

基于CNN-LSTM-Attention的国债收益率预测研究

黎悦, 李俞霖

贵州大学经济学院, 贵州 贵阳

收稿日期: 2025年4月11日; 录用日期: 2025年4月25日; 发布日期: 2025年5月29日

摘要

国债是利率债的一种重要形式, 其到期收益率在经济金融体系中扮演着关键的参考角色。当前, 市场对国债收益率的预测精度提出了更高的要求, 这不仅体现在对宏观经济趋势的判断上, 也体现在微观利率债交易中。然而, 宏观经济分析涉及的指标众多且错综复杂, 指标间的领先滞后关系也在不断变化, 这使得传统的定性分析和定量分析方法在预测的精确度和准确性方面存在一定的局限性。鉴于此, 本文采用了一种新型的深度学习预测模型来预测长期国债的收益率, 并将其与基础的LSTM模型进行比较。该模型融合了卷积神经网络(CNN)、长短期记忆网络(LSTM)和注意力机制(AM)。首先, 通过CNN对输入数据进行特征提取; 然后, LSTM利用这些特征数据来预测国债次日的收益率; 最后, 借助调幅法来捕捉不同时间特征状态对国债收益率的影响, 以提升预测的精度。根据机器学习时序预测技术的研究成果, 集成CNN-LSTM-Attention模型相比基础LSTM模型能得到更好的预测效果, 本文对此进行了验证, 得到的结果与此相符。

关键词

国债收益率预测, LSTM长短时记忆网络, CNN卷积神经网络, 注意力机制

Research on the Prediction of Treasury Bond Yield Based on CNN-LSTM-Attention Model

Yue Li, Yulin Li

School of Economics, Guizhou University, Guiyang Guizhou

Received: Apr. 11th, 2025; accepted: Apr. 25th, 2025; published: May 29th, 2025

Abstract

Treasury bond is an important form of interest rate debt, and its yield to maturity plays a key reference role in the economic and financial system. At present, the market has put forward higher

requirements for the prediction accuracy of treasury bond yield, which is not only reflected in the judgment of macroeconomic trends, but also reflected in the trading of micro interest rate bonds. However, macroeconomic analysis involves numerous and complex indicators, and the leading and lagging relationships between indicators are constantly changing, which limits the accuracy and precision of traditional qualitative and quantitative analysis methods in forecasting. In view of this, this paper adopts a new deep learning prediction model to predict the yield of long-term treasury bond, and compares it with the basic LSTM model. This model integrates Convolutional Neural Network (CNN), Long Short Term Memory Network (LSTM), and Attention Mechanism (AM). Firstly, feature extraction is performed on the input data using CNN; Then, LSTM uses these characteristic data to predict the yield of treasury bond the next day; Finally, the amplitude modulation method is used to capture the impact of different time characteristic states on treasury bond yield, so as to improve the accuracy of prediction. According to the research results of machine learning time series prediction technology, the integrated CNN-LSTM-Attention model can achieve better prediction performance compared to the basic LSTM model. This paper verifies this and the results obtained are consistent with this.

Keywords

Treasury Bond Yield Prediction, LSTM Long-Term and Short-Term Memory Network, CNN Convolutional Neural Network, Attention Mechanism

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

国债在金融市场中以其极低的信用风险和较高的流动性而著称, 其到期收益率常被视为一国的无风险利率, 是金融资产定价的关键参数, 也是市场利率的指示器。从宏观角度来看, 长期国债收益率与经济增长之间的联系比短期国债收益率更为紧密(杨华, 2022) [1], 因此, 长期国债收益率常被用作宏观经济分析的核心指标。预测其走势有助于预判经济增长趋势和未来市场的资金供需状况, 这对宏观政策制定者具有重要的研究意义。在微观层面, 国债作为一种安全性极高的利率债投资品种, 常被配置于各类资管产品和机构自营资产中, 也是商业银行、保险公司、基金等机构投资者频繁交易的对象。因此, 这些金融机构的研究部门通常会定期对国债收益率的走势进行分析, 以辅助其投资决策。

目前, 国债收益率的预测研究主要分为定性和定量两种方法。定性分析通过综合考虑宏观经济基本面、政策面、资金面等因素, 分析国债收益率在当前经济周期中的变动逻辑, 从而判断其未来走势。定量分析则通过理论分析各因素对国债收益率的影响机制, 并运用统计方法构建模型。常用的定量模型包括 ARIMA 模型和 VAR 向量自回归模型等, 这些模型在收益率预测和收益率曲线结构拟合方面取得了一定成果, 但也存在诸多数据条件限制, 如 VAR 模型要求序列平稳等。这些限制使得在模型中考虑多种因素时, 需要对数据进行多步骤的检验和处理。有研究表明, 这些方法的国债收益率预测精度较低, 预测效果一般。

随着宏观经济因素和政策信息的复杂性不断增加, 传统的利率走势分析框架面临挑战。近年来, 一些研究开始采用多种机器学习模型来预测资产收益率, 并证实部分模型具有良好的预测效果。其中, 卷积神经网络(CNN)和长短时记忆网络(LSTM)作为深度学习模型, 通过更多的隐藏层实现更精准的预测。CNN 能够通过卷积层和池化层对信息的不同特征赋予不同权重, 提取特征并减弱噪声, 而 LSTM 则能保

留信息的长短期关联特征。结合注意力机制(AM), 可以捕捉时间序列数据中过去特征状态对次日收益率的影响。本文将这三者结合成一个集成模型, 并与基础 LSTM 模型的结果进行比较。

下文的结构安排如下: 第二部分是文献综述与本文贡献; 第三部分是模型介绍; 第四部分是基于组合模型的利率预测实证研究; 第五部分是结论。

2. 文献综述与本文贡献

仅依据国债收益率自身序列的历史信息进行时序预测分析, 难以顾及其他因素的影响, 因此有研究致力于探寻可提升利率预测效果的其他变量。陈哲(2008)和帅昭文等(2017)的实证研究均表明, 宏观经济变量能显著提高利率变动预测的准确性, 尤其是对长期收益率的拟合预测效果更为显著[2][3]。当经济处于成长期时, 市场对资金的需求不断攀升, 致使企业资金紧张, 利率随之上升, 债券收益率也水涨船高; 反之, 利率则会下降。

金融时间序列数据蕴含诸多不确定因素, 除了常见的趋势性、季节性、自相关性等特征外, 还受到大量噪声的干扰。这些特性使得传统的线性技术难以实现精准预测。对国债收益率的研究主要分为两大类: 传统的计量模型时序预测研究和基于机器学习的时序预测研究。在国内学者的研究中, 杨婉茜和成力为(2016)指出 VAR 模型参数众多, 易导致过度拟合, 模型在样本外的泛化能力欠佳, 且对样本的时间区间要求严格, 预测效果不稳定; 而 Fama-Bliss 模型的预测效果甚至不如简单的随机游走模型[4]。文献[5]通过建立无套利 AFNS 模型来预测不同未来期限的国债收益率, 发现不同变体模型的 RMSE 差异较大, 样本外预测误差在预测步长较短时表现最佳。文献[6]对比了三因子 Vasicek、三因子 CIR、多项式样条、指数样条、DL、动态 SV 模型在中美两国国债收益率期限结构样本内外预测的精确度, 结果显示样本内动态 SV 模型和 DL 模型在两国收益率预测上均表现良好, 但在样本外预测时, 各模型的预测精度参差不齐, 均处于较高水平, 其中指数样条模型和 DL 模型的预测能力相对更优, 两类模型的 RMSE 值介于 0.2~0.7 之间。在国际研究方面, Vasicek (1977)较早地提出了均衡利率模型, 认为利率具有均值回归的特性, 从而使得国债收益率呈现出一定的周期性。Fama 和 Bliss (1987)提出了远期利率回归模型, 发现远期利率能够预测利率和债券回报率的变化, 尤其在长期国债收益率的预测上更具优势[7]。Brandt 和 Wang (2003)的研究发现投资者的风险厌恶程度会受到通胀非预期变化的影响, 实证结果表明通胀的非预期变动增加时, 投资者的风险厌恶水平上升, 进而推高了债券的预期回报率[8]。

然而, 随着神经网络技术在深度学习时代的飞速发展, 机器学习技术在各个领域的应用日益广泛。诸如支持向量机(SVM)、随机森林(RF)、人工神经网络(ANN)等机器学习技术与金融时序数据研究的结合日益紧密, 一定程度上突破了传统模型线性技术的局限, 在非线性拟合能力方面取得了显著进展。在利用机器学习方法预测国债收益率的研究领域, 相关文献相对较少。闫红蕾和张自力(2018)运用 NARX 神经网络模型对国债利率期限结构进行了拟合预测[9]。宋鹏、张淼等(2021)构建了 VAR-LSTM 框架, 以宏观经济变量预测值和序列滞后项作为输入因子, 有效解决了传统 LSTM 模型拟合结果“平行错位”的问题, 显著提升了预测精度, 实证结果也表明 VAR-LSTM 集成模型的预测精度优于 ARIMA 模型, 而 ARIMA 模型又优于单一 LSTM 模型[10]。王雷等(2021)采用卷积神经网络(CNN)与 DL 模型对比预测债券收益率曲线, 获得了更低的预测误差[11]。尽管如此, 目前尚未有使用 CNN-LSTM-ATTENTION 集成神经网络模型对国债收益率进行预测的研究, 本文正是试图填补这一研究空白, 这也是本文的创新之处所在。

3. CNN-LSTM-Attention 模型介绍

CNN-LSTM-Attention 模型是一种深度学习框架, 专门用于分析具备时空关联特性的序列数据。其融合了卷积神经网络(CNN)、长短期记忆网络(LSTM)以及注意力机制(Attention)这三种神经网络层, 能高效

地提取数据中的空间与时间属性。

长短期记忆网络(LSTM)，作为循环神经网络(RNN)的一种改进形式，擅长处理序列数据中的时间关联性。借助输入门、遗忘门和输出门的调控机制，LSTM 能够精准地记忆或舍弃序列中的重要信息。在进行时间序列预测时，LSTM 能够识别数据中的长期关联性。

注意力机制层(Attention)的功能是评估输入序列中各个时间点之间的相互关联。通过计算每个时间点与序列中其他时间点的注意力权重，模型能够对整个序列的信息进行加权整合。在处理序列数据时，Attention 机制使模型能够聚焦于序列中的关键时间点，同时忽略不那么重要的部分。

在 CNN-LSTM-Attention 模型中，首先利用 CNN 层来抽取序列的空间特征，然后通过 LSTM 层来把握时间上的依赖关系，最后借助 Attention 层来强化对序列中关键时间点的关注度。这种多层级的特征提取方式，整合了 CNN 在空间特征提取上的优势、LSTM 在时间序列建模上的能力以及 Attention 在捕捉全局关联性上的特长，从而能够更精准地处理复杂序列数据。凭借其出色的特征提取和序列分析能力，CNN-LSTM-Attention 模型可广泛应用于各类时空数据的预测与分析工作，为实际应用场景提供了强大的技术支撑。

(一) 卷积神经网络(CNN)

CNN 是一种深度学习架构，专注于从序列数据中提取空间特征。它通过在输入数据上移动卷积核来识别局部区域的模式，例如图像中的边缘或时间序列数据中的特定形状。在序列数据处理的过程中，CNN 能够有效捕捉局部特征，充分发挥其精准且全面地提取特征的强大能力。卷积操作通常与池化层配合使用，池化层的作用是减少特征图的空间尺寸，降低计算复杂度，同时保留关键特征。当处理包含大量信息的原始数据时，CNN 可以对其进行维度降低，从而减少参数数量，使学习过程更加精确和高效。

就国债收益率预测数据集而言，该数据集中蕴含着诸如利率期限结构等复杂的空间关联性特征。鉴于卷积神经网络在捕捉空间特征方面具备显著优势，运用 CNN 对国债收益率预测数据集中的空间特征予以捕捉从理论上具有充分的合理性与可行性，有望为国债收益率预测任务提供有力支持，提升预测模型的性能表现。

CNN 网络由以下五个主要部分组成：输入层、卷积层、池化层、全连接层和输出层。输入层将每个数据点视为一个特征点。卷积层包含多个不同尺寸的卷积核，其主要功能是从原始数据中提取关键信息。池化层对卷积层的结果进行降维和压缩，提取重要特征并减少网络参数。全连接层将池化层输出的三维张量转换为向量形式，以便传递给下一层。输出层采用 Softmax 层，Softmax 层接收全连接层的输出，并将其作为整个神经网络的最终输出。该结构如图 1 所示。

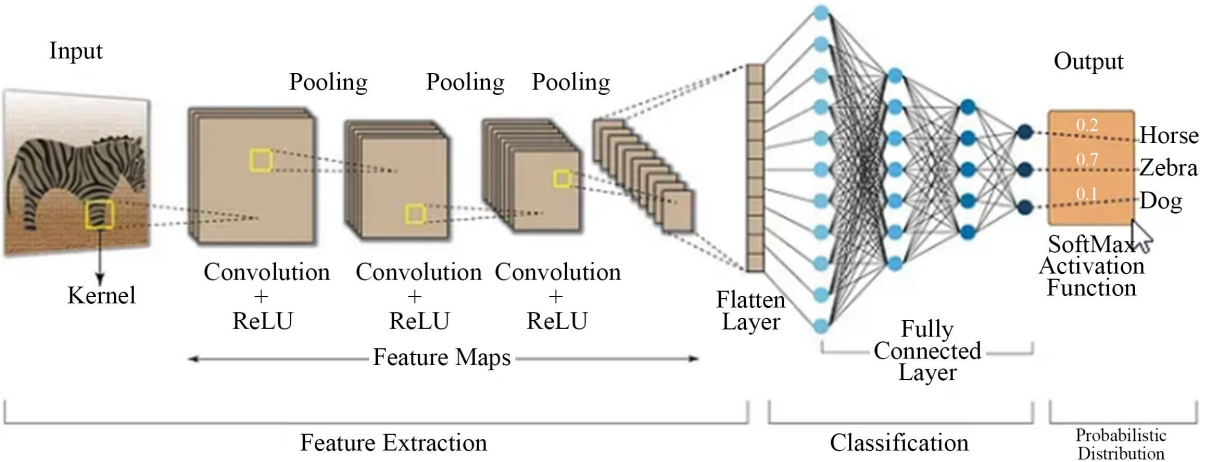


Figure 1. Structure diagram of CNN
图 1. CNN 结构图

(二) 长短时记忆神经网络(LSTM)

长短期记忆网络(LSTM)在循环神经网络(RNN)的基础上进行了改进, 它具备利用反向传播算法学习序列数据中长期依赖关系的能力。LSTM 网络通过其隐藏层中的循环连接来保存短期信息, 并能够有效地从序列数据中提取信息。不过, 标准的 RNN 面临着梯度消失的问题, 这一问题阻碍了模型对长期依赖关系的学习。为了解决这一难题, LSTM 网络在存储单元中保存有用信息的同时丢弃无用信息, 从而克服了传统 RNN 的局限, 展现出更优的性能表现。

LSTM 的每个单元由一个细胞状态(存储单元)和三个门控机制构成: 输入门、输出门以及遗忘门。当数据流经这种结构, LSTM 能够有选择性地保留有价值的信息, 决定哪些信息需要被“遗忘”, 哪些信息需要被“记住”, 以此来构建信息流动, 进而学习到序列数据中的长期依赖关系。其机制图如图 2。

针对本文所聚焦的国债收益率数据集, 国债收益率呈现出短期动量效应以及长期均值回复效应的典型变动特征。卷积神经网络(CNN)擅长提取局部特征, LSTM 在处理长期依赖关系上优势明显。将两者结合能够充分发挥 CNN 对国债收益率数据局部特征的提取能力, 以及 LSTM 对数据长短期信息的有效捕捉与处理能力。因此, 从理论层面分析, 在 CNN 的基础上结合 LSTM 模型来提取国债收益率数据集中的这些特征, 具有较高的合理性与可行性, 有望为相关研究与预测工作提供有力支撑。

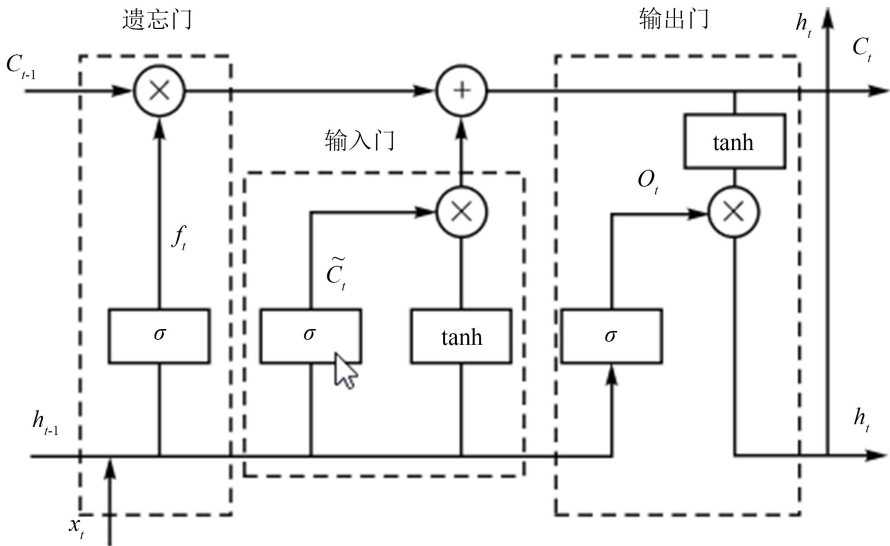


Figure 2. Mechanism diagram of LSTM
图 2. LSTM 机制图

(三) Attention 注意力机制

注意力机制(AM)最初由 Treisman 等人在 1980 年提出。该机制通过计算注意力权重的概率分布, 从海量信息中筛选出关键信息, 突出显示重要的输入内容, 以此来优化传统模型。AM 的核心理念源自人类的视觉注意力过程。人类的视觉系统能够迅速定位到关键区域, 并将注意力聚焦于这些区域, 以便获取所需的详细信息。类似地, 注意力机制有选择性地聚焦于较为重要的信息, 同时忽略那些不重要的信息, 并对信息的重要性进行分配。

(四) CNN-LSTM-Attention 网络模型

CNN-LSTM-Attention 模型是由卷积神经网络(CNN)、长短期记忆网络(LSTM)以及注意力机制(Attention)组合而成的深度学习集成模型。该模型的数据处理流程涵盖以下五个关键步骤: 首先, 将输入的时间序列数据转换成适合神经网络处理的矩阵格式; 接着, 将转换后的矩阵输入至 CNN 网络中, 以实现数

据特征的提取与维度降低; 然后, 将 CNN 提取出的特征序列送入 LSTM 网络进行进一步的训练处理; 之后, 利用 Attention 机制对 LSTM 网络的输出结果进行自动加权平均操作; 最后, 通过全连接层对处理后的数据进行计算, 以生成最终的预测结果。过程如图 3。

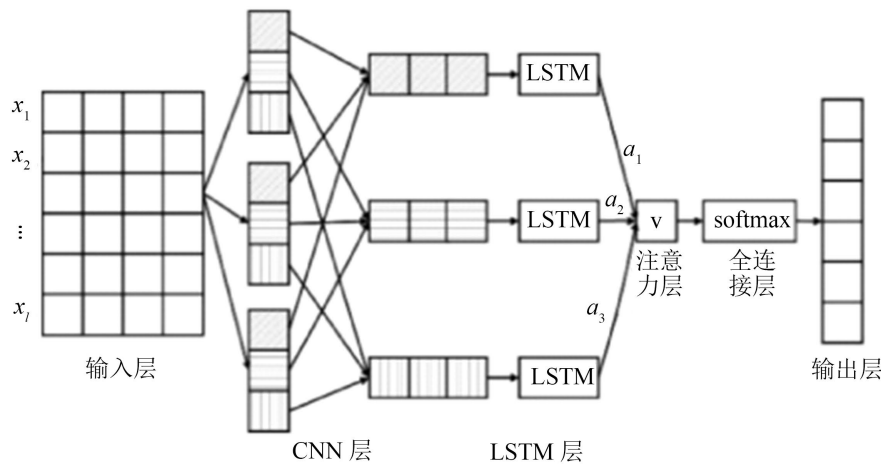


Figure 3. Structure diagram of the CNN-LSTM-attention model
图 3. CNN-LSTM-attention 模型结构图

4. 基于组合模型的收益率预测实证研究

(一) 预测变量

本文以十年期国债到期收益率为研究焦点, 将其作为各预测模型的预测变量。在众多不同期限的国债收益率中, 10 年期国债收益率因其独特的优势而被广泛用于评估经济周期和宏观经济趋势。该期限国债的发行规模庞大, 在银行间市场、交易所市场以及银行柜台市场均有广泛的交易活动, 且以其高安全性和活跃的交易特性著称。从数值角度来看, 我国通常将十年期国债收益率视为无风险利率的代表。此外, 十年期国债收益率也被认为是资金成本的体现, 即几乎无信用风险的国家主体进行融资时所需承担的成本价格, 理论上应构成整个市场资金成本的下限。同时, 作为长期利率的代表, 十年期国债收益率还隐含了市场参与者对经济前景的预期。

(二) 日频变量体系构建

在金融市场中, 国债收益率数据是按日更新的高频数据, 而经济数据指标多为中低频数据。这就导致在实际应用中, 利用低频的经济数据来分析日频的金融数据变化趋势在实践上具有困难。因为依赖中低频数据来洞察经济金融的运行状况, 往往难以捕捉到最新的动态, 进而影响研究和决策的及时性。若能将经济领域的低频数据转化为相应的日频数据, 就能实现经济与金融领域的有效对接。

本研究旨在构建一个日频自变量体系, 其主要目的有两个: 一是减少低频宏观经济指标的滞后效应, 二是尽可能满足深度神经网络对大量数据的需求。为了构建日频指标, 需要深入分析影响国债到期收益率的因素。目前, 主流的方法是把这些因素划分为三个主要类别: 经济基本面、资金流动性以及政策环境。

1) 基本面指标

在宏观经济基本面的分析中, 经济增长类指标主要以 GDP 增速为代表, 而通货膨胀类指标则以 CPI 或 PPI 为主。然而, 这些指标的发布频率相对较低, 因此需要从其计算方法出发, 拆解出更接近日频的指标。

对于经济增长类指标, 采用支出法计算的 GDP 涵盖了社会消费品零售总额、固定资产投资和出口金额三大类。社会消费品零售总额可进一步细分为多个品类, 其中汽车、石油及制品、粮油食品和住房相

关消费等品类尤为重要。众多金融研究机构为了更准确地描绘经济变动,通常会关注汽车、住房、石油等可选消费品的消费状况。参考已有分析和文献,石油消费属于价格主导型消费,其消费量波动与油价密切相关,且石油下游消费对油价波动也较为敏感。因此,本文选用布伦特原油现货价格指数作为社会消费品零售总额中的日频指标。在固定资产投资方面,为了同时反映基建和房地产投资的增长,理论上钢铁、水泥等建材类指标具有较好的代表性。例如,当基建和房地产投资增加时,对钢铁、水泥等建材的需求自然上升。本文选择全国水泥价格指数纳入日频指标体系。

通货膨胀相关的日频指标主要从消费和生产两个角度的物价指数 CPI 和 PPI 入手拆分。通货膨胀状况会影响央行的货币政策,康书隆和王志强(2010)通过实证分析得出我国长期利率与 CPI 走势具有较强相关性的结论[12]。根据兴业证券研究院对 2019 年我国 CPI 构成的各分项权重估算以及相关文献,交通通信在 CPI 中的权重和波动均较高,且交通工具用燃料主要受原油价格影响,属于石油消费的下游。因此,本文继续采用布伦特原油现货价格作为 CPI 项下的日频指标。对于 PPI(工业生产者出厂价格指数),它是衡量工业企业产品出厂价格变动趋势和变动程度的指数,分为生产资料(采掘、原材料、加工)和生活资料(食品、衣着、一般日用品、耐用消费品)两大类。袁伟鹏等人(2021)通过建立 PPI 与十年期国债收益率的 VAR 模型,并运用格兰杰因果检验及脉冲响应函数分析等方法,证实了 PPI 对十年期国债收益率走势具有预测作用[13]。信达证券研究所(2022)指出,真正显著影响 PPI 的分项为黑色金属、石化(石油+化工)、煤炭等,可选择的日度指标通常包括布伦特原油、螺纹钢等大宗商品价格指数。本文根据数据可获时间区间以及与 PPI 的相关性,仍然选择布伦特原油现货价格作为日频指标。

2) 资金面指标

魏雪梅在 2013 年对中美两国国债市场进行了研究,其实证分析发现,央行的货币政策对国债收益率有着显著的影响[14]。刘郁和田乐蒙等人在 2022 年的研究中运用了 VAR 模型,探讨了常见经济金融变量对利率中枢的作用,结果显示,短期利率的代表变量 R007 对长期收益率的影响较为直接,而社会融资等与经济增长相关的指标对利率中枢的影响则相对稳定。

目前,货币市场中存在多种利率指标,这些指标可以大致分为几个主要类别:银行间利率、交易所利率以及存贷款利率等。其中,一些较为常见的利率指标包括存款类机构质押式回购利率(DR)、银行间质押式回购利率(R)、上交所质押式国债回购利率(GC)以及上海银行间同业拆放利率(Shibor)等。在这些利率指标中,R 系列利率的参与机构涵盖了银行间市场的广泛交易者,具有不易被操控、反应迅速等特点。由于国债的主要交易场所是银行间市场,且本文的预测目标为国债到期收益率,主要关注的是银行间国债市场,同时考虑到 1 天和 7 天的交易量较为普遍,因此本文选择了 7 天的银行间质押回购利率作为研究对象。此外,为了获取与长端国债收益率更为直接相关的指标,本文还将短期 1 年期国债收益率纳入了研究体系。

3) 海外因素指标

一些实证研究已经发现美元指数与中国国债收益率之间存在显著的负相关关系,然而,这种影响的具体机制尚不明确。目前有观点认为,美元作为一种避险资产,其指数上升意味着全球市场风险的增加,进而导致市场避险情绪的上升。在这种情况下,中国国内的资金在寻求避险资产时,通常会倾向于选择国债,这使得国债收益率出现下降。因此,本文将美元指数纳入作为该类指标的一部分。

4) 其他因素指标

股票市场与债券市场之间通常存在着紧密的相互关联,特别是在股票市场出现大幅波动时,国债的避险属性尤为显著(李湛和唐晋荣, 2017)[15]。历史数据和市场经验表明,在股票市场表现强劲时,债券市场往往表现不佳;相反,当股票市场处于熊市时,债券市场则通常会迎来牛市。鉴于此,本文选用上证综指来刻画股市的走势特点。上证综指能够较为全面地体现市场股价的整体变动趋势。基于此,本文的日频指标归纳见表 1。

Table 1. Daily frequency feature index system
表 1. 日频特征指标体系

归类因素	变量名称	数据来源	变量名
资金面、政策面	银行间质押回购利率：7 天	中国货币网	FR007
	中债国债到期收益率：1 年	中国货币网	YTM1
基本面的	固定资产投资	Wind	CEM
	消费	Wind	BRT
	CPI、PPI	猪肉平均批发价	PORK
		28 种重点监测蔬菜平均批发价	VEG28
		布伦特原油现货价	BRT
海外因素	美元指数	Wind	UDI
其他因素股债跷跷板	上证综合指数	Wind	SHI
预测变量	中债国债到期收益率：10 年	中国货币网	YTM10

(三) 描述性统计

对于各变量，计算出均值、标准差等指标，在表 2 中列出：

Table 2. Sample descriptive statistics of variables in the index system
表 2. 指标体系各变量的样本描述统计

	VEG28	PORK	CEM	BRT	SHI
count	2487.00	2487.00	2487.00	2487.00	2487.00
mean	4.46	25.72	129.11	66.33	3193.85
std	0.74	8.83	29.50	19.74	339.80
min	3.05	15.62	77.76	13.28	2464.36
25%	3.88	20.09	106.46	50.80	2978.65
50%	4.37	22.60	137.75	65.54	3168.65
75%	4.96	26.56	149.50	79.26	3361.46
max	6.46	52.40	213.81	137.71	5166.35
	UDI	FR007	YTM1	YTM10	
count	2487.00	2487.00	2487.00		2487.00
mean	98.06	2.46	2.41		3.05
std	4.87	0.59	0.55		0.44
min	88.65	1.30	1.09		1.72
25%	94.54	2.05	2.08		2.77
50%	97.22	2.38	2.34		3.02
75%	101.74	2.70	2.69		3.34
max	114.16	6.00	3.81		4.05

依据以上指标，从其平均值角度进行分析，资金利率的代表指标 FR007 的平均值为 2.46，布伦特原

油的平均价格为 66.33, 水泥价格指数的平均值为 129.11, 猪肉批发价格的平均值为 25.72, 而重点蔬菜批发价格的平均值为 4.46。

(四) 数据处理

1) 缺失值处理

鉴于预测的目标是 10 年期国债到期收益率(YTM10), 首先, 必须确保该预测对象数据的真实性和连续性; 其次, 鉴于 YTM10 与一年期国债到期收益率(YTM1)之间存在高度相关性, 本文以 10 年期国债收益率的发布时点为基准, 剔除了未发布数据的时间点(一年期和十年期国债收益率的发布时点相同)。然而, 在剔除这些数据后, 其他变量仍存在一些缺失值。为了解决这一问题, 本文考虑使用各变量的样本平均值进行填充, 借助样本自身的特征来填补缺失值, 以防止产生更多的虚假数据而干扰预测结果的准确性。

2) 数据标准化

由于数据集中各个指标的维度和单位各不相同, 这可能会对研究和模型识别造成干扰。为了解决这种因维度差异引起的问题, 本文首先对数据集进行了预处理, 采用标准化方法以便于各指标之间的比较和统一。的具体公式如下:

$$x_{\text{标准化}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

3) 数据集划分

实验所用数据涵盖了中债十年期国债到期收益率、中债一年期国债到期收益率以及基本面、资金面等共 9 个变量, 数据的时间跨度为 2015 年 1 月 1 日至 2024 年 12 月 18 日, 用于模型的训练与测试。本文以 8:2 的比例将数据集划分为训练集和测试集。

(五) 实证结果分析

1) LSTM 模型

在构建 LSTM 网络模型时, 需预先确定若干关键参数以启动网络训练过程。这些参数涵盖了 LSTM 层的单元数量(units)、学习率、窗口大小(Window_size)、迭代轮次(epoch 次数)、训练集与测试集的划分比例以及输入特征的数量(fea_num)等, 参数的具体设定见表 3。

Epoch 指的是所有训练集数据完整地通过神经网络, 从输入输出到完成训练反向传播的整个过程, 即训练数据被完整训练的次数。从理论上来说, epoch 参数设置得越高, 训练集被完整训练的次数就越多, 模型的学习效果也会更好。然而, 过多的 epoch 次数可能会引发过拟合问题, 导致训练好的模型泛化能力变弱, 虽然在训练集上表现出色, 但在测试集上会出现较大误差。依据机器学习参数设置的经验, epoch 的大小应与数据集的特征相关联。对于越复杂的数据集, 应先尝试设置较大的 epoch 值。针对规模过大的数据集, 为提升训练效率和优化速度, 通常会设置 batch_size 参数。该参数表示一次完整训练中输入的样本数量, 相当于把完整的数据集进行拆分, 分批次输入网络。

LSTM 层的 unit 单元数量方面, LSTM 网络在每个时间步都有细胞 cell, 其中前馈神经网络的神经元个数就是 unit。unit 数量越多, LSTM 的记忆能力越强, 但同时也会占用更多的资源。

关于 Window_size 的选取, 通常没有固定的标准。一般可在 7~16 这个范围内选择, 偏大的值可在 24、48、96 中选择。在本文中, 首先参考与数据集相关的论文, 然后通过对比试验进行测试, 最后确定合适的 Window_size。

在 LSTM 层数的设置上, 相关资料表明, 层数并非越多越好。对于三层以上的 LSTM 神经网络, 容易出现梯度消失问题, 导致参数更新缓慢, 网络迭代速度几乎降为 0, 还有可能陷入局部极小值。因此, 本文考虑先采用两层 LSTM 构建模型, 以期获得最佳的预测结果。

具体参数设置详见下文表 3:

Table 3. Parameters of the LSTM model structure
表 3. LSTM 模型结构设定参数

参数	值
Lstm1_units	80
Lstm2_units	100
Learn_rate	0.0001
Window_size	6
batch_size	30
epoch	70
fea_num	9

观察图 4(a)中的 LSTM 模型损失函数图, 可以发现随着训练轮次(epoch)的递增, 模型的训练损失快速降低并接近于 0。同时, 测试集上的损失也呈现出缓慢下降的趋势, 并在较低水平上保持稳定。这一现象表明模型没有出现拟合的情况, 其学习过程处于正常状态。原始数据与预测值的效果对比如图 4(b)。

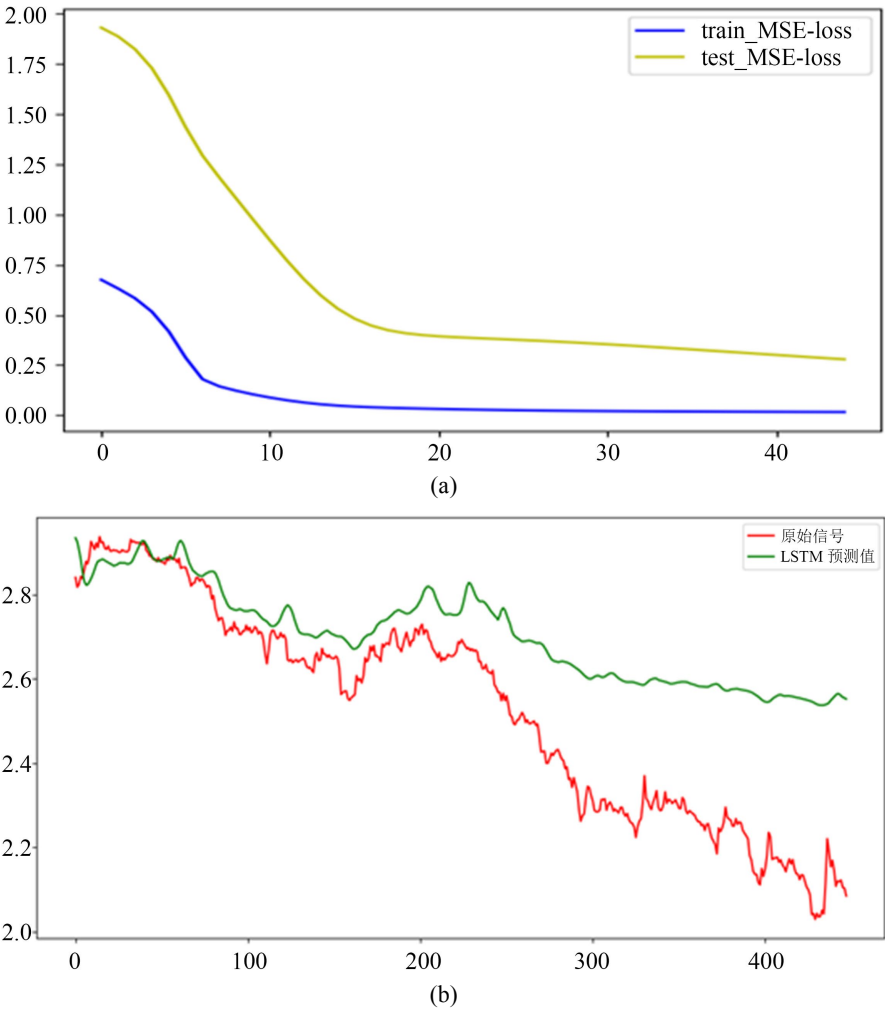


Figure 4. Loss function and prediction effect diagram of the LSTM model; (a) Loss function diagram; (b) Prediction effect diagram
图 4. LSTM 模型的损失函数及预测效果图; (a) 损失函数图; (b) 预测效果图

2) CNN-LSTM-Attention 模型

与传统的 LSTM 模型相比, CNN-LSTM-Attention 模型在数据输入阶段引入了卷积神经网络(CNN)作为前置处理。该 CNN 通过卷积操作提取时序数据在各个时间点上的空间与时间分布特征,随后将这些特征输入到 LSTM 层中以捕捉序列数据的长期和短期记忆信息。在此基础上,模型进一步融入了注意力机制(AM)层,以更精准地筛选关键信息并进行预测。因此,相较于原始的 LSTM 模型, CNN-LSTM-Attention 模型不仅能够获取特征数据之间的关联性和空间分布信息,即相当于对时序数据的横截面特征进行捕捉,还能够把握特征数据自身随时间的变化趋势。这使得模型能够更全面、更有针对性地提取信息,从而提升预测的准确性。由于在 CNN-LSTM-Attention 模型里引入了卷积神经网络与注意力机制,相较于单一模型,需要设定的参数更多。其中涵盖了卷积核数量 filters、卷积核大小 kernal_size、卷积步长 strides,以及池化方法等。模型的主要参数设定情况见表 4。

在卷积网络中,卷积核数量(filters)等同于通道数量。不同的卷积核拥有不同的权重,这使得它们能够提取出多样化的特征。卷积核数量越多,所能提取的特征类型也就越丰富。不过,对于目标对象越复杂的情况,虽然需要更多的卷积核来进行特征提取,但同时也必须考虑到算力的限制。在本文中,卷积核数量首先设置为较为常见的 64 个。

卷积核大小(kernel_size)通常会选择奇数尺寸。这是因为在卷积平移运算过程中,为了避免边缘信息的损失,会采用填充的方式,即在原数据边缘补零,这样提取的特征就能包含边缘数据的信息,保证卷积前后数据尺寸一致。此外,奇数尺寸的卷积核更便于确定中心位置。

卷积步长(strides)指的是卷积过程中的平移步长。为了更细致地过滤特征信息,本文选择将卷积步长设置为 1,以实现小步平移。

预测结果见图 5,可见,不管是前面的单一模型,还是现在的集合模型,预测结果与实际值在后续都出现较大程度的偏离,预测值仅仅在趋势上与实际值保持了一致,它仅保留了变化的相对特征,而无法把握绝对的变动幅度。

Table 4. Set parameters of the CNN-LSTM-Attention model
表 4. CNN-LSTM-Attention 模型设定参数

参数	值
filters	64
kernal_size	3
strides	1
LSTM1_units	80
LSTM2_units	100
Learn_rate	0.0003
pooling_method	maxpooling
Window_size	6
batch_size	30
epoch	70

3) 模型评价

模型具体评价指标包括 RMSE、MSE、MAE,同时也可以作为模型进行权重调整的损失函数,本文在建立神经网络模型时统一选取 MSE 作为损失函数。单一 LSTM 模型与集合 CNN-LSTM-Attention 模型

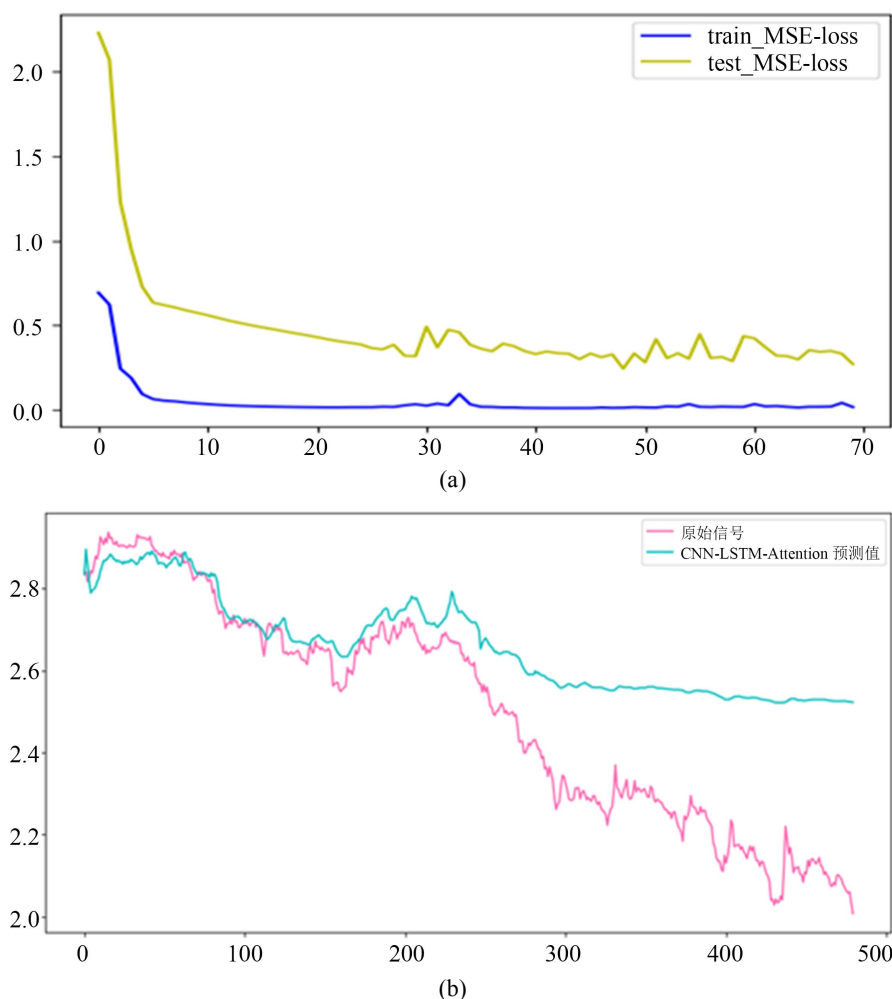


Figure 5. Loss function and prediction effect diagram of the CNN-LSTM-Attention model; (a) Loss function diagram; (b) Prediction effect diagram

图 5. CNN-LSTM-Attention 模型的损失函数及预测效果图; (a) 损失函数图; (b) 预测效果图

各自对应的评价指标值见表 5, 从本文得到的评价指标值来看, 虽然相差不大, 但相比于单一 LSTM 模型, 集成模型仍表现出较好的预测效果。可知在预测精度上集成模型要好于基础模型。

MSE (Mean Squared Error)均方误差, 以真实值和预测值的差值求平方之后进行算术平均。

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\bar{y}_i - y_i)^2$$

RMSE (Root Mean Squared Error)均方根误差, 表示预测值和观测值之间差异(即残差)的样本标准差。

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\bar{y}_i - y_i)^2}$$

MAE (Mean Absolute Error)平均绝对值误差, 表示预测值和观测值之间绝对误差的平均值。

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\bar{y}_i - y_i|$$

其中 $\bar{y}_i$ 为预测值, 为 y_i 真实值。以上两种指标越小表明预测值越接近真实值, 证明模型性能越好、特征表达能力越强。

Table 5. Comparative analysis of empirical results
表 5. 实证结果比较分析

模型	MSE	RMSE	MAE
LSTM	0.2789	0.5281	0.4156
CNN-LSTM-Attention	0.2703	0.5199	0.3896

5. 结论

本文使用了一个比较新的模型 CNN-LSTM-Attention 来预测十年期国债收益率, 并将此模型的效果与基础模型 LSTM 的预测效果进行比较。根据实证结果, 从 MSE、RMSE、MAE 这几方面来进行评价, 集成模型的预测效果优于单一的 LSTM 模型。由于 CNN-LSTM-Attention 模型对各种超参数的敏感性和高复杂性, 很有可能通过附加优化配置和特征工程来进一步改善预测能力。所以本文未来的研究方向集中在对模型参数的调节上, 并且考虑在特征变量方面进行筛检和增加, 来增加模型的预测能力。

参考文献

- [1] 杨华. 关于国债收益率曲线作为定价基准的分析及建议[J]. 财务与会计, 2022(10): 85-86.
- [2] 陈哲. 宏观经济变量预测利率期限结构变动的探讨[J]. 商业时代, 2008(32): 68-69.
- [3] 帅昭文, 沈根祥, 张碧馨. 宏观经济对国债利率期限结构的影响研究[J]. 统计与决策, 2017(17): 159-164.
- [4] 杨婉茜, 成力为. 基于贝叶斯向量自回归的中国国债收益率预测[J]. 统计研究, 2015, 32(8): 69-76.
- [5] 谈正达, 霍良安. 无套利 Nelson-Siegel 模型在中国国债市场的实证分析[J]. 中国管理科学, 2012, 20(6): 18-27.
- [6] 赵晶, 张洋, 丁志国. 利率期限结构特征的拟合与预测[J]. 统计研究, 2015, 32(2): 83-89.
- [7] Fama, E.F. and Bliss, R.R. (1987) The Information in the Long-Maturity Rates. *American Economic Review*, 77, 680-692.
- [8] Brandt, M.W. and Wang, K.Q. (2003) Time-Varying Risk Aversion and Unexpected Inflation. *Journal of Monetary Economics*, 50, 1457-1498. <https://doi.org/10.1016/j.jmoneco.2003.08.001>
- [9] 闫红蕾, 张自力. 利率期限结构预测、国债定价及国债组合管理[J]. 统计研究, 2018, 35(3): 23-37.
- [10] 宋鹏, 张淼. 国债收益率预测的 VAR-LSTM 框架[J]. 统计与决策, 2021, 37(5): 148-152.
- [11] 王雷, 闫红蕾, 张自力. 收益率曲面预测及其在信用债投资组合管理中的应用[J]. 统计研究, 2021, 38(4): 145-160.
- [12] 康书隆, 王志强. 中国国债利率期限结构的风险特征及其内含信息研究[J]. 世界经济, 2010, 33(7): 121-143.
- [13] 袁伟鹏, 张翼, 张飞, 等. 新常态下 PPI 与十年期国债收益率的动态关系研究[J]. 工程数学学报, 2021, 38(6): 809-821.
- [14] 魏雪梅. 国债收益率影响因素分析及走势预测[J]. 债券, 2013(9): 52-57.
- [15] 李湛, 唐晋荣. 股市异常波动下的股债联动关系——基于事件研究法的分析[J]. 证券市场导报, 2017(12): 24-29, 37.