资料来源: 东证衍生品研究院

图表 20: 曲率因子情景



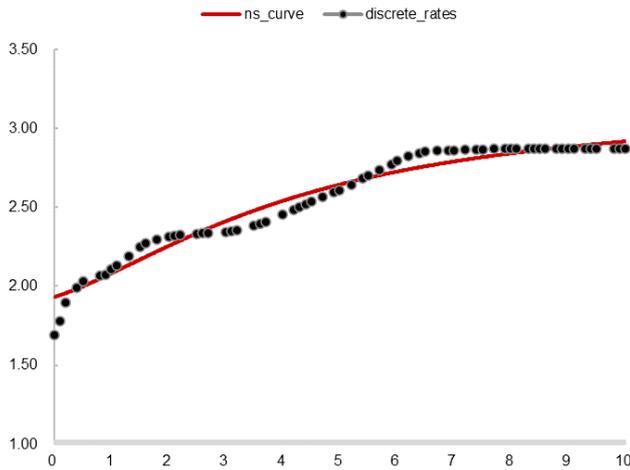
资料来源: 东证衍生品研究院

2.3. 利率曲线 NS 模型拟合情况

上文分析了基于独立的曲线因子变化蝶式组合的收益情况, 我们尝试基于 NS 模型对中债国债即期利率曲线进行建模, 初步分析表明曲线拟合结果与离散实际值之间存在一定误差, 对于 NS 模型而言, 越是高阶的因子其估计值 (相对离散收益率或相对活跃券收益率) 误差越大, 尤其是我们下图所展示的曲率因子, 模型得到的曲率因子走势呈现出波动一致性低、波幅更大的特征, 而蝶式组合策略本身就是交易收益率曲线更高阶的特征, 所以基于收益率曲线模型因子直接解释蝶式组合或引入较大的误差。

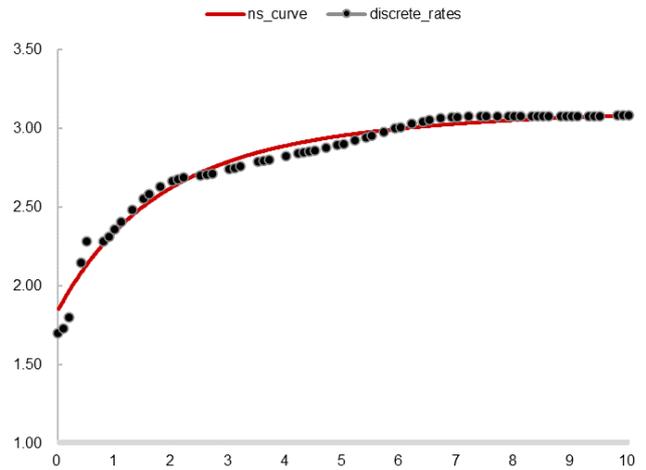
根据预测收益率曲线三因子, 得到曲线的变动预期, 进而进行债券组合策略应对或为较为理想的做法, 但由于我们基于利率曲线拟合得到的因子估计并非完全精确, 而债券组合的选择又是极为精细的操作, 基于理论上蝶式组合对利率因子的敞口, 通过预测因子的变化进而预测蝶式组合的方法或因其中的误差过大而影响策略结果。因而, 我们采取直接预测活跃券蝶式组合的预期收益率的方式进行策略建模。

图表 21: 国债收益率 NS 模型曲线拟合与离散点比较 (2022/03/03)



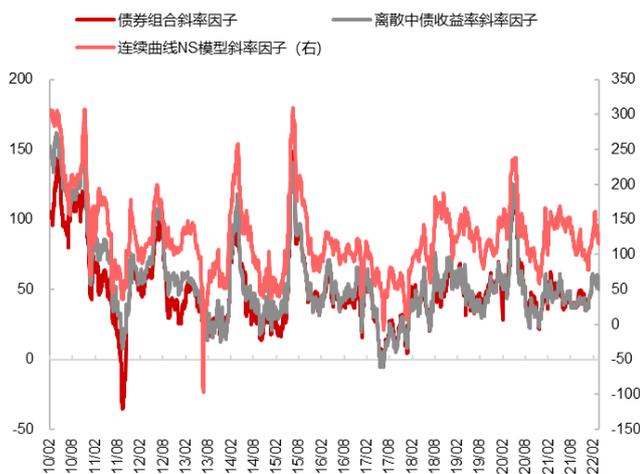
资料来源: 东证衍生品研究院

图表 22: 国债收益率 NS 模型曲线拟合与离散点比较 (2021/10/18)



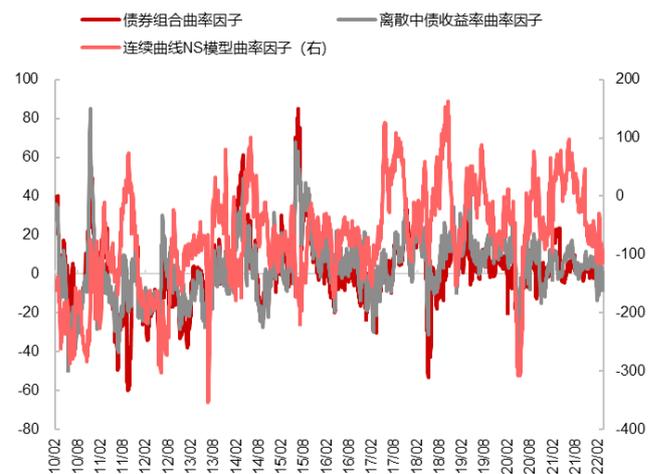
资料来源: 东证衍生品研究院

图表 23: 债券组合、离散收益率与 NS 模型斜率因子



资料来源: 东证衍生品研究院

图表 24: 债券组合、离散收益率与 NS 模型曲率因子



资料来源: 东证衍生品研究院

3. 蝶式组合多空策略建模与表现

3.1. 模型解释变量说明

我们基于实证可行性选择直接预测活跃券蝶式组合收益率。解释变量指标方面，我们对蝶式组合的潜在市场影响因素进行挖掘后得到一套可能对套利机会有预测能力的指标，其主要来自于四个大类：活跃券组合、蝶式组合、NS 模型因子、远期利率曲线，其中，前两个大类指标相关性较高，更多是从蝶式组合自身估值或自身价格规律的角度得到策略信号，后两者是从曲线与更为抽象的收益率因子中查找信息。对于所有的指标我们都进行不同方式的时序运算，以更多挖掘指标时序变化特征，并进行标准化。

图表 25：蝶式组合收益率预测主要指标与变化形式说明

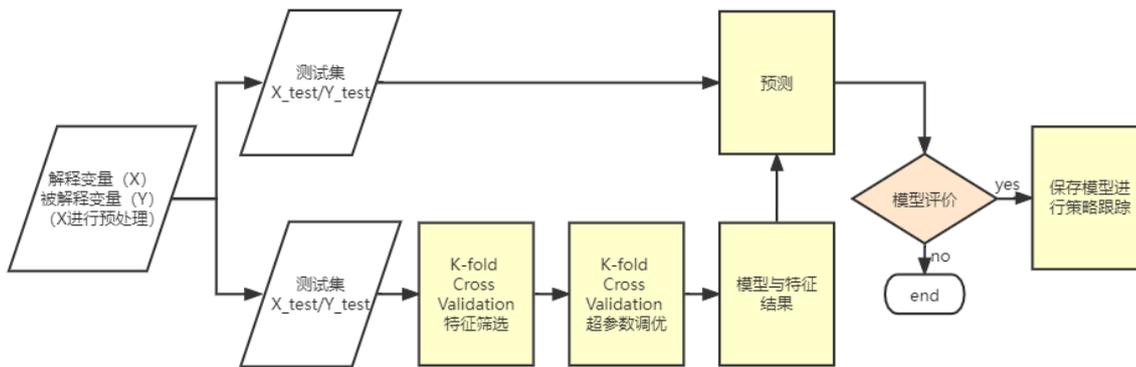
指标大类	指标代码	指标名称	指标时序运算	指标标准化
活跃券组合	ytm_lowdur	低久期组合到期收益率	原值、 差分、 5 日差分、 20 日波动率	滚动窗口 (252) : ZSCORE、 PERCENTILE、 SCALE
	ytm_middur	中久期组合到期收益率		
	ytm_highdur	高久期组合到期收益率		
	ytm_spread_lm	到期收益率利差：高久期-中久期		
	ytm_spread_ms	到期收益率利差：中久期-低久期		
	ytm_spread_ls	到期收益率利差：高久期-低久期		
	ytm_curvature	债券组合曲率		
蝶式组合	ytm_butt_body	蝶式组合躯干部分收益率		
	ytm_butt_wings	蝶式组合两翼部分加权平均收益率		
	ytm_butt_spread	蝶式组合收益率利差		
NS 模型因子	Level	水平因子		
	Slope	斜率因子		
	Curvature	曲率因子		
远期利率曲线	S_1.0	即期利率 (1)		
	F_1.0_2.0	远期利率 (1,2)		
	F_2.0_3.0	远期利率 (2,3)		
	F_3.0_4.0	远期利率 (3,4)		
	F_4.0_5.0	远期利率 (4,5)		
	F_5.0_6.0	远期利率 (5,6)		
	F_6.0_7.0	远期利率 (6,7)		
	F_7.0_8.0	远期利率 (7,8)		
	F_8.0_9.0	远期利率 (8,9)		
	F_9.0_10.0	远期利率 (9,10)		

资料来源：东证衍生品研究院

3.2. 模型构建

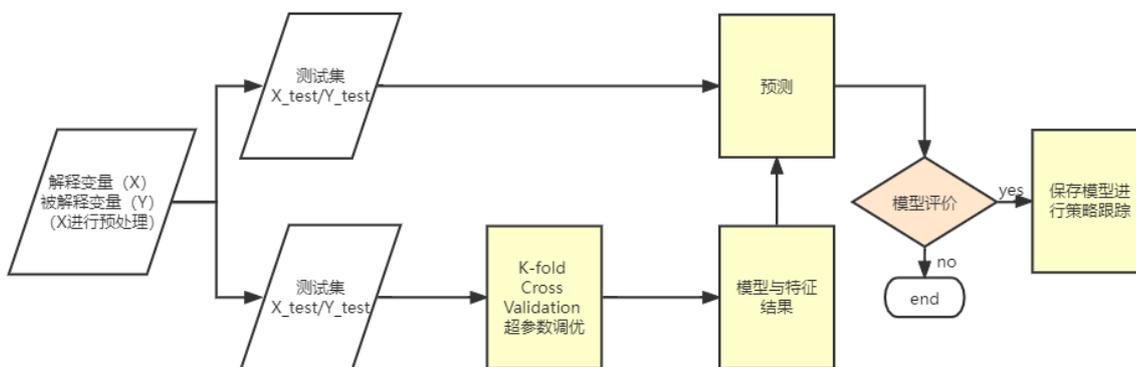
基于上述市场指标以及相应时序运算与标准化，我们得到 276 个特征因子，样本方面由于解释变量与被解释变量均为市场数据，我们以日频采样得到共计约 3000 个样本。对特征因子原始数据我们先进行预处理，分别基于滚动窗口去除极端值和填补缺失值，以 8:2 划分训练集和测试集。接下来我们基于线性降维模型（以 Lasso 为例）和回归树模型（随机森林、AdaBoost、XGBoost），探讨特征筛选、交叉验证以及超参数调优等建模流程。对于线性降维模型，特征筛选是较为必要的，因而建模过程中分别对特征筛选环境和超参数调优环境均进行训练集上的交叉验证。对于回归树模型，特征筛选并不必要且反而易导致模型过拟合，超参数调优部分进行一次训练集交叉验证。最终，通过测试集效果和测试集与训练集对比进行评价模型（是否过拟合或欠拟合）。

图表 26：基于特征筛选的线性降维模型建模流程图



资料来源：东证衍生品研究院

图表 27：回归树模型建模流程图



资料来源：东证衍生品研究院

3.3. 模型效果分析

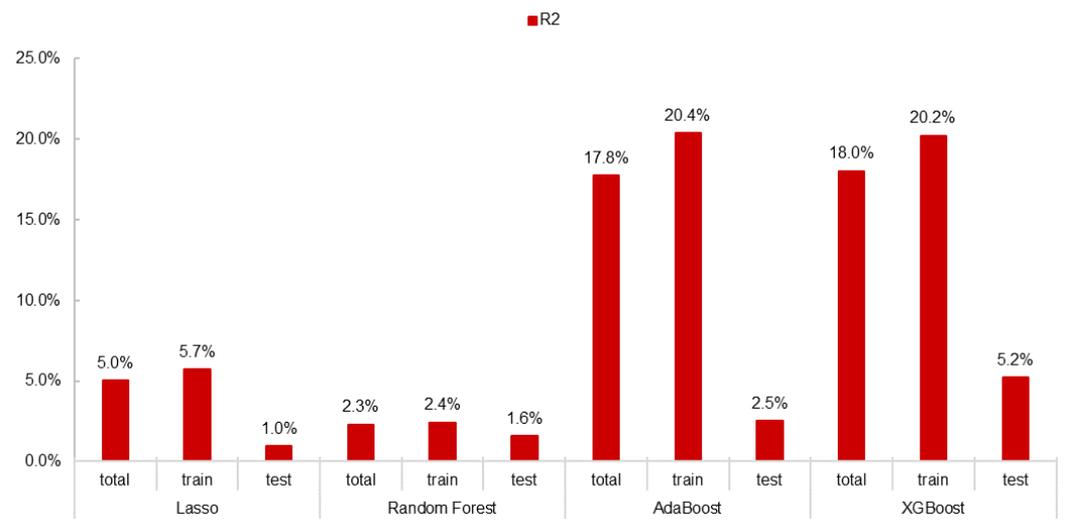
我们基于测试集与训练集模型预测值解释能力 (R2)、胜率 (HIT) 和盈亏比 (PL) 三个评价指标分析四个模型表现, 由于 R2 接近不同模型的目标函数, 因而是进行模型评价的最重要指标, 胜率和盈亏比主要侧面反应模型策略效果。从 R2 的角度, 测试集预测效果最佳的模型为 XGBoost 模型 (5.2%), 测试集与训练集预测效果差异最小的为随机森林模型(测试集为训练集的 64%), 测试集预测效果最差的为 Lasso 模型(1.0%), 测试集与训练集预测效果差异最大的为 AdaBoost 模型 (测试集为训练集的 12%)。此外, 胜率 (HIT) 的角度, 测试集随机森林模型表现最好; 盈亏比角度, Lasso 线性模型表现最佳。

图表 28: 模型训练集与测试集预测效果评价对比

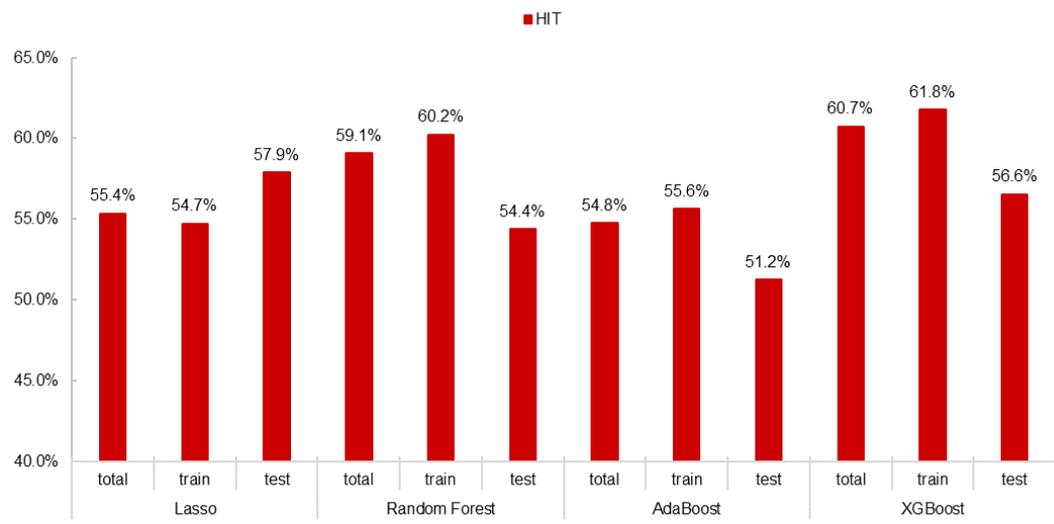
	Lasso			Random Forest			AdaBoost			XGBoost		
	total	train	test	total	train	test	total	train	test	total	train	test
R2	5.0%	5.7%	1.0%	2.3%	2.4%	1.6%	17.8%	20.4%	2.5%	18.0%	20.2%	5.2%
HIT	55.4%	54.7%	57.9%	59.1%	60.2%	54.4%	54.8%	55.6%	51.2%	60.7%	61.8%	56.6%
PL	1.27	1.24	1.47	1.26	1.25	1.25	1.34	1.34	1.35	1.59	1.63	1.38

资料来源: 东证衍生品研究院

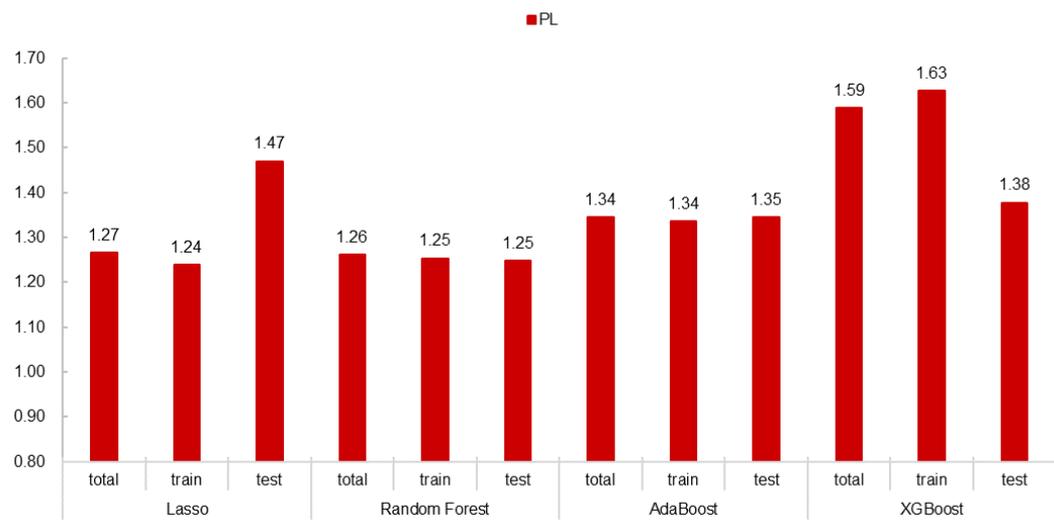
图表 29: 模型训练集与测试集预测效果评价对比 (预测值解释能力 R2)



资料来源: 东证衍生品研究院

图表 30: 模型训练集与测试集预测效果评价对比 (预测方向胜率 HIT)


资料来源: 东证衍生品研究院

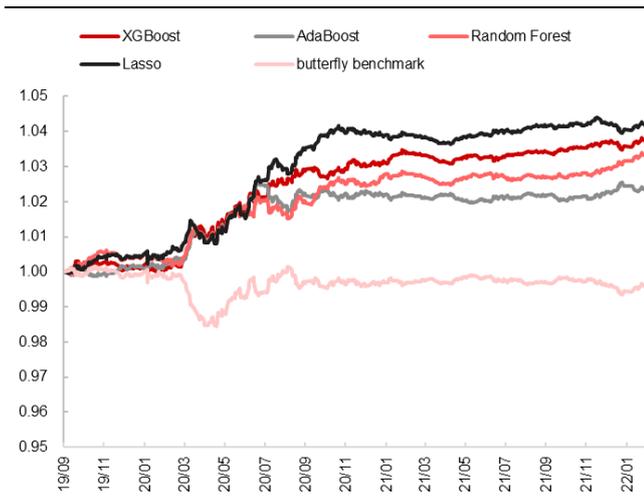
图表 31: 模型训练集与测试集预测效果评价对比 (预测盈亏比 PL)


资料来源: 东证衍生品研究院

3.4. 蝶式策略净值结果

蝶式策略表现方面, 不计债券做空成本的情况下, 测试集 Lasso、随机森林、AdaBoost、XGBoost 四个模型分别实现年化收益 1.7%、1.4%、1.0%和 1.5%, 夏普比率 2.08、1.63、1.16 和 1.84, 最大回撤-0.7%、-0.6%、-0.8%和-0.4%, 综合来看不同模型测试集净值结果差异不大, 相对而言 AdaBoost 模型略微逊色。全样本上, 四个模型的年化收益分别为 1.3%、1.6%、1.6%和 2.3%, 与上述 R2 角度的模型评价一致, Lasso 与随机森林模型全样本与测试集收益率较为接近, 两个模型过拟合风险小, AdaBoost、XGBoost 测试集表现相对全样本较弱, 可能存在一定的过拟合风险。

图表 32: 蝶式组合套利策略净值 (测试集)



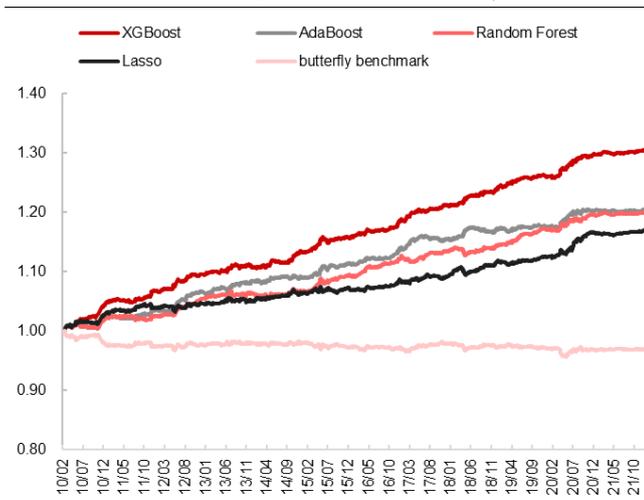
资料来源: 东证衍生品研究院

图表 33: 蝶式组合套利策略净值分析 (测试集)

	Lasso	Random Forest	Ada-Boost	XG-Boost
累计收益率	4.2%	3.3%	2.3%	3.7%
年化收益率	1.7%	1.4%	1.0%	1.5%
年化波动率	0.8%	0.8%	0.8%	0.8%
最大回撤率	-0.7%	-0.6%	-0.8%	-0.4%
胜率	54.1%	52.9%	50.8%	52.1%
盈亏比	1.26	1.22	1.21	1.31
夏普比率	2.08	1.63	1.16	1.84
Calmar 比	2.60	2.35	1.17	3.60

资料来源: 东证衍生品研究院

图表 34: 蝶式组合套利策略净值 (全样本)



资料来源: 东证衍生品研究院

图表 35: 蝶式组合套利策略净值分析 (全样本)

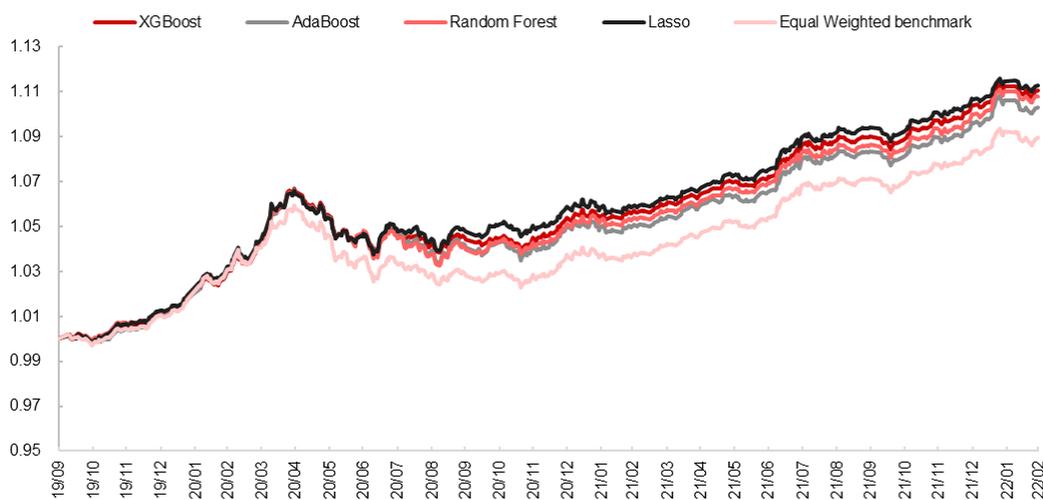
	Lasso	Random Forest	Ada-Boost	XG-Boost
累计收益率	16.7%	20.5%	20.7%	30.6%
年化收益率	1.3%	1.6%	1.6%	2.3%
年化波动率	1.1%	1.1%	1.1%	1.1%
最大回撤率	-1.3%	-1.2%	-1.1%	-1.0%
胜率	52.8%	53.6%	53.0%	54.4%
盈亏比	1.12	1.14	1.17	1.25
夏普比率	1.21	1.46	1.48	2.11
Calmar 比	0.99	1.28	1.44	2.37

资料来源: 东证衍生品研究院

4. 基于蝶式策略的债券组合多头配置

最后，我们将上述蝶式组合策略的信号结果转化为债券多头配置方案，回顾文章第一节中的债券组合介绍，债券组合多头配置方案我们即在哑铃组合与子弹组合中进行选择，配置策略中的组合构建的权重配比与蝶式组合相同（权重取绝对值）。具体而言，若蝶式组合的信号为正，即套利方面推荐多两端空中间，则对应的配置信号为配置哑铃组合，反之推荐配置子弹组合。如下，我们对测试集债券组合配置策略结果进行展示，Lasso、随机森林、AdaBoost、XGBoost 年化收益分别为 4.6%、4.4%、4.2%、4.5%，最大回撤为-2.6%、-3.1%、-3.0%、2.7%，收益风险比较等权基准有效提升。

图表 36: 基于蝶式利差预测的债券哑铃与子弹型组合轮动策略净值 (测试集)

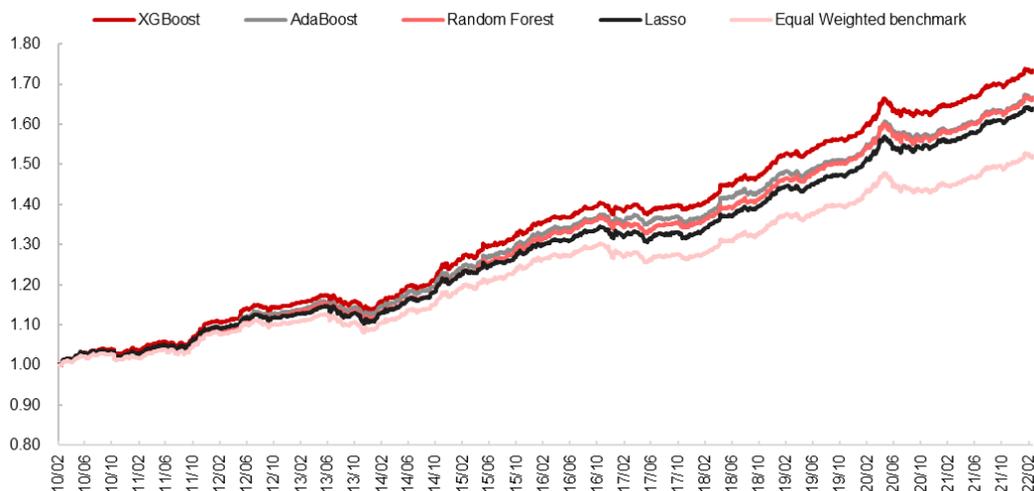


资料来源: 东证衍生品研究院

图表 37: 基于蝶式利差预测的债券哑铃与子弹型组合轮动策略净值分析 (测试集)

	Lasso	Random Forest	AdaBoost	XGBoost	Equal Weighted benchmark
累计收益率	11.3%	10.8%	10.3%	11.0%	9.0%
年化收益率	4.6%	4.4%	4.2%	4.5%	3.7%
年化波动率	1.7%	1.8%	1.8%	1.6%	1.8%
最大回撤率	-2.6%	-3.1%	-3.0%	-2.7%	-3.5%
胜率(D)	55.9%	55.9%	55.3%	56.3%	55.4%
盈亏比	1.26	1.22	1.20	1.25	1.16
夏普比率	2.74	2.51	2.37	2.73	2.09
Calmar 比	1.74	1.41	1.38	1.67	1.06

资料来源: 东证衍生品研究院

图表 38: 基于蝶式利差预测的债券哑铃与子弹型组合轮动策略净值 (全样本)


资料来源: 东证衍生品研究院

图表 39: 基于蝶式利差预测的债券哑铃与子弹型组合轮动策略净值分析 (全样本)

	Lasso	Random Forest	AdaBoost	XGBoost	Equal Weighted benchmark
累计收益率	64.0%	66.5%	66.7%	73.4%	52.8%
年化收益率	4.3%	4.4%	4.4%	4.8%	3.6%
年化波动率	1.7%	1.7%	1.7%	1.7%	1.7%
最大回撤率	-3.9%	-3.9%	-3.6%	-3.5%	-4.2%
胜率(D)	58.2%	58.4%	58.9%	59.1%	57.9%
盈亏比	1.13	1.13	1.10	1.17	1.07
夏普比率	2.53	2.55	2.55	2.85	2.16
Calmar 比	1.08	1.14	1.22	1.36	0.88

资料来源: 东证衍生品研究院

5. 风险提示

量化模型失效风险，指标的有效性基于历史数据得出，不排除失效的可能。

期货走势评级体系（以收盘价的变动幅度为判断标准）

走势评级	短期（1-3个月）	中期（3-6个月）	长期（6-12个月）
强烈看涨	上涨 15%以上	上涨 15%以上	上涨 15%以上
看涨	上涨 5-15%	上涨 5-15%	上涨 5-15%
震荡	振幅-5%-+5%	振幅-5%-+5%	振幅-5%-+5%
看跌	下跌 5-15%	下跌 5-15%	下跌 5-15%
强烈看跌	下跌 15%以上	下跌 15%以上	下跌 15%以上

上海东证期货有限公司

上海东证期货有限公司成立于 2008 年，是一家经中国证券监督管理委员会批准的经营期货业务的综合性公司。东证期货是东方证券股份有限公司全资子公司，注册资本金 23 亿元人民币，员工近 600 人。公司主要从事商品期货经纪、金融期货经纪、期货投资咨询、资产管理、基金销售等业务，拥有上海期货交易所、大连商品交易所、郑州商品交易所和上海国际能源交易中心会员资格，是中国金融期货交易所全面结算会员。公司拥有东证润和资本管理有限公司，上海东祺投资管理有限公司和东证期货国际（新加坡）私人有限公司三家全资子公司。

东证期货以上海为总部所在地，在大连、长沙、北京、上海、郑州、太原、常州、广州、青岛、宁波、深圳、杭州、西安、厦门、成都、东营、天津、哈尔滨、南宁、重庆、苏州、南通、泉州、汕头、沈阳、无锡、济南等地共设有 33 家营业部，并在北京、上海、广州、深圳多个经济发达地区拥有 134 个证券 IB 分支网点，未来东证期货将形成立足上海、辐射全国的经营网络。

自 2008 年成立以来，东证期货秉承稳健经营、创新发展的宗旨，坚持市场化、国际化、集团化的发展道路，打造以衍生品风险管理为核心，具有研究和技术两大核心竞争力，为客户提供综合财富管理平台的一流衍生品服务商。

分析师承诺

王冬黎

本人具有中国期货业协会授予的期货执业资格或相当的专业胜任能力，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接接收到任何形式的报酬。

免责声明

本报告由上海东证期货有限公司（以下简称“本公司”）制作及发布。

本研究报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本研究报告是基于本公司认为可靠的且目前已公开的信息撰写，本公司力求但不保证该信息的准确性和完整性，客户也不应该认为该信息是准确和完整的。同时，本公司不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司会适时更新我们的研究，但可能会因某些规定而无法做到。除了一些定期出版的报告之外，绝大多数研究报告是在分析师认为适当的时候不定期地发布。

在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求。客户应考虑本报告中的任何意见或建议是否符合其特定状况，若有必要应寻求专家意见。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买投资标的的邀请或向人作出邀请。

在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任，投资者需自行承担风险。

本报告主要以电子版形式分发，间或也会辅以印刷品形式分发，所有报告版权均归本公司所有。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。

如需引用、刊发或转载本报告，需注明出处为东证衍生品研究院，且不得对本报告进行任何有悖原意的引用、删节和修改。

东证衍生品研究院

地址：上海市中山南路318号东方国际金融广场2号楼21楼

联系人：梁爽

电话：8621-63325888-1592

传真：8621-33315862

网址：www.orientfutures.com

Email：research@orientfutures.com