
摘要

本文构建了包括技术因子、宏观经济因子、其它大类资产走势因子、市场资金利率因子、新闻情绪因子在内的预测因子体系。其中，创新性地使用 ChatGPT 对新闻日报中的每条新闻情绪进行判断，并将结果汇总为新闻情绪因子；创新性地采用各宏观指标是否优于或弱于一致预期来作为宏观因子。

然后，本文采用线性回归、Lasso 回归、随机森林、多层感知机、Catboost 和 XGBoost 共 6 种模型对国债期货 TS、TF 和 T 合约未来 1 天、3 天和 5 天的收益分别进行预测，即每个模型进行 9 个测试，并基于预测结果构建交易策略。

在各模型中，XGBoost 模型在预测国债期货未来收益上表现最为稳定。在 9 个测试中，基于 XGBoost 模型的策略在验证集和测试集上均获得比市场基准更高的收益风险比。

本文还进一步探索了部分因子与国债期货未来收益之间的关系。新闻情绪因子总体上与国债期货的未来收益呈负相关关系，且 TS 合约对新闻情绪因子的敏感性低于 T 和 TF 合约；沪深 300 指数前一日的收益率越高，国债期货合约未来 3 天的收益倾向于越低，体现出了股债“跷跷板”的效应；前一日利率互换的数值与未来 3 天国债期货收益总体上呈正相关关系，体现了对未来利率的预期具有“均值回归”属性。

关键词：国债期货，量化投资，机器学习，资产定价

ABSTRACT

The article built a factor pool consisting of technical factors, macroeconomic factors, other asset trend factors, fund rate factors and news sentiment factor. The news sentiment factor was built by using ChatGPT to evaluate the sentiment of news. The macroeconomic factors were defied by whether macro indicators exceed their consensus expectations.

Then, future 1-day, 3-day and 5-day returns of Treasury futures were predicted based on linear regression, Lasso regression, random forest, multilayer perceptron, CatBoost and XGBoost, respectively. Trading strategies were built from prediction results. Therefore, each model was evaluated in 9 scenarios.

Among all model tested, XGBoost was the most resilient one because the trading strategies based on XGBoost realized higher ratio of return to risk on both valid and test sets in all scenarios.

The article further investigated the relationship between some factors and future returns of Treasury futures. There was a negative correlation between news sentiment factor and future returns of Treasury futures. The sensitivity of TS return towards news sentiment factor was lower than that of TF and T. The 3-day forward returns of Treasury futures tended to be lower when historical returns of HS300 index were higher. There was a positive correlation between IRS points on last day and future returns of Treasury futures, which showed the expectation of future interest rate was mean-reverting.

Key words: Treasury futures, quantitative investment, machine learning, asset pricing

目 录

摘要.....	I
ABSTRACT	II
第一章 引言.....	1
1.1 研究主题	1
1.2 研究意义	2
1.3 文章结构	2
1.4 本章小结	3
第二章 理论框架与文献综述.....	4
2.1 文献综述	4
2.1.1 国外关于国债期货的研究.....	4
2.1.2 国内关于国债期货的研究.....	5
2.1.3 国外关于线性模型和机器学习用于资产定价的研究.....	5
2.1.4 国内关于线性模型和机器学习用于资产定价的研究.....	6
2.1.5 文献综述总结.....	7
2.2 理论框架	7
2.2.1 资产未来收益预测.....	7
2.2.2 Lasso 回归	8
2.2.3 随机森林	8
2.2.4 多层感知机	9
2.2.5 CatBoost	10
2.2.6 XGBoost	10
2.3 本文主要研究角度	10
2.4 创新点	11

2.5 研究难点与解决方案.....	11
2.6 本章小结.....	12
第三章 数据描述及研究设计.....	14
3.1 国债期货主力合约.....	14
3.2 因子构造.....	15
3.2.1 技术因子	15
3.2.2 资金利率	16
3.2.3 新闻情绪	17
3.2.4 宏观因子	18
3.2.5 其它大类资产价格走势.....	19
3.2.6 因子筛选	19
3.3 模型预测、策略构建和回测.....	20
3.4 本章小结.....	21
第四章 实证分析.....	22
4.1 相关性分析.....	22
4.2 策略回测.....	27
4.2.1 线性模型	27
4.2.2 Lasso 模型	30
4.2.3 随机森林	32
4.2.4 多层感知机.....	35
4.2.5 CatBoost	38
4.2.6 XGBoost	40
4.3 本章小结.....	43
第五章 因子与国债期货未来收益的关系	45
5.1 新闻情绪因子与国债期货未来收益的关系.....	45
5.2 股债“跷跷板”	46

5.3 利率互换价格对国债期货的影响.....	47
5.4 本章小结.....	48
第六章 总结.....	49
6.1 主要结论.....	49
6.2 研究展望.....	50
参考文献.....	51
致 谢.....	53

第一章 引言

1.1 研究主题

随着众多做市商的加入，中国国债期货市场日渐发展成熟，交易量和持仓量也屡创新高。图 1-1 为国债期货各合约持仓量，国债期货合约的持仓量自 2020 年以来实现了快速上涨，截止 2023 年末，10 年期、5 年期和 2 年期国债期货持仓量涨幅分别为 150%、227% 和 306%。一方面，随着债券投资的日益增多，利率风险管理的需求愈发增多，国债期货作为中国市场上交易最为活跃的利率衍生品，覆盖了 2 年、5 年、10 年、30 年四个重要期限，是债券投资者进行利率风险管理的重要工具；另一方面，国债期货本身可以作为一项投资工具。不同于中国的债券主要在场外市场（银行间债券市场）交易，国债期货是标准化的场内交易的证券。因此，与直接投资于传统的债券相比，国债期货的交易摩擦更小，市场透明度更高。并且，由于国债期货的交易遵循保证金交易制度，投资国债期货占用的资金少，投资者能够以很低的成本增加杠杆。而且国债期货还可以比较便捷地进行多空双向交易，而在传统债券投资中，若要进行做空交易，则只能通过逆回购借券卖出，这会增加额外的借券成本。



图 1-1 国债期货持仓量

国债期货在中国金融市场中的重要性日益增加，越来越多的投资者将国债期货纳入到其投资组合中。因此，国债期货的价格走势预测也成为了投资者面临的一个重要问题。

线性模型和机器学习模型在权益资产收益预测的问题中取得了十分出色的表现。因此，本文的研究将线性模型和机器学习模型应用于国债期货的收益预测问题中，根据预测结果构建交易策略，并对比基于不同预测模型的交易策略的回测结果。

预计通过参数调优后，本文所构建的部分交易策略能够战胜市场基准，获得超额收益。

1.2 研究意义

从学术角度上看，本文的研究成果能进一步扩展线性模型和机器学习模型用于资产定价领域的研究成果，发掘线性模型和机器学习模型在中国国债期货市场上的预测能力，并对比在中国国债期货收益预测问题中不同预测模型之间的差别。此外，由于机器学习模型能够拟合变量之间的非线性关系，本文也将通过机器学习模型的拟合结果，探究各因子与国债期货未来收益之间的关系。

从实践角度看，本文的研究成果能够为国债期货做市商、套期保值者和其它国债期货投资者的实操提供一定的指导，为其提供了一种预测国债期货价格走势的方法，这对投资者的收益获取、头寸管理和交易时机的选择都将提供一定的帮助。

1.3 文章结构

本文第一章是本文的引言，提出了本文要研究的实证问题及其研究意义；本文的第二章进行了相关文献综述，介绍了本文研究的理论基础，并提出了本文的研究角度；本文的第三章进行了研究设计，具体包括因子体系构建、因子降维、模型预测和交易策略生成的具体方法；本文的第四章是实证分析，使用各模型对国债期货未来收益进行预测，构建交易策略并对其进行回测；本文的第五章对部分因子与国债期货合约之间的关系进行探究，并给出了经济学上的解释；本文的第六章是对全文的总结。

1.4 本章小结

本章介绍了本文的研究主题，即国债期货价格走势预测。除此之外，本章还介绍了本文研究的学术和实践意义。

从学术角度上看，本文的研究成果能进一步扩展线性模型和机器学习用于资产定价领域的研究成果，发掘线性模型和机器学习模型在中国国债期货市场上的预测能力；从实践角度看，本文的研究成果能够为做市商、套期保值者和其它国债期货投资者的收益获取、头寸管理和交易时机的选择都将提供一定的帮助。

第二章 理论框架与文献综述

2.1 文献综述

2.1.1 国外关于国债期货的研究

国内外学者对国债期货进行过丰富的研究。国外文献中国债期货相关研究如下。Smales (2021) 通过研究发现美国国债的拍卖对 10 年期美国国债期货的走势有显著影响；在国债拍卖之后的一个区间中，10 年期国债期货市场表现出更大的波动性以及更高的市场交易量；当国债拍卖过程中出现更高的投标覆盖率时，表明市场对所拍卖国债有更强烈的需求，进而导致接下来一个区间中国债期货更容易获得正收益以及更低的收益波动性。Zhang, Zhao, Nan (2021) 使用 GARCH-MIDAS 模型研究了国债期货波动和经济政策不确定性之间的关系；作者用国债期货的已实现波动率和经济不确定性指数作为模型变量，研究发现国债期货的已实现波动率和经济政策不确定性对国债期货的长期波动性有显著影响，此结论对中国和美国的国债期货市场均成立。Ruan, Zhou, Yin et al (2021) 采用多重分形去趋势波动分析法 (MF-DFA) 和多重分形去趋势互相关性分析法(MF-DCCA)，研究中国国债期货的对冲有效性；研究发现 10 年期国债期货和现货具有持续互相关性，而 5 年期国债期货和现货不具有显著的互相关性；然而 10 年国债期货和 5 年以及更长期限的现货间具有持续互相关性，因此 10 年国债期货可作为对冲利率风险的良好工具。Tang, Yang, Yu (2018) 研究了中国国债期货和现货市场的价格发现和波动溢出效应，其研究发现中国 5 年期国债期货市场发挥了重要的价格发现功能；其研究还发现了国债期货和现货市场存在波动率溢出效应，而且，国债期货市场对现货市场的波动率溢出效应强于现货市场向期货市场的波动率溢出效应。Dungey, Hvozdyk (2012) 研究了美国国债期货和现货的同步跳水行为；其研究采用高频数据，研究发现同步跳水的行为在短期限的品种中出现得更加频繁；当宣布重要的宏观经济数据时，国债期货与现货同步跳水的概率会有所改变，尤其当非农就业数据、CPI、GDP 和零售数据超预期时，会增加国债期货和现货同步跳水的概率。Bansal, Connolly, Stivers (2010) 研究了 1997-2005 年间美国股票指数和 10 年期国债期货的表现；作者将这段时期分为高压时期和低压时期，通过对比研究发现股票市场压力对国债期货定价有显著影响；以及在股票市场压力相对较高的时期，股票和

债券组合的分散化带来的益处更大。Zou, Rose, Pinfold (2007) 采用线性和非线性 GARCH 模型研究了信息流对 3 年期澳大利亚国债期货收益波动性的影响；其研究发现了显著的非对称信息效应，坏消息比好消息对国债期货的收益波动性有更大的影响。Brandt, Kavajecz, Underwood (2007) 研究了美国国债期货市场的价格发现功能，作者分析了订单流和信息流在不同市场间的传导，发现在国债期货市场和现货市场上的订单流驱动了国债期货和现货价格的变动，从而促进了国债期货的价格发现功能；作者还发现价格发现的方向和强度受到回购利率和市场流动性的影响。Holder, Qi, Sinha (2004) 研究了美国国债期货市场中的机构结构对国债期货价格的影响；研究发现交易时长和交易员的数量对国债期货的收益有正相关关系，而交易量和国债期货的收益有负相关关系。Park, Switzer (1996) 研究了不同期限的利差的均值回归行为，基于对收益率曲线形态变动的观点开发交易策略，并采用不同期限的国债期货组合来实现基于利差均值回归现象的交易策略，其交易策略在实现了明显高于无风险收益率的收益的同时，保持了较小的风险。

2.1.2 国内关于国债期货的研究

国内文献中也有一些国债期货相关的研究。李骏琪、杨垚立 (2022) 采用卷积神经网络 (CNN) 和时序卷积神经网络 (TCN)，对国债期货订单簿数据中的特征进行学习，以此来预测国债期货的价格变动，其研究结果发现采用 CNN+TCN 模型来预测国债期货价格变动，模型在样本外的预测能力显著高于支持向量机等模型。周凯文、黄泽松 (2022) 利用遗传规划从国债期货数据中提取特征，并使用机器学习模型对国债期货未来收益进行预测。

2.1.3 国外关于线性模型和机器学习用于资产定价的研究

在传统的研究中，线性模型被广泛地应用于资产定价领域。Sharpe (1964) 提出了 CAPM 模型，CAPM 模型拟合了股票收益与市场组合超额收益之间线性关系。Fama, French (1993) 将市值因子和账面市值比因子加入到模型中，将模型扩展为三因子模型。在后续研究中，Fama, French (2015) 将盈利因子和投资因子加入到模型中，将模型进一步扩展为五因子模型。

机器学习模型十分适合用于处理复杂的预测问题，机器学习模型在模式识别、图像识别、自动驾驶等人工智能领域取得了巨大的成就。机器学习模型采用复杂的模型结构去识别因变量与多个自变量之间的复杂非线性关系。因此，机器学习模型在人工

智能领域取得了优异的表现。金融资产的价格与证券估值、公司财务情况、宏观经济、交易者情绪等多种变量有联系，且部分变量与金融资产价格之间具有非线性关系。因此，机器学习模型正好也适用于金融资产定价领域。近年来，国内外学者将机器学习引入到资产定价领域。以下是国外文献中的一些尝试。Zhu, Wu, Wells(2023)采用了两个稀疏机器学习模型来解释和预测基于金融新闻的股票收益和行业收益。其研究采用新闻嵌入的 UMAP 系数选择模型、行业相关的新闻稀疏编码器模型对股票收益和行业收益进行预测，模型表现优于传统的资产定价模型。Drobertz, Otto (2021) 研究了使用线性模型和机器学习模型预测欧洲市场股票收益的表现；其研究发现机器学习模型能够实现比线性模型更好的预测力，但必须经过充分的训练和微调，机器学习模型才能够克服高维度问题和过拟合问题；其研究发现最具预测力的因子为股票趋势因子和估值比率相关的因子；其研究发现预测模型中，表现最好的模型为神经网络模型；然而，如果不直接使用未来收益，而将预测标签进行类别化处理，并使用支持向量机来预测，则可以得到比神经网络模型更好的策略表现。Wang (2021) 研究了使用机器学习模型进行数字货币的资产定价，作者在研究了波动率风险因子（如已实现波动率、已实现偏度、跳跃因子）和流动性因子（买卖价差、交易量增速），使用了股票因子和数字货币特有的因子，并使用机器学习模型来对资产收益进行预测，其研究发现机器学习模型能很好地使用于数字货币定价领域。Gu, Kelly, Xiu (2020) 采用对比研究的方法，使用线性模型、树模型和神经网络预测股票收益，发现机器学习能够提高实证资产定价的效果；因为金融资产数量较少和信噪比较低的原因，深度学习模型在资产定价领域表现不佳；其研究还发现机器学习在预测规模更大、流动性更好的股票和组合时具有更好的表现；同时，研究发现最有效的预测因子是收益动量和反转的相关因子，其次是波动率和估值因子。Gabriel ANGHEL (2020) 研究了机器学习模型在新兴股票市场资产定价中的应用，其研究使用机器学习分类模型，发现流动性、条件偏度、波动率因子对股票长期收益具有预测力。国外学者在采用机器学习模型进行权益领域的资产定价时，取得了不错的效果，预计在国债期货领域，机器学习模型也能较好地预测国债期货的未来收益，后续的研究将进一步验证这个猜想。

2.1.4 国内关于线性模型和机器学习用于资产定价的研究

国内文献中也有一些使用线性模型和机器学习进行资产定价的相关研究。许杰、祝玉坤、邢春晓（2022）对机器学习在金融资产定价领域的应用进行了综述，文章介绍了传统机器学习模型、迁移学习、集成学习、深度学习等方法。方毅、陈煜之（2022）

对比了 CAPM、DNN、LSTM、SVM 等多种模型在资产定价中的使用，并分别测试了固定窗口预测和滚动窗口预测；其研究结果表明，无论是在滚动窗口预测问题中还是在固定窗口预测问题中，机器学习模型的样本外预测能力均优于线性模型，且机器学习模型的预测误差显著低于线性模型的预测误差。李仁宇、叶子谦（2023）研究了机器学习对基金收益预测的能力，其研究采用了 30 个基金异象因子，采用 Light GBM 作为机器学习模型的代表，验证了机器学习模型在预测基金收益方面的能力优于传统的线性模型。李斌、邵新月、李玥阳（2019）研究了机器学习在基本面量化投资中的应用，其研究采用了 A 股市场 96 个异象因子，使用 Lasso 回归、岭回归、支持向量机、梯度提升树等机器学习算法来预测股票收益，基于其预测结果得到的投资组合优于传统的线性算法。孙达昌、毕秀春（2018）研究了基于深度学习算法的高频交易策略；其研究采用了每秒两次的 tick 级别沥青期货数据，并使用卷积神经网络和 LSTM 神经网络分别进行涨跌分类预测；其研究结果发现基于 LSTM 网络的策略的胜率和期望收益最高，其次是卷积神经网络，但是两者的收益均优于普通的浅层人工神经网络。

2.1.5 文献综述总结

一方面，国内外学者对影响国债期货的因素进行过丰富的研究，也对国债期货市场和现货市场联动性、国债期货市场和权益市场的波动率溢出效应等国债期货相关的问题进行过充分的研究；另一方面，国内外学者对使用线性模型和机器学习进行资产收益的预测进行过很多尝试，这些研究主要集中在权益资产和商品期货领域。纵观国内外相关研究，很少有研究聚焦于采用线性模型和机器学习模型对国债期货的收益进行预测的文献。

2.2 理论框架

2.2.1 资产未来收益预测

基于当前时点 t ，对资产未来收益的预期为

$$E[P_{t+k} - P_t | I_t] \quad (2-1)$$

其中， $P_{t+k} - P_t$ 为资产的未来收益， I_t 为 t 时刻的信息集。基于对资产未来的收益的预期，金融市场的投资者做出相应的投资决策。

I_t 是 t 时刻的信息集， I_t 具有丰富的内涵，包含了 t 时刻可以获得的所有信息。在

实际预测中，我们无法直接使用 I_t ，需要从 I_t 中提取出一些与资产的未来收益具有相关性的变量，并通过这些变量来预测资产的未来收益，我们将这些变量称为因子。

在获得了因子集后，我们需要拟合资产组合未来收益与因子集之间的关系。可以使用线性模型来拟合因子集与未来收益的关系，也可以使用非线性模型来拟合因子集与未来收益之间的关系。

早期的资产定价领域的研究大多是拟合因子与资产的收益或收益率之间的线性关系。如 Sharpe (1964) 提出的 CAPM 模型和 Fama, French (1993) 提出的三因子模型。

随着资产定价的发展，用于资产定价的因子数量越来越多，且不同因子之间可能存在多重共线性，有的因子与预测目标之间的关系也并非简单的线性关系。因此，为了更好地拟合因子集与资产收益或收益率之间的关系，学界开始将机器学习模型引入资产定价领域。如 Gu, Kelly, Xiu (2020) 采用对比研究的方法，发现机器学习模型能够提高实证资产定价的效果。

2.2.2 Lasso 回归

Tibshirani (1996) 提出了一种对线性回归进行改进的方法 Lasso 回归，Lasso 回归在最小化数据的残差平方和的同时，限定所有参数绝对值之和小于某个特定值，Lasso 回归可以表示为

$$\min_{\beta} \frac{1}{N} \|y - X\beta\|_2^2 \quad s.t. \quad \|\beta\|_1 < t \quad (2-2)$$

采用拉格朗日乘数法，可将上述表达式写为

$$\min_{\beta} \frac{1}{N} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \quad (2-3)$$

我们可以将上述表达式中的 λ 看作一个惩罚项系数，当 λ 值变大时，惩罚项在优化目标中占的比重越大，此时会得到更简单的模型，即更多变量被丢弃。 λ 的选择是一个权衡， λ 值过小，则模型容易过拟合； λ 值过大，则模型容易欠拟合。

相比于传统的线性回归模型，Lasso 模型能够降低模型出现过拟合的可能性，提高模型的样本外泛化能力。

2.2.3 随机森林

Ho (1995) 提出了一种基于传统决策树的改进方法，这种改进方法是在随机的特征集上训练决策树，并将决策树的结果组合起来形成模型最终的结果。

在对一颗决策树进行训练时，若采取较低的树深，则决策树无法对数据实现很好的拟合，若采取较高的树深，则决策树容易出现过拟合，模型在样本外的预测能力会大大降低。在随机的特征集上训练多颗树深较低的决策树，并将不同决策树组合起来的方法，可以在避免过拟合的同时，提升模型的拟合效果和预测效果。

Ho (1998) 提出了可以在决策树的每个节点分裂时采用不同的随机特征子集，提升模型的随机性，从而进一步提升模型的泛化能力。

相比于传统的决策树，随机森林的机制可以保证在训练更多的决策树以提升模型的拟合效果时，模型的泛化能力仍然能得到保证。

2.2.4 多层感知机

多层感知机是一种具有仿生特征的机器学习模型，模型通过输入层输入信号，并将信号通过权重传导到下一层，下一层的数值经过激活函数处理后，再传导到下一层。

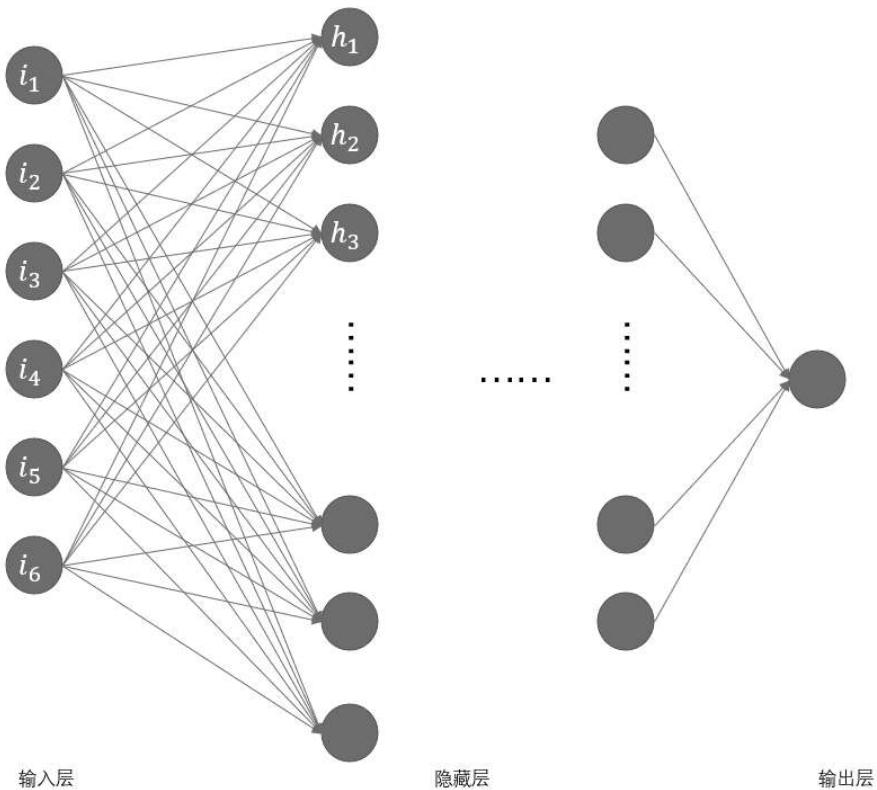


图 2-1 神经网络结构示意图

如上图所示，第一个隐藏层中的第 k 个元素的计算方法为

$$h_k = a(\sum_j w_{jk} i_j) \quad (2-4)$$

其中, w_{jk} 是多层感知机两层神经元之间的连接权重; i_j 是上一层神经元的数值, h_k 是下一层神经元的数值; $a(\cdot)$ 是激活函数, 常见的激活函数包括 Sigmoid、ReLU 等。多层神经网络结构配合激活函数使得该模型能够捕捉输入与输出之间的非线性关系。

在模型训练的过程中, 使用反向传播和梯度下降, 不断降低损失函数的数值, 得到最终模型参数。

2.2.5 CatBoost

CatBoost 是一种对类别特征具有较强处理能力的 boosting 算法, CatBoost 通过将类别特征转化为数值特征、特征交叉、排序提升和对称二叉树, 提供具有高精度和高泛化能力的预测算法。

CatBoost 模型采用如下公式将类别特征转化为数值特征:

$$\hat{x}_k^i = \frac{\sum_{x_j \in D_k} \mathbb{1}_{x_j^i = x_k^i} y_j + ap}{\sum_{x_j \in D_k} \mathbb{1}_{x_j^i = x_k^i} + a} \quad (2-5)$$

其中, 将训练集数据 D 随机打乱, \hat{x}_k^i 是第 k 个样本的第 i 个类别特征的转化值, D_k 是打乱的数据集中排在 k 样本之前的所有样本, a 是先验样本的数量, p 是样本的先验值。基于这样的处理, CatBoost 模型巧妙地将类别特征转化为数值特征。

2.2.6 XGBoost

XGBoost 模型是由 Chen、Guestrin (2016) 提出的, XGBoost 模型是在 GBDT 的基础上, 进行了一系列改进得到的机器学习算法。

XGBoost 模型采用多个回归树对预测目标进行预测。模型采用的损失函数为

$$L^t = \sum_{i=1}^n l(y_i, \hat{y}_i^{t-1} + f_t(x_i)) + \Omega(f_t) \quad (2-6)$$

其中, 前面一项是对模型残差的拟合, 后一项是正则项。为了求解上述的最优化问题, 用损失函数的二阶泰勒展开作为损失函数的近似替代, 并进行优化求解。

XGBoost 模型具有精度高、运行速度快等特点, 在机器学习领域表现出色。

2.3 本文主要研究角度

目前, 国内外对使用线性模型和机器学习对国债期货收益进行预测的研究很少。因此, 本文将构建各大类预测因子 (技术指标、宏观经济指标、其它大类资产价格走

势、市场资金利率、新闻情绪等)，采用线性模型和机器学习模型对国债期货进行收益预测和交易策略构建，并详细地对比线性模型及各种机器学习类模型在预测国债期货收益与构建投资策略上的表现。

2.4 创新点

1、采用的预测因子种类全面，包括技术指标、宏观经济指标、其它大类资产价格走势、市场资金利率、新闻情绪等，这些指标包括了影响国债期货价格变化的各种因素，指标体系全面。

2、新闻情绪指标创新性地使用 ChatGPT 对新闻日报中的每条新闻情绪进行判断，并将结果汇总为新闻情绪指数。

3、对于宏观指标，为了更好地反映宏观指标对国债期货收益的影响，没有直接采用宏观指标本身的值作为因子，而是创新性地采用各宏观指标是否优于或弱于一致预期来作为因子。

4、在使用线性模型和机器学习模型对国债期货的未来收益进行预测时，采用未来 1 天收益、未来 3 天收益和未来 5 天收益三种不同的预测目标，并对比模型对国债期货不同期限未来收益的预测效果。

5、在使用机器学习模型对国债期货的未来收益进行拟合后，使用机器学习模型的拟合结果对部分变量与国债期货未来收益的关系进行探究，探究二者之间的线性及非线性关系，并寻求现象背后的经济学解释。

6、初始构建的因子池中因子数量较多，而国债期货日频数据的数量有限，为了避免模型过拟合，采用分步回归的方法进行因子筛选。对于不同国债期货合约、不同预测目标，分别采用分步回归筛选出最合适的因子集。

2.5 研究难点与解决方案

相比于在权益资产上使用机器学习模型进行收益预测，在国债期货上使用机器学习模型进行收益预测存在一些新的挑战。在权益市场上，标的数量较多，标的数量能达到几千甚至上万，在绝大多数将机器学习用于权益资产定价的相关研究中，学者将不同股票的数据都作为样本点放到同一个机器学习模型的训练集中，并将训练得到的模型用于对所有权益资产的预测，最后做多预测收益率最高的一部分股票并做空预测

收益率最低的一部分股票，进而获取相对较强的股票与相对较弱的股票之间的收益差额。因此，机器学习模型并不一定要特别好地拟合股票的未来收益，模型只需要能较好地拟合不同股票未来收益的相对强弱关系，这样的投资策略就能够取得较好的收益。而在国债期货市场中，现行中国国债期货市场上仅存在 2 年期、5 年期、10 年期和 30 年期共 4 个品种。其中，2 年期国债期货于 2018 年上市，5 年期国债期货于 2013 年上市，10 年期国债期货于 2015 年上市，30 年期国债期货于 2023 年上市，各合约上市至今的存续时间也较短。因此，标的数量较少和上市时间较短意味着数据量相对较少，这使得机器学习模型的训练可能面临数据不足的问题。为了解决这个问题，本文采用模型结构相对简单，不容易过拟合的机器学习模型和线性模型进行国债期货未来收益的预测。

另一方面，国债期货每个期限的品种都有当季、次季和远季三个合约。通常来说，三个合约中只有一个被大多数交易者交易的主力合约，其余两个合约交易不活跃，其价格也不能很好地反映市场的供求关系。因此，在对国债期货合约的价格进行分析时，我们最好只分析主力合约的价格序列。不过，主力合约并不是一直持续的，每个季度都会进行主力合约的切换，在主力合约切换的时点，想要继续持有国债期货头寸的投资者会将原主力合约平仓并在新的主力合约上开仓。因为期货合约的基差会随着到期时间的临近而逐步收敛，所以主力合约切换带来的问题是，在主力合约切换的当日，主力合约的价格会有一个明显的跳跃。如果我们要进行连续性的预测，则需要判断出每个交易日的主力合约，并生成主力合约的复权数据，这样就可以避免价格跳跃，避免出现异常数据点。

由于标的较少，国债期货合约难以实现权益资产上的择券策略（做多相对较好的一部分券，做空相对较差的一部分券），更合理地是在单个品种上进行预测的择时策略。更具体来说，就是当预测收益率大于 0 时做多期货，当预测收益率小于 0 时做空期货。

虽然使用线性模型和机器学习模型进行国债期货的收益预测存在一些难点，但是这样的研究是具有现实意义的。

2.6 本章小结

本章对国债期货相关文献、线性模型和机器学习模型用于资产定价相关文献进行了综述，进而提出了本文的主要研究角度——采用线性模型和机器学习模型对国债期

货进行收益预测和交易策略构建。

本文将构建全面的因子体系，并通过创新性的方法构建新闻情绪因子和宏观因子，然后通过线性模型和机器学习模型对国债期货未来1、3、5天的收益进行预测。本文还将对比不同模型对国债期货的预测效果，并探讨部分因子与国债期货未来收益之间的关系。

第三章 数据描述及研究设计

国债期货合约是在中国金融期货交易所交易的标准化期货合约。国债期货合约共有 4 种不同的期限，分别为 2 年期、5 年期、10 年期和 30 年期国债期货合约。随着国债期货市场的发展，今年来国债期货合约的持仓量也逐渐升高。截止 2023 年末，2 年期国债期货合约持仓量为 71532；5 年期国债期货合约持仓量为 118543；10 年国债期货合约持仓量为 205045；30 年期国债期货合约持仓量为 48512。10 年期以内的国债期货合约随着期限的增加，持仓量也更多；今年上市的 30 年期国债期货合约目前持仓量最少，但是增长速度可观，其持仓量有望逐步赶上其它合约。



图 3-1 国债期货合约持仓量

3.1 国债期货主力合约

对于某一特定期限的国债期货合约，在每个交易日都有当季、次季、远季三个合约可交易。例如在 2023 年 12 月 29 日，10 年期国债期货有 T2403、T2406、T2409 三个合约可以交易（T 是 10 年期国债期货合约的代码，后面的 4 个数字代表合约的交割月份）。

国债期货的每个合约的存续期都不足一年。当我们对国债期货的价格走势进行

连续性的研究时，主力合约更换日的价格跳跃会给我们的研究带来问题。因此需要自行合成一条连续的国债期货合约的价格曲线，以用于后续的研究。

此处将前一交易日三个存续合约中持仓量最大的国债期货合约定义为当日的主力合约，这个合约是市场参与者主要参与交易的合约，其价格走势最能反映国债期货合约整体的价格走势。在主力合约切换的时点，由于新旧合约的价格不一致，在该时点会形成一个价格的跳跃。为了得到一条连续的价格曲线，在合约切换时进行价格后复权，这样就可以得到一条连续的价格曲线。

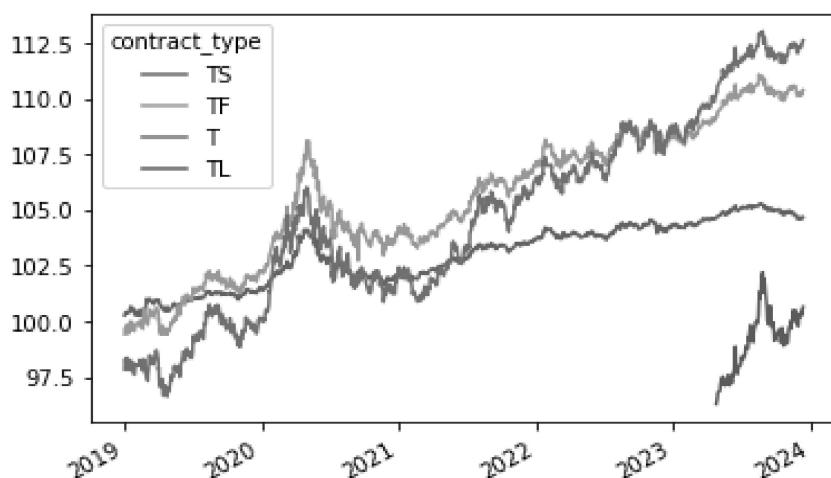


图 3-2 主力合约价格走势

TS、TF、T 和 TL 分别代表 2 年、5 年、10 年和 30 年期国债期货。观察以上曲线的走势可知，2020 年疫情突发后的降息潮带动国债期货价格达到了顶峰，后随着疫情的逐渐好转而逐步下降。自 2021 年开始至今，国债期货价格指数整体呈上升走势。

3.2 因子构造

下方将构造用于预测模型的各种因子。数据选取的时间区间为 2019 年 6 月 13 日-2023 年 12 月 1 日。其中 2019 年 6 月 13 日-2022 年 10 月 31 日的数据划分为训练集；2022 年 11 月 1 日-2023 年 5 月 12 日的数据划分为验证集；2023 年 5 月 15 日至 2023 年 12 月 1 日的数据划分为测试集。

3.2.1 技术因子

技术分析是基于证券历史交易的量价数据生成指标，来分析证券的未来价格走势

的一种方法。在市场有效性达不到弱式有效市场前提的下，技术分析可以帮助我们获取超额收益。

采用 pandas_ta 程序包生成技术因子，生成的技术因子总数为 172 个。

3.2.2 资金利率

资金利率是持有债券的资金成本，资金利率对债券收益率尤其是短久期债券的收益率有较大影响。因为国债期货价格和国债收益率有很强的相关性。因此，资金利率也会对国债期货的价格变动造成很大的影响。资金利率反映了当下使用资金的价格，而资金的交易量则体现了资金的供给需求关系及债券市场的杠杆程度，反映了市场参与者对债券市场的情绪，因此银行间和交易所债券借贷的交易量也会对国债期货的价格变动产生一定影响。

我们将资金利率相关的指标也加入到国债期货价格的预测因子中。使用到的资金利率指标如下：

表 3-1 资金面因子

指标	含义
R001:加权平均	隔夜银行间质押式回购加权平均利率
成交量:R001	隔夜银行间质押式回购成交量
R007:加权平均	一周银行间质押式回购加权平均利率
成交量:R007	一周银行间质押式回购成交量
R014:加权平均	两周银行间质押式回购加权平均利率
成交量:R014	两周银行间质押式回购成交量
GC001:加权平均	隔夜上交所回购加权平均利率
GC001:成交量	隔夜上交所回购成交量
GC007:加权平均	一周上交所回购加权平均利率
GC007:成交量	一周上交所回购成交量
GC014:收盘价	一周上交所回购加权平均利率
GC014:成交量	两周上交所回购成交量
SHIBOR:1 个月	1 月期上海银行间同业拆放利率
SHIBOR:3 个月	3 月期上海银行间同业拆放利率
SHIBOR:6 个月	6 月期上海银行间同业拆放利率
利率互换:FR007:1 年	1 年期 FR007 利率互换，体现对未来 1 年内的资金价格预期
利率互换:FR007:5 年	5 年期 FR007 利率互换，体现对未来 5 年内的资金价格预期

3.2.3 新闻情绪

财经新闻是金融交易员们十分关注的信息，新闻中所蕴含的情绪会影响交易员们买入或卖出金融资产的意愿，从而对金融资产的价格走势造成影响。国债期货价格走势也受到宏观新闻情绪的影响，因此此处构建宏观新闻情绪指数，作为国债期货价格的预测因子之一。

华尔街见闻是中国领先的金融信息平台。华尔街见闻会日常推送各类资产行情和热点财经新闻。华尔街见闻网站上的新闻十分繁多，重要性并不一致，有的新闻的重要性可能远低于另一些。不过，华尔街见闻每天上午 7 点左右会发布“华尔街见闻早餐”系列新闻汇总。此系列的汇总在每日开盘前发布，涵盖了前一天交易日及隔夜发生的重要财经新闻。每天的“华尔街见闻早餐”又包括市场概述、市场收报、要闻和要闻详情几个部分。要闻详情部分的是对前一天交易日及隔夜财经要闻的汇总，该部分的新闻是最能对当日开盘后交易产生影响的新闻汇总。因此，本文选取每个交易日“华尔街见闻早餐”中的要闻详情下的新闻，来构建新闻情绪指标。

要闻详情	要闻详情
国内宏观	
中国5月70城房价：各线房价同比涨幅回落或降幅扩大	A股冲高回落！成交额创2月底以来最高 大金融尾盘回落 多只券商股炸板
一线城市新房和二手房价格环比一涨一平，二、三线城市均为下降。	6月15日，A股冲高回落，截至收盘，沪指涨0.5%，深成指涨0.95%，创业板涨1.05%。板块方面，券商、保险等大金融板块全线爆发临近尾盘有所回落，教育股、中船系冲高，油气、煤炭、磷化工跌幅居前。
发改委：CPI将延续运行在合理区间，能完成全年3%的目标	港股全天高位震荡 券商股午后暴涨 光大证券尾盘跳水 新东方在线一周四倍
发改委政研室副主任、委新闻发言人孟玮表示，将全力推动扎实稳住经济一揽子政策措施尽快落地见效，确保二季度经济实现合理增长；确保火电出力水平好于常年，多发水电节省燃煤消耗；随着下半年猪肉消费需求逐步增加，生猪价格有望总体回升。	盘面上，科技、教育、券商、汽车股集体冲高。
财政部：1-5月全国一般公共预算支出98059亿元，比上年同期增长5.9%	习近平同俄罗斯总统普京通电话
社会保障和就业、教育、卫生健康等民生重点领域支出得到有力保障。	中方愿同俄方继续在涉及主权、安全等核心利益和重大关切问题上相互支持，密切两国战略协作，加强在联合国、金砖国家、上海合作组织等重要国际和地区组织的沟通协调，促进新兴市场国家和发展中国家团结合作，推动国际秩序和全球治理朝着更加公正合理的方向发展。
国内公司	
茅台召开股东大会，丁雄军称今年目标营收1259亿元，增速15%	国常会：既果断加大力度、稳经济政策应出尽出，又不超发货币、不透支未来
针对投资者关于提高分红比例的问题，茅台表示，希望股东对分红能有更加平和的心态。	立足当前稳增长，着眼长远促进经济持续健康发展。抓住时间窗口，注重区间调控，既要加大力度、稳经济政策应出尽出，又不超发货币、不透支未来，着力保市场主体保就业稳物价，稳住宏观经济大盘，支持实体经济健康发展，支持民间资本发展创业投资。会议强调，地下综合管廊是城市“里子”工程，带动能力强，是一举多得的代表性项目，要结合老旧管网改造统筹推进。科学规划，健全收费回报机制，吸引社会资本参与。引导金融机构发放长期贷款。
美国央行加息25基点，利率升至十三年新高，暗示必要时会采取更强有力行动	韩正：积极推动新增产能落地，加强煤炭储备能力建设
美国央行公布的利率决议显示，英国央行如期加息25个基点，将政策利率提高至1.25%。前值为1%。1.25%的关键利率也是2009年1月英国经济受到全球金融危机冲击以来的最高水平。英国央行发出迄今为止最强烈的加息信号，称准备在必要时采取更大行动来遏制通胀。	要依法加快产能核增、用地用草、安全生产、环境影响评价等手续办理进度，积极推动新增产能落地，加强煤炭储备能力建设，促进国内能源市场平稳运行。
瑞士央行意外加息50个基点，为七年来首次上调利率，不排除外汇干预	中国央行等额续作MLF，利率不变
瑞士央行强调，不排除未来进一步加息，以稳定通胀；准备干预外汇市场，以遏制瑞士法郎的过度升值或贬值。	中信证券：随着经济的复苏，利率有上行风险，货币政策重心也可能从宽货币向宽信用转移。
为避免加剧通胀，欧洲央行可能会选择中和新的危机应对工具下的债券购买行为。欧洲央行通过任何新工具进行的债券购买，可能都会涉及到出售投资组合中的其他证券，以便不会破坏欧	新华社：坚定信心，巩固恢复态势——前5个月中国经济走势观察
	展望全年，摆在中国经济面前的，是一张难度不小的“考卷”。放眼全球，国际环境依然复杂严峻，乌克兰危机加剧新冠肺炎疫情对全球经济的影响，外部不确定性增大；纵观国内，疫情防

图 3-3 华街见闻要闻详情示例

2022 年 6 月 16 日以前，要闻详情模块直接对所有要闻进行陈列，因此本文选取

要闻详情中所陈列的所有要闻进行分析；2022年6月17日开始，要闻详情模块对新闻进行归类，分为国内宏观、国内公司、海外宏观和海外公司几大类。国债期货的底层资产是国债，国债主要与宏观经济整体的相关度较高。因此，从2022年6月17日开始，本文仅将要闻详情中的国内宏观和海外宏观的新闻纳入分析范围。

在获取了每日的新闻列表后，接下来需要将新闻文本这种非结构化的数据转化为结构化的数值，这样方可作为因子加入到国债期货的预测模型中。Chat GPT是Open AI研发的聊天机器人，Chat GPT在文本信息处理方面具有极强的能力。因此，本文选择用Chat GPT对每一条新闻对应的情绪进行判断。若新闻包含积极情绪，则返回1；若新闻包含消极情绪，则返回-1；当无法判断时则返回0。先采用Chat GPT对所有新闻的进行判断，然后取每日各新闻情绪的均值，作为当日的新闻情绪指数，将此新闻情绪指数作为一个因子用于后续国债期货的预测。

3.2.4 宏观因子

国债期货的标的资产是国债，国债期货的价格走势包含着市场参与者对未来市场无风险利率的预期。而市场无风险利率的走势则蕴含着对未来宏观经济情况的预期。例如，若预期未来宏观经济增长强劲、通胀高企，则对未来无风险利率的预期较高，国债期货价格较低；若预期未来宏观经济增长乏力、市场通缩，则对未来无风险利率的预期较低，国债期货价格较高。

市场上的研究机构会定期发布对各种宏观经济数据的预测，供市场参与者借鉴。市场参与者会结合这些预测，进行国债期货合约的定价以及日常的投资交易决策。当宏观经济数据公布时，若其与此前的预测不一致，则会引起市场参与者修正自己的预期，并通过对国债期货的买卖行为来影响国债期货的价格走势。因此，本文将宏观经济数据也结合到了因子体系中。

将宏观指标结合到因子体系中面临一个棘手的问题，由于宏观指标的更新频率较低（通常为每月或每季度发布），因此需要一些特殊的处理才能将其结合到日频的因子体系中。另一方面，宏观指标的绝对水平对国债期货价格的影响不大，其与预期的差别才会对国债期货的价格造成影响。新的宏观数据公布时，若宏观数据超出或低于预期，则会对市场造成一定的冲击，从而导致国债期货价格的上升或下降。因此，本文采用各宏观指标是否超出或弱于万得一致预期作为国债期货价格走势的预测因子。

选取的宏观指标包括：GDP同比、CPI同比、PPI同比、规模以上工业增加值同比、固定资产投资累计同比、社会消费品零售总额同比、进口金额同比、出口金额同

比、贸易差额、M2 同比、新增人民币贷款、社会融资规模存量同比、PMI、失业率。

GDP 同比、CPI 同比、PPI 同比、PMI 的发布时间在开盘时或开盘前。因此，若数据发布日超预期，则指标设为 1；若数据发布日弱于预期，则指标设为 -1，其余日期指标均设为 0。

其余数据发布时间均在开盘后，因此将数据发布日下一个交易日的指标按超过或弱于预期分别设置，其余日期的指标设为 0。

3.2.5 其它大类资产价格走势

金融市场上的资产可以大体分作权益、债券、商品、外汇四大类。四大类资产在不同的市场环境和经济环境下的表现具有千丝万缕的联系。

通常来说，当投资者减持某一大类资产时，为了维持投资组合的头寸会增持别的大类资产，这种不同大类资产市场间的资金流动带来了“跷跷板”效应。例如，当权益资产下跌时，资金流出权益市场，转而进入债券市场寻求避险，这就会助推债券市场的价格走高，进而推动国债期货的价格走高。

通常来说，国债被投资者作为安全性很高的防御资产，当其它资产下跌时，更多资金寻求进入国债市场获取稳定的收益。当其它资产上涨时，资金从国债市场流出以寻求更高的收益。国债期货价格会与其它大类资产价格呈此消彼长的“跷跷板”关系。因此，在预测国债期货的因子池中加入其它大类资产价格走势有助于提升预测效果。

具体的，将沪深 300 指数、南华商品指数和人民币兑美元即期汇率定盘的日频收益率滞后一期，并作为因子加入到因子池中。

3.2.6 因子筛选

本研究使用日频数据，从 2019 年 6 月 13 日-2023 年 12 月 1 日，共有 1088 个样本点。由于日频样本的数量较少，而技术因子和资金面因子的数量都较多，因此训练中容易造成模型的过拟合，为了解决这个问题，此处决定对因子进行降维，并将总因子个数控制在 50 左右。

表 3-2 因子筛选数量

因子类	原始数量	筛选后数量
技术因子	172	25
资金面因子	17	5
情绪因子	1	1
宏观因子	14	14
大类资产价格因子	3	3

为了选出对预测目标最具预测能力的因子池，本文采用了分步回归进行因子筛选。分步回归的流程如下：

1. 对于 n 维特征 X ，和预测目标 y ，计算各特征与 y 的相关系数
2. 选取和 y 相关系数绝对值最高的特征 X_{i_1}
3. 进行 $y \sim 1 + X_{i_1}$ 的回归，计算待选特征和残差之间的相关系数，选取相关系数绝对值最高的特征 X_{i_2} ，再将 X_{i_2} 也加入到回归模型中
4. 一直重复步骤 3，直至选取的特征数量满足需求

为了避免“透视未来”，采用训练集数据进行特征筛选，并采用筛选出的特征进行模型训练、验证和测试。

3.3 模型预测、策略构建和回测

在构造因子并进行因子降维后，我们需要对机器学习模型进行训练并将训练好的模型用于预测，此后根据预测结果构建交易策略，并对策略的历史表现进行回测。

流程如下：

- a. 将整个数据划分为训练集、验证集和测试集。

训练集	2019/6/13-2022/10/31
验证集	2022/11/01-2023/05/12
测试集	2023/05/15-2023/12/01

- b. 使用训练集的数据对机器学习模型进行训练。

X_{train}	自变量：筛选后因子的集合
y_{train}	因变量：国债期货合约未来收益（考虑未来 1 天、3 天、5 天的收益）

- c. 将验证集和测试集的自变量 X_{valid} 和 X_{test} 带入训练好的模型，得到未来收益的预测值 y_{valid} 和 y_{test} 。
- d. 预测未来收益大于 0 时，做多国债期货；预测未来收益小于 0 时，做空国债期货；预测未来收益等于 0 时，保持空仓。
- e. 若预测目标为未来一天的收益（即持有期为 1），则每日对整个组合进行仓位调整；若预测目标为未来 N 天的收益（即持有期为 N ），则每日对 $1/N$ 的组合进行仓位调整，并将这部分仓位维持 N 天。据此规则可得到组合每日的仓位。
- f. 根据组合的每日仓位，计算策略的每日损益，并计算出策略表现的各种评价

指标。(对组合在验证集和测试集上的表现分别计算评价指标)。

表 3-3 策略评价指标

指标	含义	计算方式
收益风险比	$\frac{E(r_p)}{\sigma_p}$	$\frac{\text{日度收益均值}}{\text{日度收益标准差}} \sqrt{250}$
收益风险比的稳定性	1周滚动收益风险比的均值/标准差	1周滚动窗口计算收益风险比，然后用均值/标准差
总收益	期间内总收益	日度收益累加
最大回撤	策略净值的最大回撤	策略净值的最大回撤
胜率	衡量策略取得盈利的概率	$\frac{\text{策略盈利天数}}{\text{总天数}}$
盈亏比	衡量策略盈利亏损幅度的大小	$\frac{\text{盈利日的日均盈利}}{\text{亏损日的日均亏损}}$
日均交易	衡量策略调仓的频率	$\frac{\text{每日仓位变化绝对值累加}}{\text{总天数}}$

- g. 使用不同的模型超参数，通过 a-f 进行策略构建。比较使用不同超参数的策略在验证集上的表现，挑选每个持有期上表现最好的模型超参数，作为最终的策略参数。下文采用策略在验证集上实现的收益风险比的稳定性对不同超参数组进行排序，并对每个持有期的策略选择一组最优超参数，最优超参数为在验证集内的收益风险比的稳定性最高的一组超参数。

3.4 本章小结

本章构建了技术因子、资金利率因子、新闻情绪因子、宏观因子和其它大类资产价格走势因子，这些因子较为全面地涵盖了影响国债期货收益的各类因素。在得到初始的因子池后，还提出了分步回归作为因子筛选的方法进行因子降维，避免模型过拟合。

本章还提出了模型训练、超参数选择、投资策略构建和回测的具体方法。

第四章 实证分析

4.1 相关性分析

以下是各因子与国债期货未来 1 日收益、未来 3 日收益和未来 5 日收益的相关性。下文中的 r1、r3、r5 分别代表未来 1 日、3 日、5 日收益。

表 4-1 因子与 T 收益相关系数

	r1	r3	r5	
技术因子	SLOPE_1	-9.81%	DMP_14	-10.10%
	CDL_3INSIDE	3.66%	PVO_12_26_9	10.19%
	CDL_INSIDE	2.75%	AR_26	2.03%
	CDL_LONGLINE	0.18%	CDL_CLOSINGMARUBOZU	-8.80%
	PVO_12_26_9	5.97%	CDL_HARAMICROSS	9.76%
	DMP_14	-7.45%	CDL_DRAGONFLYDOJI	-3.69%
	AR_26	-0.21%	ISB_26	-3.65%
	CDL_STICKSANDWICH	4.86%	ICS_26	4.43%
	CDL_PIERCING	-4.39%	QQEI_14_5_4.236	-2.23%
	PWMA_10	-2.17%	CG_10	-10.75%
	ICS_26	2.42%	CDL_HOMINGPIGEON	-0.54%
	MOM_10	-8.18%	QQEs_14_5_4.236	0.51%
	CDL_XSIDEGAP3METHODS	-1.26%	SLOPE_1	-7.18%
	CDL_ONNECK	2.70%	CDL_DOJISTAR	2.91%
	CDL_DOJISTAR	4.04%	SQZPRO_ON_NARROW	4.64%
	STOChd_14_3_3	-1.28%	SQZ_ON	-3.72%
	SQZPRO_ON_NARROW	3.14%	SMIo_5_20_5	-4.39%
	PSL_12	-7.91%	CDL_HARAMI	7.83%
资金面因子	CDL_SHORTLINE	3.40%	CDL_THRUSTING	-1.71%
	ER_10	2.44%	CDL_ADVANCEBLOCK	2.91%
	CDL_MORNINGDOJISTAR	1.65%	CDL_HIKKAKE	0.40%
	CDL_SPINNINGTOP	0.12%	CDL_MORNINGSTAR	1.03%
	CDL_SEPARATINGLINES	-2.44%	CDL_MORNINGDOJISTAR	-3.35%
	CDL_EVENINGSTAR	1.43%	CDL_ONNECK	-4.35%
	SQZ_20_2.0_20_1.5	-4.70%	CDL_MARUBOZU	-2.27%
	GC007:成交量	-2.95%	利率互换:FR007:5 年	10.57%
	利率互换:FR007:5 年	4.60%	GC001:加权平均	-5.01%
	R007:加权平均	-3.12%	GC007:成交量	-2.93%
	成交量:R007	2.85%	成交量:R007	2.68%
				GC007:成交量
				-2.21%

续表 4-1

	r1		r3		r5	
资金面因子	R001:加权平均	2.56%	GC014:成交量	2.89%	成交量:R007	0.79%
新闻情绪	scores	-5.01%	scores	-9.98%	scores	-7.76%
大类资产走势	HS300_1d	6.31%	HS300_1d	-6.44%	HS300_1d	-4.49%
	南华商品指数_1d	1.29%	南华商品指数_1d	-1.48%	南华商品指数_1d	0.54%
	人民币兑美元_1d	-2.69%	人民币兑美元_1d	-0.14%	人民币兑美元_1d	1.30%
宏观因子	GDP	0.06%	GDP	-2.65%	GDP	-1.44%
	PPI	-2.34%	PPI	-1.22%	PPI	1.32%
	CPI	-4.71%	CPI	-0.11%	CPI	0.29%
	工业增加值	1.91%	工业增加值	0.58%	工业增加值	5.58%
	固定资产投资	1.83%	固定资产投资	-3.24%	固定资产投资	-4.96%
	社会消费品零售	4.05%	社会消费品零售	1.92%	社会消费品零售	-2.34%
	进口	0.33%	进口	1.27%	进口	1.46%
	出口	2.68%	出口	0.95%	出口	4.80%
	贸易差额	-0.99%	贸易差额	-3.28%	贸易差额	-1.37%
	M2	1.86%	M2	2.59%	M2	2.48%
	人民币贷款新增	3.57%	人民币贷款新增	4.15%	人民币贷款新增	2.18%
	社融	0.46%	社融	1.18%	社融	-0.57%
	PMI	1.18%	PMI	-5.89%	PMI	-7.07%
	失业率	-0.75%	失业率	0.44%	失业率	0.44%

上方列出了 10 年期国债期货未来收益和各因子的相关系数情况，技术因子整体和 r1、r3 和 r5 都有较强相关性。

新闻情绪因子和 r1、r3、r5 均有较强相关性，尤其和 r3 的相关性达到了-10%左右。新闻情绪因子和 T 未来收益的相关性为负，说明新闻情绪较差时对国债类资产的追捧度会更高，符合逻辑。

沪深 300 过去一天的收益率与 r1 呈正相关性，与 r3 和 r5 呈负相关性，说明股债“跷跷板”效应完全反映出来需要一定的时间。南华商品指数和美元汇率与国债期货收益的相关性较低。

宏观经济指标中，CPI 同比、固定资产投资同比、社会消费品零售同比、人民币贷款新增、PMI 与 T 的未来收益相关性较大。

表 4-2 因子与 TF 收益相关系数

	r1		r3		r5	
技术因子	CFO_9	-8.34%	SQZPRO_ON_WIDE	-7.56%	SQZPRO_ON_WIDE	-6.89%
	CDL_INSIDE	6.66%	SLOPE_1	-10.09%	DMP_14	-9.22%
	CDL_ADVANCEBLOCK	-5.53%	DMN_14	-6.85%	PPO_12_26_9	7.72%
	CDL_DARKCLOUDCOVER	-7.17%	PVOH_12_26_9	7.94%	QQEI_14_5_4.236	-3.82%
	DMP_14	-5.53%	QQEI_14_5_4.236	-1.56%	PVO_12_26_9	6.63%
	CDL_BELTHOLD	2.04%	CDL_ONNECK	-7.86%	RVGIs_14_4	1.36%
	PVO_12_26_9	3.50%	PVOs_12_26_9	3.47%	HILOS_13_21	-8.66%
	PPO_12_26_9	2.55%	KAMA_10_2_30	-6.46%	ICS_26	4.62%
	CDL_ONNECK	-5.27%	ICS_26	3.59%	J_9_3	-1.59%
	CDL_HAMMER	-2.73%	CDL_ADVANCEBLOCK	-5.98%	CDL_STALLEDPATTERN	10.83%
	CDL_CLOSINGMARUBOZU	-7.24%	J_9_3	-3.64%	CDL_CLOSINGMARUBOZU	-6.65%
	CDL_SEPARATINGLINES	-1.34%	CDL_HARAMICROSS	3.95%	DMN_14	-10.08%
	CDL_MORNINGSTAR	4.29%	SQZPRO_ON_NARROW	3.20%	CDL_HARAMI	4.93%
	ICS_26	1.91%	CDL_INVERTEDHAMMER	5.93%	CDL_SHORTLINE	-2.14%
	HILOS_13_21	-3.51%	CDL_HAMMER	-0.02%	CDL_DARKCLOUDCOVER	-5.66%
	CG_10	-2.62%	CDL_LONGLINE	-3.44%	AR_26	5.07%
	CDL_SHOOTINGSTAR	5.12%	CDL_DARKCLOUDCOVER	-4.56%	SUPERTd_7_3.0	1.39%
	CDL_LONGLINE	0.49%	CDL_UNIQUE3RIVER	-5.35%	CDL_SEPARATINGLINES	-3.09%
	QQEI_14_5_4.236	0.96%	CDL_CLOSINGMARUBOZU	-8.49%	CDL_EVENINGSTAR	0.47%
	CDL_TRISTAR	-6.50%	CDL_MATCHINGLOW	-4.30%	CDL_ENGULFING	-6.94%
	SQZPRO_ON_WIDE	-2.09%	CDL_DOJISTAR	6.33%	CDL_LONGLINE	-1.59%
	CDL_3OUTSIDE	-1.14%	CDL_3LINESTRIKE	-3.42%	SQZPRO_ON_NARROW	0.69%
	CDL_DOJISTAR	5.01%	CDL_3OUTSIDE	-2.75%	CDL_SHOOTINGSTAR	2.44%
	CDL_THRUSTING	3.08%	BR_26	4.83%	BOP	-5.32%
	CDL_XSIDEGAP3METHODS	2.14%	CDL_THRUSTING	-1.78%	CFO_9	-0.91%
资金面因子	GC007:成交量	-2.98%	利率互换:FR007:5 年	9.27%	利率互换:FR007:5 年	12.19%
	利率互换:FR007:5 年	3.83%	GC001:加权平均	-5.39%	GC001:加权平均	-8.80%
	R007:加权平均	-2.02%	GC007:成交量	-3.13%	成交量:R001	2.67%
	成交量:R007	2.49%	成交量:R007	0.69%	GC007:成交量	-4.14%
	R001:加权平均	2.50%	GC014:成交量	3.37%	GC014:成交量	2.42%
新闻情绪	scores	-5.77%	scores	-9.98%	scores	-8.00%
大类资产走势	HS300_1d	3.92%	HS300_1d	-7.35%	HS300_1d	-4.65%
	南华商品指数_1d	0.05%	南华商品指数_1d	-0.22%	南华商品指数_1d	0.35%
	人民币兑美元_1d	-0.87%	人民币兑美元_1d	1.48%	人民币兑美元_1d	2.26%
宏观因子	GDP	0.27%	GDP	-2.76%	GDP	-0.59%
	PPI	-0.90%	PPI	1.26%	PPI	2.13%
	CPI	-1.37%	CPI	3.92%	CPI	1.10%
	工业增加值	2.72%	工业增加值	1.19%	工业增加值	4.74%
	固定资产投资	3.41%	固定资产投资	0.25%	固定资产投资	-2.02%
	社会消费品零售	3.15%	社会消费品零售	2.21%	社会消费品零售	0.38%

续表 4-2

	r1		r3		r5	
宏观因子	进口	-0.99%	进口	-0.36%	进口	0.72%
	出口	4.62%	出口	2.17%	出口	5.72%
	贸易差额	1.62%	贸易差额	-0.67%	贸易差额	0.64%
	M2	-0.79%	M2	2.57%	M2	3.28%
	人民币贷款新增	4.30%	人民币贷款新增	6.18%	人民币贷款新增	3.75%
	社融	-0.11%	社融	1.15%	社融	0.84%
	PMI	3.40%	PMI	-6.51%	PMI	-7.48%
	失业率	-1.15%	失业率	-0.74%	失业率	-1.29%

上方列出了 5 年期国债期货未来收益和各因子的相关系数情况，技术因子整体和 r1、r3 和 r5 都有较强相关性。

新闻情绪因子和 r1、r3、r5 具有较强相关性，尤其和 r3 的相关性达到了-10%左右。新闻情绪因子和 TF 未来收益的相关性为负，说明新闻情绪较差时对国债类资产的追捧度会更高，符合逻辑。

沪深 300 过去一天的收益率与 r1 呈正相关性，与 r3 和 r5 呈负相关性，说明股债“跷跷板”效应完全反映出来需要一定的时间。南华商品指数和美元汇率与国债期货收益的相关性较低。

宏观经济指标中，工业增加值同比、出口同比、人民币贷款新增、PMI 与 TF 的未来收益相关性较大。

表 4-3 因子与 TS 收益相关系数

	r1		r3		r5	
技术因子	CDL_XSIDEGAP3METHODS	-6.14%	SUPERTs_7_3.0	-10.58%	AR_26	12.38%
	CDL_INVERTEDHAMMER	7.44%	ICS_26	3.31%	ICS_26	4.12%
	CDL_HARAMI	7.05%	DMP_14	-3.13%	SUPERT_7_3.0	-12.64%
	CDL_DARKCLOUDCOVER	6.66%	BOP	-9.26%	RSI_14	5.65%
	SUPERTs_7_3.0	-6.17%	KSTs_9	5.47%	SQZPRO_ON_WIDE	-8.98%
	ICS_26	1.92%	SQZPRO_ON_NARROW	-8.13%	CDL_XSIDEGAP3METHODS	-8.77%
	DMP_14	-3.59%	STOCHRSId_14_14_3_3	-4.84%	CDL_HOMINGPIGEON	4.63%
	DMN_14	-2.41%	SQZPRO_OFF	8.19%	BOP	-6.24%
	SQZPRO_ON_WIDE	-5.27%	CDL_PIERCING	-7.60%	CDL_RICKSHAWMAN	-5.26%
	CDL_STICKSANDWICH	5.20%	CFO_9	-9.41%	CDL_DOJISTAR	-1.28%
	PPOh_12_26_9	1.50%	AR_26	8.13%	PVOh_12_26_9	-3.18%
	INERTIA_20_14	-3.57%	CDL_HOMINGPIGEON	4.67%	PVOs_12_26_9	-1.02%
	CDL_SPINNINGTOP	6.63%	CDL_RICKSHAWMAN	-6.27%	DMP_14	-3.84%

续表 4-3

	r1	r3	r5	
技术因子	CDL_MATCHINGLOW	3.41%	CDL_3OUTSIDE	-4.74%
	CDL_3OUTSIDE	-4.42%	SUPERTI_7_3.0	-9.84%
	CDL_CLOSINGMARUBOZU	-4.57%	QQE_14_5_4.236	6.73%
	CDL_HOMINGPIGEON	3.98%	PSL_12	-3.19%
	CDL_STALLEDPATTERN	4.38%	INERTIA_20_14	-3.67%
	CDL_INSIDE	4.73%	PPOh_12_26_9	4.98%
	CDL_TASUKIGAP	3.61%	CDL_MATCHINGLOW	2.35%
	CDL_LADDERBOTTOM	-2.88%	ER_10	4.33%
	AR_26	3.82%	CDL_GRAVESTONEDOJI	-1.93%
	CDL_SHOOTINGSTAR	3.50%	CDL_INVERTEDHAMMER	3.78%
资金面因子	CDL_HARAMICROSS	7.17%	CDL_3INSIDE	-5.53%
	PVOs_12_26_9	-0.30%	CDL_2CROWS	-2.54%
	利率互换:FR007:5 年	6.88%	利率互换:FR007:5 年	12.34%
	GC014:收盘价	-0.07%	GC001:加权平均	-1.57%
	成交量:R014	0.88%	GC014:成交量	6.04%
新闻情绪	成交量:R007	2.66%	GC007:成交量	-1.30%
	GC001:成交量	-3.71%	成交量:R007	1.18%
	scores	-7.04%	scores	-10.19%
大类资产走势	HS300_1d	4.99%	HS300_1d	-7.27%
	南华商品指数_1d	2.40%	南华商品指数_1d	-0.10%
	人民币兑美元_1d	1.01%	人民币兑美元_1d	5.64%
宏观因子	GDP	-2.78%	GDP	-4.40%
	PPI	-0.79%	PPI	0.81%
	CPI	-2.23%	CPI	1.40%
	工业增加值	4.83%	工业增加值	0.84%
	固定资产投资	5.37%	固定资产投资	0.62%
	社会消费品零售	2.96%	社会消费品零售	2.13%
	进口	-2.90%	进口	-1.12%
	出口	3.31%	出口	2.42%
	贸易差额	1.46%	贸易差额	-0.48%
	M2	1.20%	M2	4.19%
	人民币贷款新增	4.12%	人民币贷款新增	6.26%
	社融	0.68%	社融	1.79%
	PMI	1.89%	PMI	-4.73%
	失业率	0.00%	失业率	0.31%
			失业率	-0.46%

上方列出了 2 年期国债期货未来收益和各因子的相关系数情况，技术因子整体和 r1、r3 和 r5 都有较强相关性。

新闻情绪因子和 r1、r3、r5 具有较强相关性，尤其和 r3 的相关性达到了 -10%。

新闻情绪因子和 TS 未来收益的相关性为负，说明新闻情绪较差时对国债类资产的追捧度会更高，符合逻辑。

沪深 300 过去一天的收益率与 r1 呈正相关性，与 r3 和 r5 呈负相关性，说明股债“跷跷板”效应完全反映出来需要一定的时间。南华商品指数与国债期货收益的相关性较低。与 TF 和 T 不同的是，美元汇率对 TS 的 r3 和 r5 的相关性较高，说明外汇的升值会导致资金更多流向短期国债。

宏观经济指标中，GDP 同比、工业增加值同比、固定资产投资同比、出口同比、人民币贷款新增、PMI 与 TS 的未来收益相关性较大。

综上，结合之前的各类因子与国债期货合约的未来收益的相关性来看。技术因子与国债期货未来收益之间具有较强的相关性，说明市场未达到弱式有效市场。新闻情绪因子总体来说与国债期货相关性较强，说明新闻情绪是影响国债期货价格的重要因素。其它大类资产价格中，沪深 300 对国债期货的影响较大，南华商品指数影响较小。宏观因子中，GDP 同比、出口同比、人民币贷款新增、PMI 与国债期货未来收益的相关性也较强。

4.2 策略回测

在展示各策略的回测结果前，现在先给出市场基准策略，此处将市场基准策略定义为每日持续全仓做多的策略。市场基准策略在验证集和测试集上的表现如下：

表 4-4 市场基准在验证集和测试集上的表现

品种	valid					test				
	收益 风险 比	总收 益	最大 回撤	胜率	盈亏 比	收益 风险 比	总收 益	最大 回撤	胜率	盈亏 比
TS	0.90	0.40	0.54	0.54	0.99	-0.56	-0.24	0.70	0.50	0.90
TF	0.81	0.85	1.26	0.56	0.93	0.41	0.43	1.19	0.53	0.95
T	1.51	2.24	1.30	0.60	0.89	0.90	1.22	1.43	0.55	0.96

4.2.1 线性模型

首先，我们采用线性回归模型作为基准模型来构建策略，后续可以比较机器学习模型策略和线性回归策略的差异。

策略参数：

参数	含义	测试范围
hold_period	持有期	1, 3, 5

回测结果：

表 4-5 线性回归模型 TS 策略表现

	策略表现											
	valid						test					
持有期	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易
1	0.57	0.25	0.53	0.57	0.84	0.55	2.29	0.95	0.37	0.59	1.01	0.69
3	2.52	0.72	0.17	0.54	1.35	0.31	2.46	0.86	0.27	0.52	1.41	0.19
5	1.39	0.33	0.20	0.53	1.14	0.22	2.01	0.68	0.26	0.50	1.40	0.13

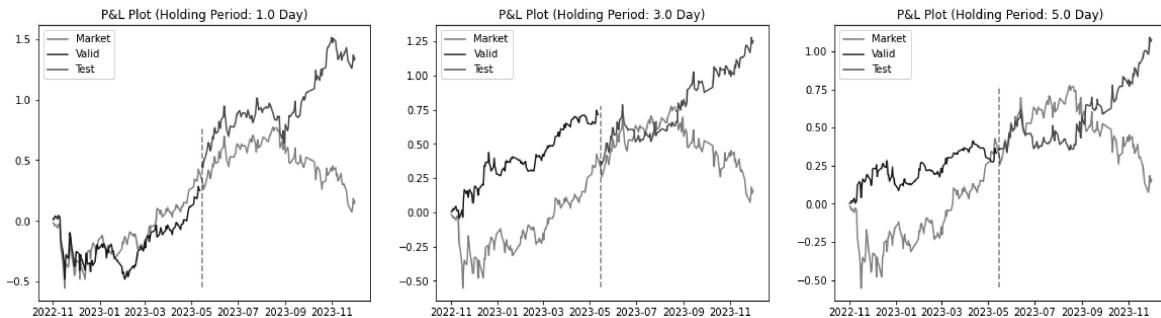


图 4-1 线性回归模型 TS 策略 P&L

基于线性回归模型的策略在 TS 上总体表现出色。当持有期为 3 天和 5 天时，策略表现较好，在测试集和验证集上的收益风险比均大幅优于市场基准。

表 4-6 线性回归模型 TF 策略表现

	策略表现											
	valid						test					
持有期	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易
1	3.19	3.29	0.86	0.60	1.22	0.72	1.55	1.63	0.81	0.52	1.19	0.72
3	1.94	1.48	0.87	0.58	1.07	0.26	1.12	0.90	0.79	0.46	1.44	0.24
5	1.86	1.46	0.87	0.56	1.11	0.11	-1.03	-0.74	1.32	0.43	1.12	0.14

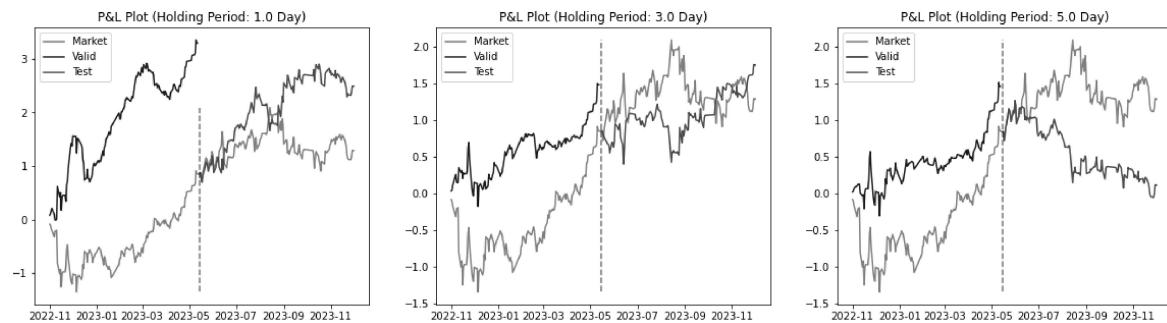


图 4-2 线性回归模型 TF 策略 P&L

总体来看,线性回归模型对 TF 合约的预测效果不错。当持有期为 1 天和 3 天时,策略在验证集和测试集上的收益风险比均略大于市场基准。当持有期为 5 天时,策略在测试集上的收益风险比远小于市场基准。

表 4-7 线性回归模型 T 策略表现

持有期	策略表现											
	valid			test								
收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	
1	1.40	2.08	1.84	0.59	0.91	0.73	0.76	1.03	0.91	0.52	1.05	0.74
3	1.96	2.02	1.03	0.57	1.10	0.24	1.41	1.40	1.20	0.52	1.21	0.26
5	2.38	2.83	1.03	0.61	1.00	0.08	0.28	0.25	0.56	0.47	1.17	0.16

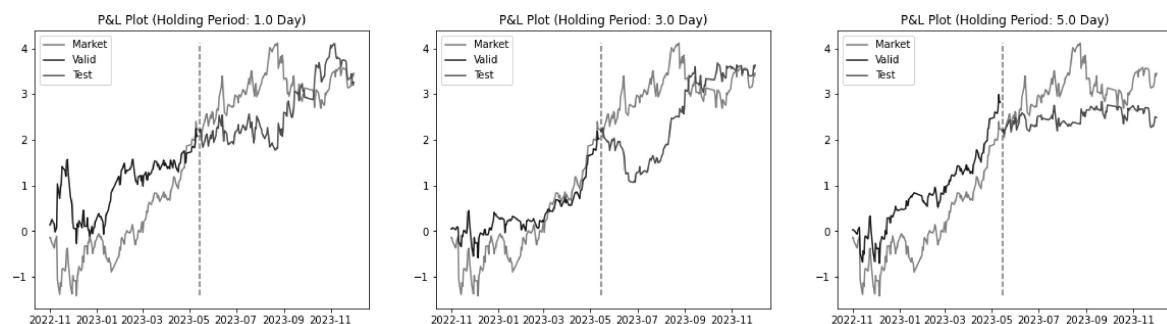


图 4-3 线性回归模型 T 策略 P&L

总体来看,线性模型对 T 合约的预测效果一般。仅持有期为 3 天时策略在验证集和测试集上的收益风险比略大于市场基准。在其它持有期下,策略在验证集和测试集上的收益风险比均未能同时高于市场基准。

综上所述,线性模型对 TS 合约的预测效果较为出色,对其它国债期货合约的预测效果一般。

4.2.2 Lasso 模型

Lasso 回归在线性回归的基础上加入了正则项，理论上能减少线性回归模型的过拟合，提高模型样本外表现。

策略参数：

参数	含义	测试范围
hold_period	持有期	1, 3, 5
alpha	正则项系数	0.00001, 0.0001, 0.001

回测结果：

表 4-8 Lasso 回归模型 TS 策略表现

	策略表现											
	valid						test					
持有期	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易
1	0.57	0.25	0.53	0.57	0.84	0.55	2.29	0.95	0.37	0.59	1.01	0.69
3	2.60	0.74	0.17	0.55	1.32	0.31	2.46	0.86	0.27	0.52	1.41	0.19
5	1.44	0.35	0.20	0.54	1.11	0.20	2.16	0.74	0.22	0.52	1.35	0.12

持有期	模型参数											
	alpha											
1	0.00001											
3	0.0001											
5	0.0001											

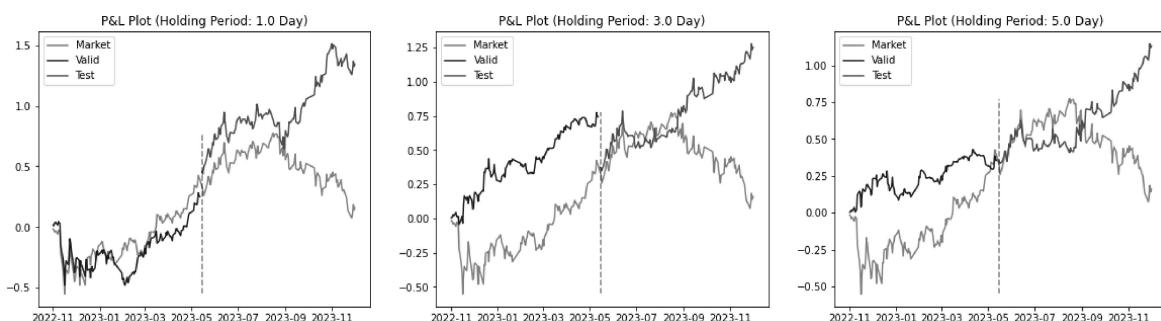


图 4-4 Lasso 回归模型 TS 策略表现

Lasso 模型在 TS 合约的预测上表现总体不错，与线性回归模型类似。当持有期为 3 天和 5 天时，策略在验证集和测试集上的收益风险比均优于市场基准。

表 4-9 Lasso 回归模型 TF 策略表现

策略表现												
	valid						test					
持有期	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易
1	3.17	3.27	0.71	0.58	1.30	0.75	1.10	1.17	0.81	0.51	1.13	0.72
3	2.11	1.62	0.87	0.59	1.07	0.25	0.90	0.68	0.81	0.47	1.33	0.25
5	1.86	1.46	0.87	0.56	1.11	0.11	-1.03	-0.74	1.32	0.43	1.12	0.14

模型参数												
持有期	alpha											
1	0.001											
3	0.001											
5	0.00001											

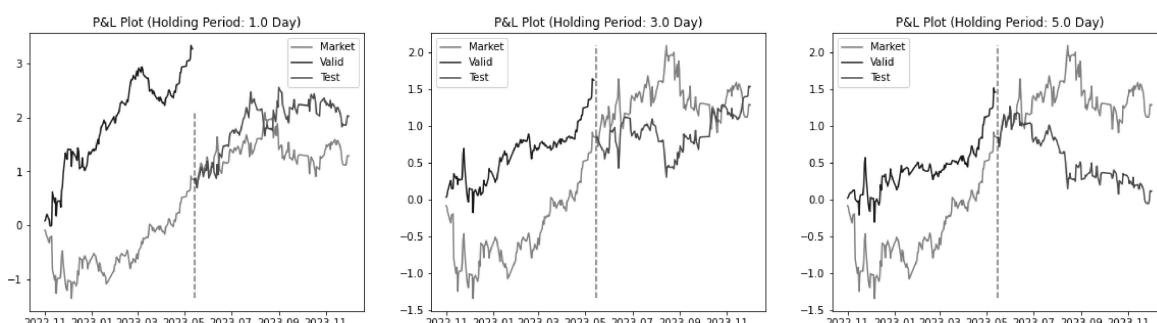


图 4-5 Lasso 回归模型 TF 策略 P&L

当持有期为 1 天和 3 天时，策略在验证集和测试集上的收益风险比均优于市场基准。当持有期为 5 天时，策略在测试集上弱于市场基准。

表 4-10 Lasso 回归模型 T 策略表现

策略表现												
	valid						test					
持有期	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易
1	1.50	2.22	1.84	0.60	0.89	0.77	0.76	1.03	0.91	0.52	1.05	0.74
3	1.96	2.02	1.03	0.57	1.10	0.24	1.41	1.40	1.20	0.52	1.21	0.26
5	2.38	2.83	1.03	0.61	1.00	0.08	0.28	0.25	0.56	0.47	1.17	0.16

模型参数	
持有期	alpha
1	0.0001
3	0.00001
5	0.00001

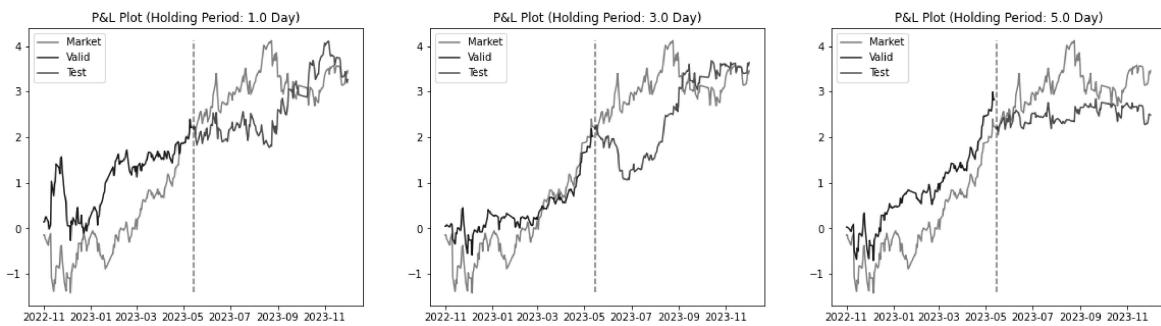


图 4-6 Lasso 回归模型 T 策略 P&L

将 Lasso 回归用于 T 合约的预测，当持有期为 3 天时，策略在验证集和测试集的收益风险比均高于市场基准；当持有期为 3 天和 5 天时，策略在测试集上弱于市场基准。

总体来看，基于 Lasso 回归构建的策略效果和线性模型策略基本一致。

4.2.3 随机森林

随机森林模型使用了样本重采样和特征随机选择等技巧，在小数据集中的表现出色。

策略参数：

参数	含义	测试范围
hold_period	持有期	1, 3, 5
n_estimators	森林中决策树的数量	3000
max_depth	决策树的最大深度	2, 3, 5, 9
max_features	决策树每次分裂时考虑的最大特征数	sqrt
max_samples	训练每颗决策树的样本数比例	0.05, 0.2
oob_score	是否采用袋外数据检测模型	True

回测结果：

表 4-11 随机森林模型 TS 策略表现

策略表现												
	valid						test					
持有期	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易
1	1.33	0.58	0.54	0.56	0.99	0.06	-0.29	-0.12	0.57	0.50	0.94	0.06
3	0.91	0.40	0.53	0.54	0.99	0.02	-0.27	-0.11	0.57	0.50	0.94	0.01
5	0.95	0.42	0.53	0.54	1.00	0.01	-0.56	-0.24	0.70	0.50	0.90	0.00

模型参数												
持有期	n_estimators	max_depth	max_features	max_samples	oob_score							
1	3000	2	sqrt	0.05	TRUE							
3	3000	2	sqrt	0.05	TRUE							
5	3000	2	sqrt	0.05	TRUE							

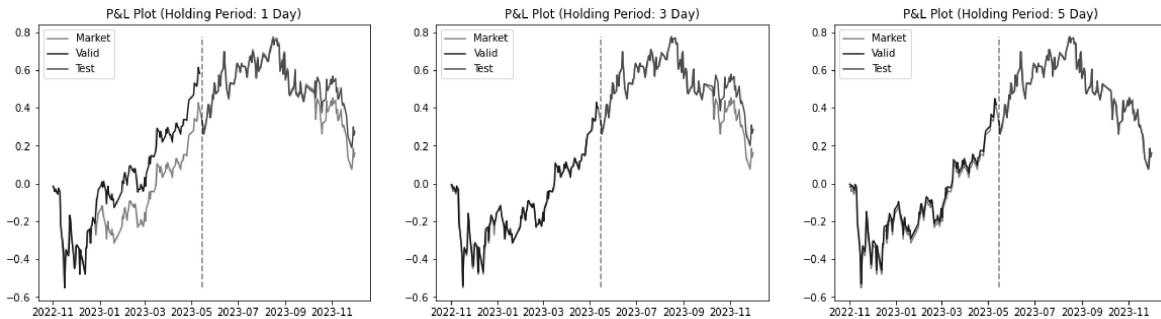


图 4-7 随机森林模型 TS 策略 P&L

随机森林对预测 TS 合约总体表现较好，当持有期为 1、3 或 5 天时，策略在验证集和测试集上的表现均优于市场基准。

表 4-12 随机森林模型 TF 策略表现

策略表现												
	valid						test					
持有期	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易
1	1.76	1.84	1.17	0.58	1.00	0.19	1.03	1.09	0.87	0.56	0.94	0.09
3	1.82	1.46	0.73	0.55	1.15	0.17	2.32	1.72	0.39	0.56	1.18	0.28
5	0.93	0.97	1.21	0.56	0.95	0.01	0.37	0.39	1.19	0.53	0.95	0.01

模型参数					
持有期	n_estimators	max_depth	max_features	max_samples	oob_score
1	3000	5	sqrt	0.05	TRUE
3	3000	9	sqrt	0.2	TRUE
5	3000	2	sqrt	0.05	TRUE

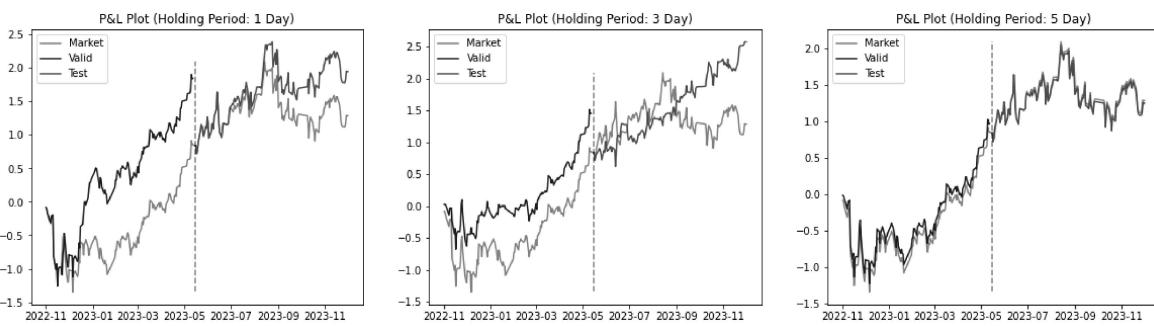


图 4-8 随机森林模型 TF 策略 P&L

随机森林在预测 TF 合约时表现整体较好，除了持有期为 5 天的策略在测试集上收益风险比弱于基准外，其它持有期下策略收益风险比均强于市场基准。

表 4-13 随机森林模型 T 策略表现

策略表现											
持有期	valid						test				
	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比
1	2.18	3.20	1.30	0.61	0.94	0.09	1.15	1.56	1.28	0.56	0.97
3	1.81	2.63	1.30	0.60	0.95	0.05	1.93	2.36	1.16	0.56	1.10
5	1.62	2.39	1.30	0.60	0.91	0.01	0.88	1.19	1.43	0.55	0.96

模型参数					
持有期	n_estimators	max_depth	max_features	max_samples	oob_score
1	3000	2	sqrt	0.2	TRUE
3	3000	5	sqrt	0.05	TRUE
5	3000	2	sqrt	0.05	TRUE

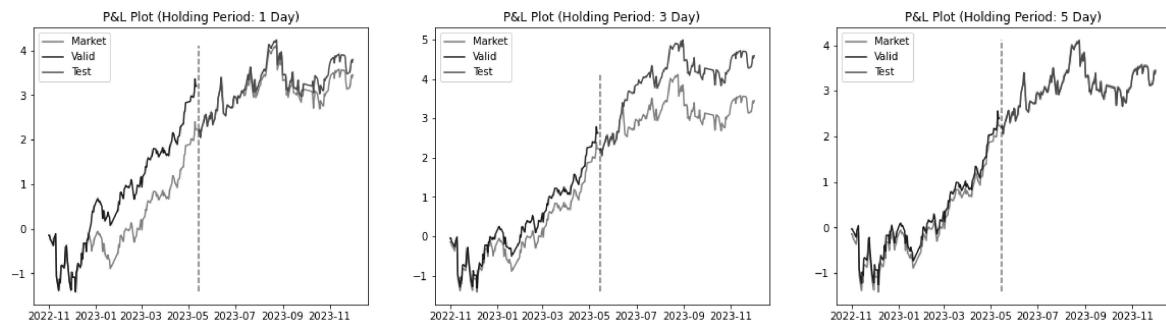


图 4-9 随机森林模型 T 策略 P&L

使用随机森林对 T 合约收益进行预测，当持有期为 1 天和 3 天时，策略在验证集和测试集上收益风险比均优于市场基准；当持有期为 5 天时，收益风险比与市场基准基本一致。

4.2.4 多层感知机

策略参数：

参数	含义	测试范围
hold_period	持有期	1, 3, 5
max_iter	最大迭代次数	1000
hidden_layer_sizes	隐层结构	[15, 3], [40], [50], [100]
alpha	L2 正则项	0.0001, 0.001, 0.01, 0.1
learning_rate_init	初始学习率	0.0001, 0.001, 0.01, 0.1
batch_size	每次迭代使用的样本量	5, 10, 100, 200

回测结果：

表 4-14 多层感知机模型 TS 策略表现

策略表现												
持有期	valid						test					
	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易
1	3.18	1.37	0.54	0.63	1.04	0.42	0.12	0.05	0.67	0.53	0.92	0.75
3	4.18	1.22	0.14	0.59	1.66	0.31	-1.13	-0.31	0.39	0.49	0.86	0.25
5	2.28	0.77	0.31	0.62	0.97	0.10	-0.70	-0.15	0.35	0.54	0.75	0.20
模型参数												
持有期	max_iter		hidden_layer_sizes		alpha		learning_rate_in_it		batch_size			
1	1000		[40]		0.0001		0.01		5			
3	1000		[100]		0.01		0.1		200			
5	1000		[100]		0.001		0.01		100			

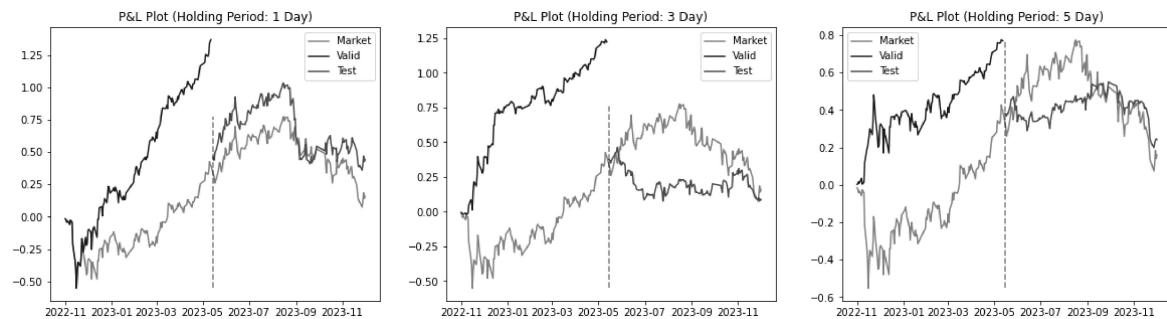


图 4-10 多层感知机模型 TS 策略 P&L

多层感知机预测 TS，当持有期为 1 天时，策略在验证集和测试集上的收益风险比均高于市场基准；当持有期为 3 天或 5 天时，策略在测试集上的表现不及市场基准。

表 4-15 多层感知机模型 TF 策略表现

策略表现												
持有期	valid						test					
	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易
1	3.26	3.35	1.17	0.64	1.04	0.47	-0.55	-0.59	1.93	0.49	0.97	0.53
3	2.93	2.49	0.87	0.63	1.07	0.21	1.10	0.78	0.80	0.56	0.96	0.28
5	1.94	1.50	0.87	0.59	1.08	0.15	-0.47	-0.33	1.05	0.46	1.09	0.16

模型参数											
持有期	max_iter		hidden_layer_sizes		alpha		learning_rate_in_it		batch_size		
	1		[15, 3]		0.1		0.01		200		
3	1000		[50]		0.1		0.001		10		
5	1000		[40]		0.1		0.01		200		

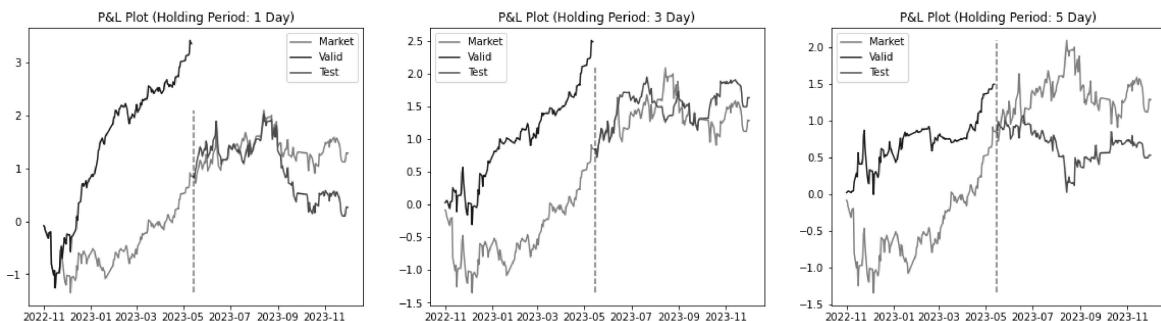


图 4-11 多层感知机模型 TF 策略 P&L

多层感知机预测 TF，当持有期为 3 天时，策略在验证集和测试集上的收益风险比均高于市场基准；当持有期为 1 天或 5 天时，策略在测试集上的表现不及市场基准。

表 4-16 多层感知机模型 T 策略表现

策略表现												
持有期	valid						test					
	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易
1	3.25	4.73	1.02	0.62	1.11	0.94	-0.55	-0.75	1.58	0.50	0.93	0.72
3	3.78	4.23	1.03	0.64	1.12	0.19	3.41	3.52	0.45	0.59	1.28	0.23
5	3.58	3.36	1.02	0.63	1.17	0.14	1.00	0.82	0.56	0.51	1.15	0.13

模型参数												
持有期	max_iter		hidden_layer_sizes		alpha		learning_rate_in_it		batch_size			
	1000		[40]		0.01		0.001		5			
3	1000		[100]		0.001		0.01		10			
5	1000		[15, 3]		0.1		0.01		200			

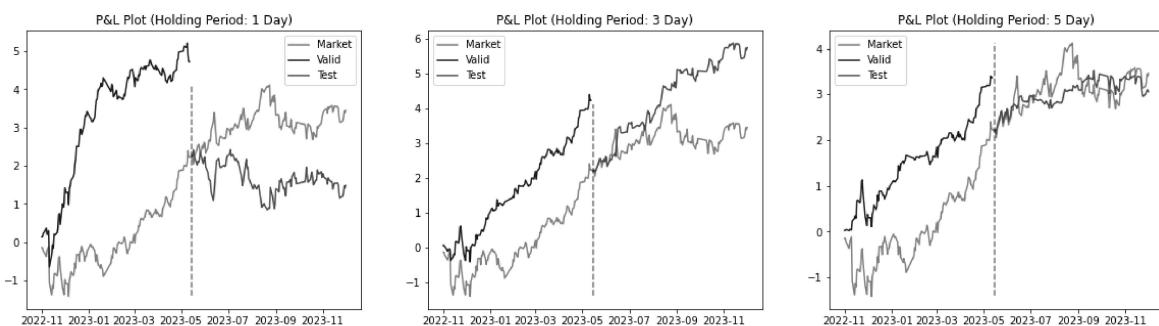


图 4-12 多层感知机模型 T 策略 P&L

多层感知机对 T 合约的预测效果相对较好。当持有期为 3 天或 5 天时，策略在验证集和测试集上的收益风险比均高于市场基准。

综上所述，神经网络模型对 T 合约具有较好的预测效果，但对 TS 和 TF 合约的预测效果一般。

4.2.5 CatBoost

策略参数：

参数	含义	测试范围
hold_period	持有期	1, 3, 5
iterations	最大树的数量	200, 500, 1000
depth	树深	3, 6, 10

回测结果：

表 4-17 CatBoost 模型 TS 策略表现

	策略表现											
	valid						test					
持有期	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易
1	0.16	0.07	0.67	0.48	1.10	0.72	-0.58	-0.24	0.75	0.51	0.87	0.81
3	1.18	0.50	0.53	0.56	0.96	0.04	-0.04	-0.01	0.53	0.53	0.87	0.22
5	0.34	0.14	0.53	0.57	0.80	0.11	1.54	0.46	0.39	0.56	1.01	0.17

持有期	模型参数											
	iterations						depth					
1	500						6					
3	500						6					
5	1000						6					

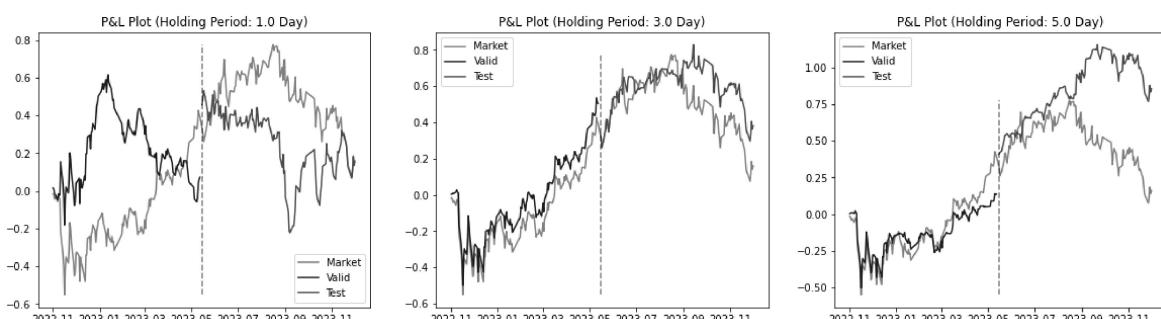


图 4-13 CatBoost 模型 TS 策略 P&L

将 CatBoost 模型用于 TS 合约的预测效果不佳。仅当持有期为 3 时，策略在验证集和测试集上的收益风险比优于市场基准。

表 4-18 CatBoost 模型 TF 策略表现

	策略表现												
	valid							test					
持有期	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	
1	2.24	2.33	0.65	0.56	1.21	0.50	1.30	1.37	1.36	0.52	1.14	0.74	
3	2.13	1.54	0.73	0.60	1.03	0.23	1.32	0.91	0.52	0.49	1.32	0.31	
5	1.47	1.23	0.87	0.57	1.00	0.10	1.04	0.75	0.70	0.51	1.16	0.18	

持有期	模型参数												
	iterations							depth					
1	500							6					
3	200							10					
5	1000							10					

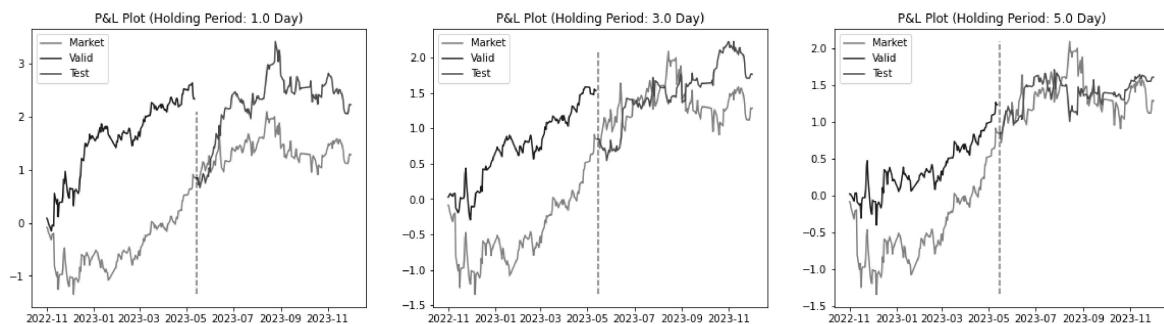


图 4-14 CatBoost 模型 TF 策略 P&L

基于 CatBoost 的预测构建的 TF 策略效果出色，持有期为 1、3 或 5 天时，策略在验证集和测试集上的收益风险比均强于市场基准。

表 4-19 CatBoost 模型 T 策略表现

	策略表现												
	valid							test					
持有期	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	
1	1.39	2.05	1.56	0.60	0.87	0.52	0.48	0.64	1.78	0.59	0.77	0.68	
3	1.58	2.20	1.30	0.61	0.86	0.09	1.91	1.99	0.70	0.57	1.05	0.17	
5	1.61	2.29	1.30	0.60	0.92	0.04	0.68	0.61	0.94	0.53	1.02	0.16	

模型参数		
持有期	iterations	depth
1	200	6
3	500	6
5	500	6

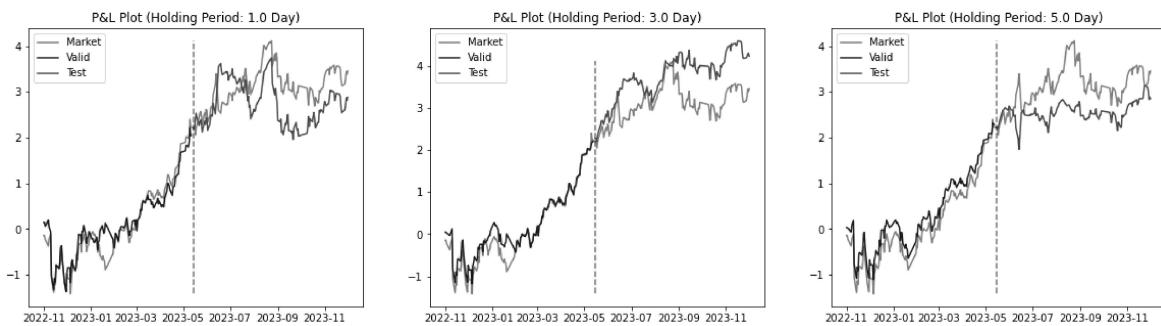


图 4-15 CatBoost 模型 T 策略 P&L

CatBoost 模型对 T 合约预测效果一般，当持有期 3 天时，策略在验证集和测试集上的收益风险比强于市场基准。当持有期为 1 天或 5 天时，策略在测试集上弱于市场基准。

综上所述，Catboost 模型对 TF 合约预测效果较好。

4.2.6 XGBoost

策略参数：

参数	含义	测试范围
hold_period	持有期	1, 3, 5
eta	学习率	0.01, 0.015, 0.025, 0.05, 0.1
max_depth	最大树深	3, 5, 6, 7, 9

回测结果：

表 4-20 XGBoost 模型 TS 策略表现

策略表现												
	valid							test				
持有期	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易
1	1.71	0.75	0.34	0.58	0.99	0.56	1.54	0.64	0.38	0.58	0.93	0.71
3	0.93	0.38	0.53	0.57	0.91	0.15	1.78	0.55	0.32	0.58	1.00	0.22
5	1.48	0.57	0.40	0.56	1.04	0.08	0.91	0.27	0.43	0.56	0.93	0.16

模型参数												
持有期	eta						max_depth					
1	0.1						9					
3	0.1						2					
5	0.1						2					

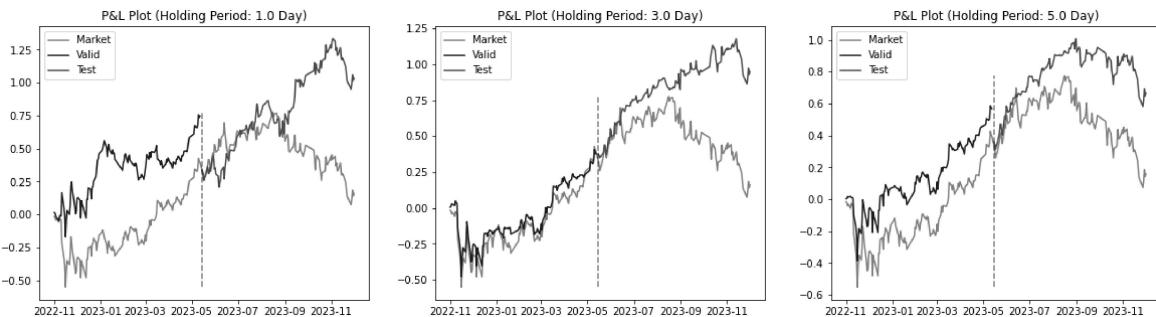


图 4-16 XGBoost 模型 TS 策略 P&L

使用 XGBoost 进行 TS 合约的预测，当持有期为 1 天、3 天或 5 天时，策略在测试集和验证集上的收益风险比均高于市场基准。

表 4-21 XGBoost 模型 TF 策略表现

策略表现												
	valid							test				
持有期	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易
1	1.86	1.94	1.41	0.60	0.95	0.33	0.58	0.62	1.04	0.51	1.07	0.44
3	1.28	1.21	1.21	0.58	0.94	0.15	1.61	1.47	1.08	0.58	0.96	0.16
5	1.75	1.33	0.87	0.60	0.92	0.14	1.22	1.01	0.68	0.53	1.11	0.13

模型参数		
持有期	eta	max_depth
1	0.05	9
3	0.05	9
5	0.025	9

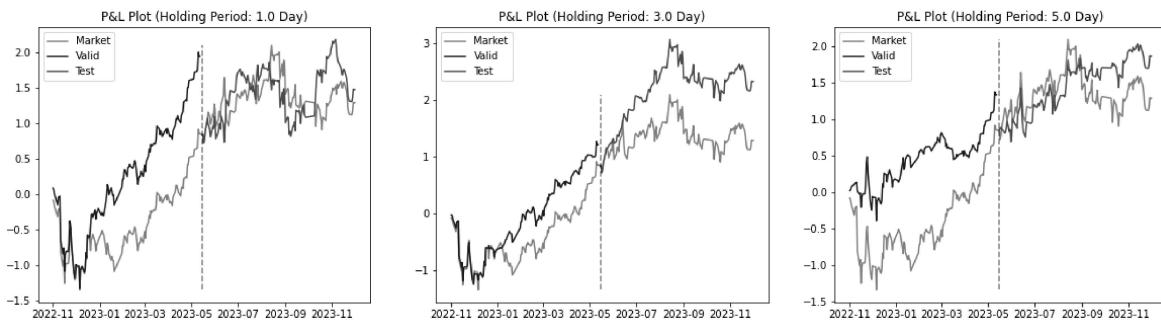


图 4-17 XGboost 模型 TF 策略 P&L

使用 XGBoost 进行 TF 合约的预测，当持有期为 1 天、3 天或 5 天时，策略在测试集和验证集上的收益风险比均高于市场基准。

表 4-22 XGBoost 模型 T 策略表现

策略表现												
持有期	valid						test					
	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易	收益风险比	总收益	最大回撤	胜率	盈亏比	日均交易
1	1.94	2.85	1.30	0.61	0.90	0.16	1.43	1.93	1.17	0.58	0.93	0.19
3	1.70	2.32	1.30	0.60	0.92	0.12	1.92	2.23	0.70	0.52	1.32	0.23
5	1.53	2.16	1.30	0.60	0.88	0.04	0.97	0.93	1.07	0.54	1.01	0.15

模型参数												
持有期	eta						max_depth					
1	0.025						9					
3	0.05						5					
5	0.1						5					

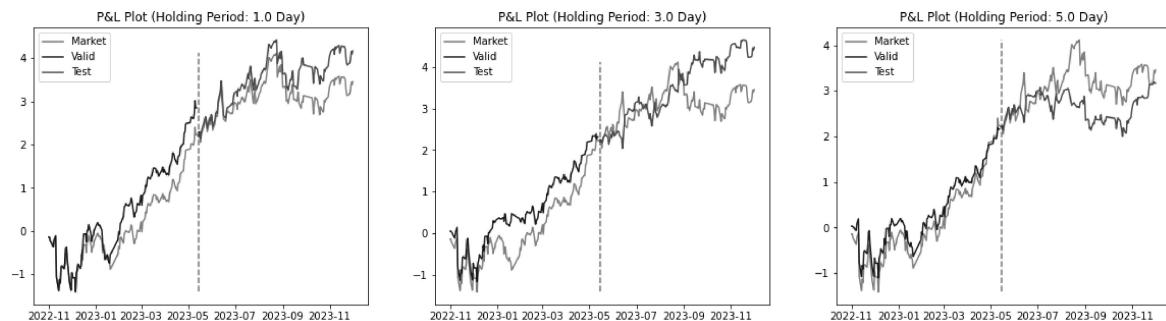


图 4-18 XGBoost 模型 T 策略 P&L

使用 XGBoost 进行 T 合约的预测，当持有期为 1 天、3 天或 5 天时，策略在测试集和验证集上的收益风险比均高于市场基准。

综上所述，XGBoost 模型在预测国债期货合约时具有比较稳定的表现，基于此模型构建的策略在各国债期货合约、各持有期上均获得了超越市场基准的表现。

4.3 本章小结

表 4-23 基于各模型策略的收益风险比（红色代表超过市场基准，蓝色代表弱于市场基准）

		T			TF			TS		
		1	3	5	1	3	5	1	3	5
LR	Valid	1.40	1.96	2.38	3.19	1.94	1.86	0.57	2.52	1.39
	Test	0.76	1.41	0.28	1.55	1.12	-1.03	2.29	2.46	2.01
Lasso	Valid	1.50	1.96	2.38	3.17	2.11	1.86	0.57	2.60	1.44
	Test	0.76	1.41	0.28	1.10	0.90	-1.03	2.29	2.46	2.16
RF	Valid	2.18	1.81	1.62	1.76	1.82	0.93	1.33	0.91	0.95
	Test	1.15	1.93	0.88	1.03	2.32	0.37	-0.29	-0.27	-0.56
MLP	Valid	3.25	3.78	3.58	3.26	2.93	1.94	3.18	4.18	2.28
	Test	-0.55	3.41	1.00	-0.55	1.10	-0.47	0.12	-1.13	-0.70
CatBoost	Valid	1.39	1.58	1.61	2.24	2.13	1.47	0.16	1.18	0.34
	Test	0.48	1.91	0.68	1.30	1.32	1.04	-0.58	-0.04	1.54
XGBoost	Valid	1.94	1.70	1.53	1.86	1.28	1.75	1.71	0.93	1.48
	Test	1.43	1.92	0.97	0.58	1.61	1.22	1.54	1.78	0.91

表 4-23 汇总了基于各模型的策略的收益风险比。本文对每个模型都进行了 9 个测试，分别是针对 TS、TF、T 未来 1、3、5 天收益的预测，并基于预测构建交易策略。

在每个测试中，若策略在验证集和测试集上的收益风险比均优于市场基准，则认

定为优于市场基准。各模型优于市场基准的测试个数如下：

模型	LR	Lasso	RF	MLP	CatBoost	XGBoost
优于市场基准的测试数	5	5	7	4	5	9

在测试的 6 个模型中，优于市场基准的测试数量多于线性回归模型的为随机森林和 XGBoost 两个模型；Lasso 回归和 CatBoost 模型优于市场基准的测试数和线性回归模型一样；多层次感知机模型优于市场基准的测试数少于线性回归模型。

战胜市场是非常困难的，XGBoost 模型在 9 个测试中，均优于市场基准，说明 XGBoost 模型对国债期货合约未来收益的拟合效果较好，预测效果稳定。

本文测试了对国债期货未来 1 天、3 天和 5 天收益的预测。在基于各模型对国债期货未来 3 天收益预测而构建的交易策略中，仅多层次感知机的 TS 策略的收益风险比没有强于市场基准，其余各模型的 3 天持有期策略均强于市场基准。收益风险比超越市场组合的策略中，3 天持有期策略的占比高于 1 天或 5 天持有期策略，说明基于本文的因子集，国债期货未来 3 天的收益能更好地被预测。

本文的下一章将探究部分因子与国债期货未来收益率的关系，并寻找其中蕴含的经济学原理。

第五章 因子与国债期货未来收益的关系

通过传统线性模型的拟合，我们仅能拟合各因子与预测目标之间的线性关系，而通过机器学习模型，可以拟合各因子与预测目标之间的非线性关系。此章将对部分因子与国债期货未来收益的关系进行探究，并从经济学上解释这些关系。

通过表 4-23 可以看出，除了多层感知机在持有期 3 天的 TS 策略上没有优于市场基准外，基于各模型的 3 天持有期策略均优于市场基准，说明本文中的模型和因子对国债期货未来 3 天的收益具有更加稳定的预测能力。因此，此章的研究主要采用模型对国债期货未来 3 天收益的拟合。

在机器学习模型方面，选取在 9 次不同的测试中均优于市场基准的 XGBoost 模型。

5.1 新闻情绪因子与国债期货未来收益的关系

前文中提到，本文构建的新闻情绪因子与 TS、TF 和 T 合约未来 3 天的收益的相关系数均接近 -0.1，相关性很高。下图绘制了基于 XGBoost 模型，新闻情绪因子从 -1 至 1 变化时，国债期货未来 3 天收益率的变化。

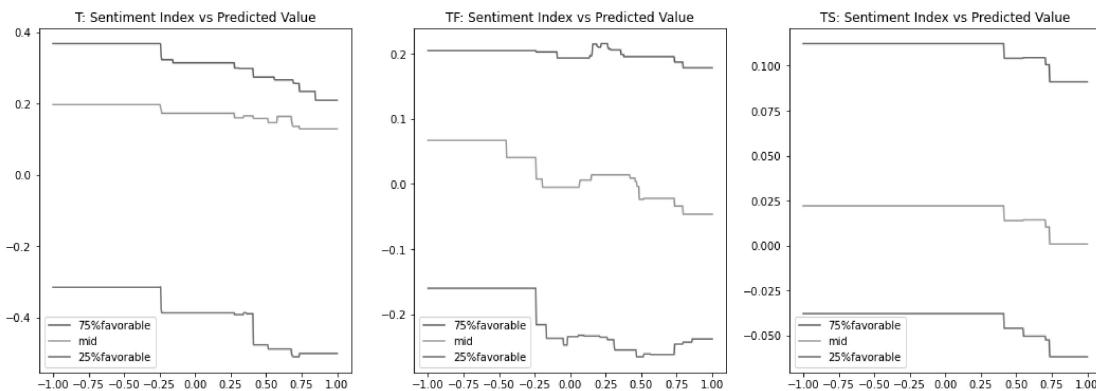


图 5-1 新闻情绪指数和国债期货未来 3 天收益的关系¹

¹ 图中的 mid 指其余各因子取中位数；25%favorable 指与收益正相关的因子取 25% 分位数，与收益负相关的因子取 75% 分位数；75%favorable 指与收益正相关的因子取 75% 分位数，与收益负相关的因子取 25% 分位数，下同

我们可以发现，新闻情绪因子总体上与国债期货的未来收益呈负相关关系。这是符合经济学直觉的。当宏观经济新闻情绪较好时，投资者倾向于将更多的资产分配到风险资产中，因此不利于国债期货价格的上涨。

对比几个合约未来收益与新闻情绪因子的关系，我们可以发现与 T 和 TF 合约相比，TS 合约对新闻情绪因子的变动敏感性相对较低，新闻情绪因子从 -1 至 0.4 变化时，对 TS 合约未来收益几乎无影响，只有当新闻情绪因子的值较高时，TS 合约的未来收益才会随新闻情绪因子的增加而降低。从经济学上看，宏观新闻情绪更多影响了市场参与者对经济基本面的情绪，经济基本面相关的信息对收益率曲线长端影响较大，而收益率曲线短端受资金面和技术面影响较大，对经济基本面信息相对不敏感。因此，相比于 T 和 TF 合约，TS 合约对新闻情绪因子的敏感度较低。

5.2 股债“跷跷板”

权益资产和利率债资产的联动性是最受市场参与者关注的大类资产关系。权益资产和利率债资产分别作为风险资产和避险资产的代表，常常呈现出此消彼长的关系。权益资产上涨时，投资者风险偏好增强，不利于避险资产；权益资产下跌时，市场避险情绪高涨，避险资产收到追捧。以下绘制了沪深 300 指数前一日收益率与国债期货合约未来 3 日收益的关系。

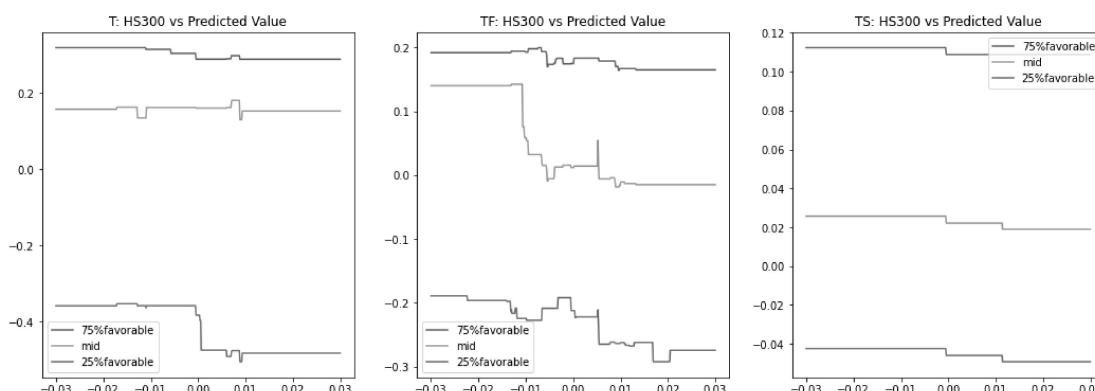


图 5-2 沪深 300 指数前一日收益率和国债期货未来 3 天收益的关系

总体来看，沪深 300 指数前一日的收益率越高，国债期货合约未来 3 天的收益越低，体现出了股债“跷跷板”的效应。

当其它变量处于 75% favorable，即比较有利于国债期货的位置时，国债期货合约

未来收益对沪深 300 指数收益率的斜率绝对值较小，曲线相对平坦。反映了当其它变量处在有利于国债期货的点位时，国债期货受股债“跷跷板”效应的影响相对较弱；反之则受股债“跷跷板”效应的影响相对较强。

5.3 利率互换价格对国债期货的影响

从相关性表中，可以看出，利率互换 FR007: 5Y，即 5 年期利率互换的价格与 TS、TF 和 T 的未来收益均呈正相关。以下是二者关系的散点图。

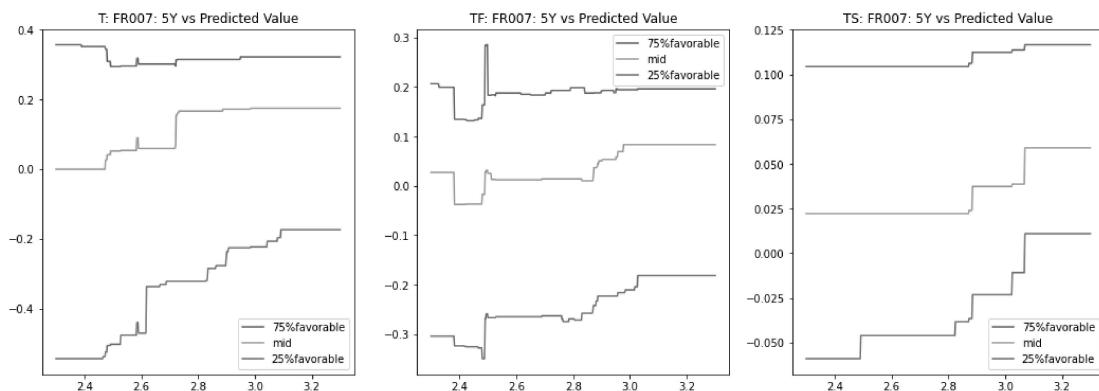


图 5-3 5 年期利率互换和国债期货未来 3 天收益的关系

前一日的利率互换数值反映了前一日市场参与者对未来利率的预期水平，而未来 3 天国债期货的价格变化则反映了市场参与者对未来利率预期的变动。从图 5-3 中可以看出，前一日利率互换的数值与未来 3 天国债期货收益总体上呈正相关关系，即前一日利率互换数值高时，未来国债期货的收益也较高。

这实际上反映了当前一日对未来利率的预期较高时，未来 3 日对未来利率的预期会下降，这体现了对未来利率的预期具有均值回归的属性。

权益市场中，通常来说对未来市场的预期会具有动量叠加的效应，而非均值回归的效应。利率市场中这种对未来利率预期的均值回归效应来源于利率本身的属性。利率作为金融市场上最重要的变量之一，受到宏观调控的影响较大，而宏观调控通常希望将利率维持在一个区间内，而非让利率随着趋势持续上升或下降，这就能解释对未来利率预期的这种均值回归现象。

5.4 本章小结

本章通过 XGBoost 模型的拟合结果，探究了部分因子与国债期货未来 3 天收益之间的关系。

新闻情绪因子总体上与国债期货的未来收益呈负相关关系。且与 T 和 TF 合约相比，TS 合约对新闻情绪因子的敏感性相对较低。

沪深 300 指数前一日的收益率越高，国债期货合约未来 3 天的收益倾向于越低，体现出了股债“跷跷板”的效应。当其它变量处于有利于国债期货的点位时，国债期货受股债“跷跷板”效应的影响相对较弱；反之则受股债“跷跷板”效应的影响相对较强。

前一日利率互换的数值与未来 3 天国债期货收益总体上呈正相关关系，即前一日利率互换数值高时，未来国债期货的收益也较高，体现了对未来利率预期的“均值回归”属性。

第六章 总结

6.1 主要结论

本文构建了丰富的因子体系，包括技术指标、宏观经济指标、其它大类资产价格走势、市场资金利率、新闻情绪，这些指标包括了影响国债期货价格变化的各种因素，指标体系全面。其中，新闻情绪指标创新性地使用 ChatGPT 对新闻日报中的每条新闻情绪进行判断，并将结果汇总为新闻情绪指数；对于宏观指标，为了更好地反映宏观指标对国债期货未来收益的影响，没有直接采用宏观指标本身的值作为因子，而是创新性地采用各宏观指标是否优于或弱于万得一致预期来作为因子。

在得到初始的因子池后，为了避免模型过拟合又进行了因子降维。采用分步回归筛选出了部分预测力较强的因子。

之后，本文采用线性回归、Lasso 回归、随机森林、多层感知机、CatBoost 和 XGBoost 共 6 种模型进行国债期货未来收益的预测。在预测 TS、TF 和 T 合约的未来收益时，预测周期分别为 1 天、3 天和 5 天。综上所述，每个模型分别进行 9 个测试，分别为对 TS、TF 和 T 合约未来 1 天、3 天和 5 天收益的预测。基于预测结果，当预测未来收益大于 0 时做多国债期货合约，当预测结果小于 0 时做空国债期货合约。

在各模型中，对国债期货的预测表现最为稳定的是 XGBoost 模型，XGBoost 模型在 9 个测试中，策略均实现在验证集和测试集上获得比市场基准更高的收益风险比。因为国债期货合约是保证金交易，可以很容易地增加杠杆，所以比起策略的绝对收益，我们更注重承担单位风险所获得的收益，即收益风险比。

相比于 1 天和 5 天期，本文的模型和因子集对国债期货未来 3 天期的收益拟合效果更好。除了多层感知机模型的 TS 策略外，基于各模型的 3 天持有期策略在验证集和测试集上的收益风险比均高于市场基准。优于市场基准的策略中，3 天持有期的策略占比高于 1 天和 5 天持有期的策略。

对于部分变量与国债期货未来收益的关系，本文有如下发现：

1. 新闻情绪因子总体和国债期货未来 3 天收益呈负相关关系。且相比于 TF 和 T 合约，TS 合约对新闻情绪因子变动的敏感性相对较低，只有当新闻情绪因子的值较高时，TS 合约的未来收益才会随新闻情绪因子的增加而降低。

2. 沪深 300 指数前一日的收益率越高，国债期货合约未来 3 天的收益倾向于越低，体现出了股债“跷跷板”的效应。当其它因子处于有利于国债期货的点位时，国债期货受股债“跷跷板”效应的影响相对较弱。
3. 前一日利率互换的数值与未来 3 天国债期货收益呈正相关关系。反映了当前一日对未来利率的预期较高时，未来 3 日对未来利率的预期会下降。这个现象说明对未来利率的预期具有“均值回归”的属性。

6.2 研究展望

本文构建了全面的因子体系，使用线性模型和机器学习模型对中国国债期货的未来收益进行预测，并构建交易策略。本文对比了不同模型构建的交易策略的盈利能力，并研究了部分因子与国债期货未来收益之间的关系。本文的研究还有如下扩展方向：

- 1、采用国债期货高频数据。本文研究采用的是国债期货日频数据，目前市场上可得的中国国债期货最高频率的数据为间隔 500ms 的订单簿切片数据。采用更高频率的数据可以弥补国债期货数据量较小的问题。
- 2、采用深度学习模型。本文受制于数据量较小，没有使用深度学习模型对国债期货未来收益进行预测。若采用国债期货高频数据，则可配合深度学习模型进行预测。
- 3、进一步丰富因子体系。采用另类数据、国债期货订单簿（Limit Order Book）多档数据等，进一步丰富因子体系。
- 4、探索由预测结果转化为交易策略的法则。本文当预测未来收益大于 0 时做多，预测未来收益小于 0 时做空。可探索更加细化的交易法则，如预测未来收益大于 t ($t>0$) 时才做多，预测未来收益小于 $-t$ ($t>0$) 时才做空，或可提高策略胜率。还有可以在预测未来收益的绝对值较大时设置仓位较大，反之仓位较小，或可提高策略盈亏比。

参 考 文 献

- [1] Bansal, N., Connolly, R.A., Stivers, C. Regime-switching in stock index and Treasury futures returns and measures of stock market stress[J]. *The Journal of Futures Markets*. 2010, 30(8):753-779.
- [2] Brandt, M.W., Kavajecz, K.A., Underwood, S.E. Price discovery in the treasury futures market[J]. *The Journal of Futures Markets*. 2007, 27(11):1021-1051.
- [3] Chen, T., Guestrin, C. XGBoost: A Scalable Tree Boosting System[C]. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016: 785-794.
- [4] Drobetz, W., Otto, T. Empirical asset pricing via machine learning: evidence from the European stock market[J]. *Journal of Asset Management*. 2021, 22(7):507-538.
- [5] Dungey, M., Hvozdyk, L. Cojumping: Evidence from the US Treasury bond and futures markets[J]. *Journal of Banking & Finance*. 2012, 36(5):1563-1575.
- [6] Fama, E.F., French, K.R. A five-factor asset pricing model[J]. *Journal of Financial Economics*. 2015, 116(1):1-22.
- [7] 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.
- [8] Gabriel ANGHEL, D. Predicting Intraday Prices in the Frontier Stock Market of Romania Using Machine Learning Algorithms[J]. *International Journal of Economics and Financial Research*. 2020, (67):170-179.
- [9] Gu, S., Kelly, B., Xiu, D. Empirical Asset Pricing via Machine Learning[J]. *The Review of Financial Studies*, 2020, 33(5):2223-2273.
- [10] Hancock, J.T., Khoshgoftaar, T.M. CatBoost for big data: an interdisciplinary review[J]. *Journal of Big Data*. 2020, 7(1):94.
- [11] Ho, T. The random subspace method for constructing decision forests[C], *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1998, 20(8):832-844.
- [12] Ho, T.. Random decision forests[C]. *Proceedings of 3rd International Conference on Document Analysis and Recognition*. 1995, 1:278-282.
- [13] Holder, M.E., Qi, M., Sinha, A.K. The impact of time duration between trades on the price of treasury note futures contracts[J]. *The Journal of Futures Markets*, 2004, 24(10):965-980.
- [14] Park, T.H., Switzer, L.N. Mean reversion of interest-rate term premiums and profits from trading strategies with treasury futures spreads[J]. *The Journal of Futures Markets*, 1996, 16(3):331-352.
- [15] Ruan, Q., Zhou, M., Yin, L. et al. Hedging effectiveness of Chinese Treasury bond futures: New evidence based on nonlinear analysis[J]. *Physica A*. 2021, 565:125553.
- [16] Sharpe, W.F. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk[J]. *The Journal of Finance* 1964, 19(3):425-442.
- [17] Smales, L.A. The effect of treasury auctions on 10-year Treasury note futures[J]. *Accounting and*

- Finance. 2021, 61(1):1517-1555.
- [18] Tang, D., Yang, Y., Yu, Y. Price Discovery and Volatility Spillover Effect in Treasury Bond Futures and Spot Markets: Evidence from China[J]. IOP Conference Series. Materials Science and Engineering, 2018, 439(3):32056.
- [19] Tibshirani, R. Regression Shrinkage and Selection via the Lasso[J]. Journal of the Royal Statistical Society. 1996, Series B, Methodological, 58(1):267-288.
- [20] Wang, Q. Cryptocurrencies asset pricing via machine learning[J]. International Journal of Data Science and Analytics. 2021, 12(2):175-183.
- [21] Zhang, M., Zhao, Y., Nan, J. Economic policy uncertainty and volatility of treasury futures[J]. Review of Derivatives Research. 2022, 25(1):93-107.
- [22] Zhu, L., Wu, H., Wells, M.T. News-Based Sparse Machine Learning Models for Adaptive Asset Pricing. Data Science in Science, 2023, 2(1).
- [23] Zou, L., Rose, L.C., Pinfold, J.F. ASYMMETRIC INFORMATION IMPACTS: EVIDENCE FROM THE AUSTRALIAN TREASURY-BOND FUTURES MARKET[J]. Pacific Economic Review. 2007, 12(5):665-681.
- [24] 方毅, 陈煜之. 机器学习与线性模型在资产收益率预测中的比较[J]. 统计与决策. 2022, 38 (16): 180-183.
- [25] 李斌, 邵新月, 李玥阳. 机器学习驱动的基本面量化投资研究[J]. 中国工业经济. 2019 (8): 61-79.
- [26] 李骏琪, 杨垚立. “CNN+TCN” 模型在高频国债期货市场走势预测中的应用[J]. 统计与决策. 2022, 38 (1): 145~148.
- [27] 李仁宇, 叶子谦. 基于机器学习的基金收益预测[J]. 统计与决策. 2023, 39 (11): 156-161.
- [28] 孙达昌, 毕秀春. 基于深度学习算法的高频交易策略及其盈利能力[J]. 中国科学技术大学学报. 2018, 48 (11): 923-932.
- [29] 许杰, 祝玉坤, 邢春晓. 机器学习在金融资产定价中的应用研究综述[J]. 计算机科学. 2022, 49 (6): 276-286.
- [30] 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.
- [31] 周志华. 机器学习[M]. 北京: 清华大学出版社. 2016: 98-107

致 谢

首先，非常感谢我的导师资超老师！在毕业论文选题和论文撰写的过程中，资老师给我提供了细致入微的指导。在资老师的帮助下，我的论文得到了很大的提升。

在论文写作的过程中，资老师时常与我就论文内容进行交流，当我论文撰写遇到困难时，资老师会给予我针对性的指导。资老师的指导总是一针见血，使我醍醐灌顶。

我还要感谢我的室友和同学，在论文写作的过程中，我和同学们互相帮助，遇到疑难问题时，我们积极讨论，互帮互助；我们互相加油打气、互相督促。

最后，我要向上海高级金融学院的所有老师和 22 级金融硕士班的同学表达深深的感谢。两年的时光里，是老师们的教导和同学们的陪伴使我成为更好的自己！愿各位在今后的工作和生活里一切顺利，越来越好！