

基于 XGBoost 的国债期货择时策略

摘要

本文基于日内尾盘信息、“股债跷跷板”效应、短期动量等量价特征构造择时因子，在模型方面使用 XGBoost 模型对输入因子信号进行处理，建立了国债期货择时模型。

模型在 2015 年至 2023 年验证集上年化收益率为 4.5%，夏普率 1.34，最大回撤 4.2%，在 2024 年上半年测试集上年化收益达到 7.6%，夏普率 3.84，最大回撤 0.7%，两个区间内均优于多头表现。在后续的研究中，我们将考虑纳入更多的因子并使用更为完善的模型对策略进行改进。

量化

专题报告

2024 年 7 月 9 日

国贸期货·研究院
金融衍生品中心

李泽钜

从业资格号：F0251925

投资咨询号：Z0000116

研究助理 许郑霏

从业资格号：F03128092

欢迎扫描下方二维码
进入国贸投研小程序



期市有风险，入市需谨慎

1 国债期货介绍

目前，我国共有四个国债期货品种在中国金融期货交易所（CFFEX）上市，可供投资者进行交易，分别为 2 年期国债期货（TS）、5 年期国债期货（TF）、10 年期国债期货（T）和 30 年期国债期货（TL）。5 年期和 10 年期国债期货运行时间较长，分别于 2013 年和 2015 年上市，成交量也相对较高，市场较活跃；30 年期国债期货上市最晚，于 2023 年 4 月 21 日上市交易。各品种合约细则如下：

图表 1：国债期货各品种合约细则

品种代码	TS	TF	T	TL
上市时间	2018 年 8 月 17 日	2013 年 9 月 6 日	2015 年 3 月 20 日	2023 年 4 月 21 日
合约标的	面值为 200 万元人民币、票面利率为 3% 的名义中短期国债	面值为 100 万元人民币、票面利率为 3% 的名义中期国债	面值为 100 万元人民币、票面利率为 3% 的名义长期国债	面值为 100 万元人民币、票面利率为 3% 的名义超长期国债
可交割国债	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	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	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	发行期限不高于 30 年，合约到期月份首日剩余期限不低于 25 年的记账式付息国债

数据来源：中金所，国贸期货整理

2 极限梯度提升算法（XGBoost）

XGBoost 全称为 eXtreme Gradient Boosting，中文名可称为极限梯度提升算法，使用梯度提升（Gradient Boosting）的思想，这是一个迭代的机器学习算法，通过逐步添加新的模型来修正现有模型的错误。其特点主要有三个：运用 boost 的思想，使用加法模型进行预测；通过二阶泰勒展开计算损失函数，提升精确性；在目标函数中加入正则化。

在预测方面，其加法模型结构如下：

$$\hat{y}_i = \phi(\mathbf{x}_i) = \sum_{k=1}^K f_k(\mathbf{x}_i)$$

每个新模型被训练来拟合残差，即实际值与当前模型预测值之间的差异。这个过程持续进行，直到达到预设的迭代次数或者模型性能不再显著提高。

其目标函数由二阶泰勒展开的损失函数和复杂度惩罚项构成：

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \Omega(f_t)$$

对损失函数进行二阶泰勒展开可以得到：

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^n \left[l(y_i, \hat{y}^{(t-1)}) + g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i) \right] + \Omega(f_t)$$

其中括号内第一项为常数项，简化后得到：

$$\tilde{\mathcal{L}}^{(t)} = \sum_{i=1}^n \left[g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i) \right] + \Omega(f_t)$$

复杂度惩罚项 $\Omega(f)$ 如下：

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$$

T为叶子节点数，第二项为L2正则化项，也可采用L1正则化或同时使用。

综上，最终得到的目标函数为：

$$\begin{aligned} \tilde{\mathcal{L}}^{(t)} &= \sum_{i=1}^n \left[g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \\ &= \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \end{aligned}$$

从而得到每片叶子的权重：

$$w_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}$$

3 策略构建

本文构建国债期货择时策略的流程如下：第一步，从逻辑出发构造多维度择时因子，作为模型的特征输入；第二步，使用XGBoost模型和交叉验证，将数据进行划分后在训练集和验证集上进行训练并调试模型；第三步，用得到的模型在测试集上生成信号并回测策略表现。

3.1 特征选取

本文围绕日间收益、股债跷跷板、尾盘信息、短期动量四个维度构建模型的输入特征。

3.1.1 股债跷跷板

股债跷跷板效应是指在金融市场中，股票和债券这两种主要资产类别价格变动之间经常观察到的一种负相关关系。当股市上涨时，债券市场往往表现疲软；相反，当股市下跌或表现不佳时，债券市场可能会表现得更好。这种现象就像跷跷板一样，一头上升时另一头就会下降。此现象背后的逻辑包括市场的风险偏好、经济周期的影响以及利率变化等。因此，本文从宽基股指的景气度角度出发构造择时特征。

3.1.2 日间收益

收益率是重要的量价特征，短期的动量或反转效应均可由日间收益率捕捉。反转效应可解释为交易者的反应过度，动量效应则是交易者反应不足的体现。选取收盘价计算的日间收益率作为特征。

3.1.3 尾盘信息

国债期货的开盘和收盘往往是市场成交最活跃的时候，同时由于国债期货的交易时间规则，其尾盘最后15分钟位于股市收盘后，蕴含了交易者对于今日股市行情的反应，因此本文采用尾盘数据构造择时特征。

3.1.4 短期动量

动量策略广泛运用于择时和 CTA 策略中，我们采用 MACD 指标对标的短期趋势进行刻画，捕捉国债期货的短期动量。

3.2 样本选择和模型构建

3.2.1 数据预处理

本文使用 10 年期国债期货主力合约作为标的，使用其日频数据进行策略构建。由于期货合约存在换月的情况，统一采用后复权对合约价格进行处理，使合约行情连续，避免信号判断失真。复权因子的计算方式如下所示：

$$\text{Adj_factor}_i = \text{Adj_factor}_{i-1} * \frac{\text{Close}_{i-1, \text{old}}}{\text{Close}_{i-1, \text{new}}}$$

3.2.2 样本区间与集合划分

本文力求使用尽量长的时间区间进行研究。10 年期国债期货上市时间为 2015 年 3 月 20 日，扣除特征计算所需的数据长度，样本区间起始于 2015 年 7 月 1 日，使用 2024 年以前的数据进行模型的训练和验证，2024 年以后的数据进行策略测试(数据截至于 2024 年 6 月 30 日)。

在训练集和验证集的划分上，常用的方法有简单交叉验证、k 折交叉验证等。简单交叉验证按时间顺序前后划分训练集和验证集，能够避免未来数据的泄露，但由于 10 年期国债期货日频数据量本身较少，模型的表现受区间划分的影响大，得到的模型很可能并不稳健。本文采用带间隔的 5 折交叉验证进行训练集与验证集的划分，通过在每次划分时在训练集和验证集

之间设置足够长的间隔来避免数据的泄露。

3.2.3 模型训练与预测过程

如前所述，本文采用带间隔的 k 折交叉验证进行模型训练，取 $k=5$ ，将每日特征输入 XGBoost 模型，预测目标为未来一日以开盘价计算的收益率正负，使用早停和正则化避免模型的过拟合。训练得到 5 个模型，在测试时使用 5 个模型预测结果投票得到最终的预测结果。

3.3 策略表现

3.3.1 交易规则

策略每日调仓，每日生成第二日的交易方向信号，以第二日的开盘价交易。由于交易频率较低，策略表现受交易成本和滑点的影响较小，本文暂不考虑手续费及滑点。杠杆率设为 1。同时为了对比策略表现，使用国债期货多头作为基准。

3.3.2 策略表现汇总

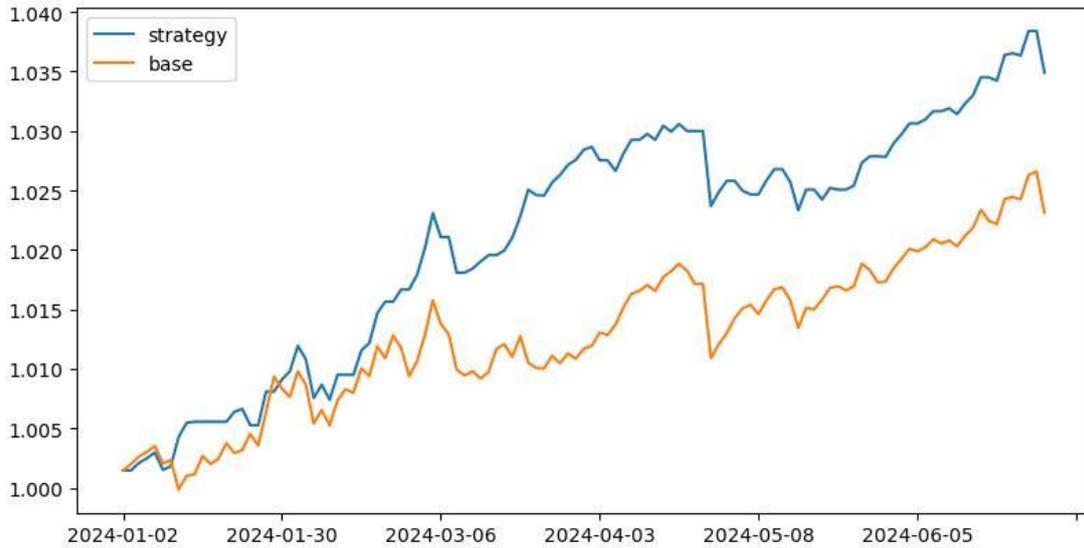
策略结果的统计分为两部分：一是策略在验证集上的表现，二是在测试集上的表现。

图表 2：策略验证集表现



数据来源：国贸期货

图表 3：策略测试集表现



数据来源：国贸期货

图表 4：各策略表现汇总

	年化收益率	夏普率	最大回撤
验证集表现	4.5%	1.34	4.2%
多头表现（验证集）	2.3%	0.61	7.5%
测试集表现	7.6%	3.84	0.7%
多头表现（测试集）	5.0%	2.28	0.8%

数据来源：国贸期货

3.4 总结及展望

本文基于日内尾盘信息、“股债跷跷板”效应、短期动量等量价特征构造择时因子，在模型方面使用 XGBoost 模型对输入因子信号进行处理，建立了国债期货择时模型。模型在 2015 年至 2023 年验证集上年化收益率为 4.5%，夏普率 1.34，最大回撤 4.2%，在 2024 年上半年测试集上年化收益达到 7.6%，夏普率 3.84，最大回撤 0.7%，两个区间内均优于多头表现。同时，策略也存在使用的特征较少，模型结构较为简单的改进空间。在后续的研究中，我们将考虑纳入更多的因子并使用更为完善的模型对策略进行改进。

风险提示：量化研究结果基于历史数据，过往表现不代表未来收益。

本报告中的信息均源于公开可获得的资料，国贸期货力求准确可靠，但不对上述信息的准确性及完整性做任何保证。

本报告不构成个人投资建议，也未针对个别投资者特殊的投资目标、财务状况或需要，投资者需自行判断本报告中的任何意见或建议是否符合其特定状况，据此投资，责任自负。

本报告仅向特定客户推送，未经国贸期货授权许可，任何引用、转载以及向第三方传播的行为均构成对国贸期货的侵权，我司将视情况追究法律责任。

免责声明