

基于LSTM和GRU神经网络的上海原油期货 价格预测研究

徐瑞敏, 贝乐敏*, 王舒琪, 徐星瑶, 何佳琪

嘉兴南湖学院现代金融学院, 浙江 嘉兴

收稿日期: 2025年8月25日; 录用日期: 2025年9月19日; 发布日期: 2025年9月26日

摘要

本研究针对上海原油期货价格预测问题, 系统对比了ARIMA、VARIMA、LSTM与GRU四类模型的预测性能。实证结果表明: 传统线性模型ARIMA和VARIMA因受限于线性假设, 在测试集上预测误差显著 ($RMSE > 9.5$), 尤其对突发性波动事件响应滞后; 而深度学习模型展现出显著优势, 其中GRU以双门控耦合机制实现效率与精度的最优平衡 ($MAE = 4.6928$, 参数量较LSTM减少25.6%), LSTM则凭借三门分立结构在长期趋势捕捉中表现稳健 ($R^2 = 0.9647$)。消融实验进一步验证了门控设计的必要性——移除LSTM遗忘门导致误差激增50%, 印证了其在噪声过滤中的核心作用。研究成果为油价预测提供了兼具精度与效率的深度学习解决方案, 并为构建智能化风控体系提供了理论依据。

关键词

ARIMA, VARIMA, LSTM, GRU, 原油期货, 价格预测

Forecasting Shanghai Crude Oil Futures Prices with LSTM and GRU Neural Networks

Ruimin Xu, Lemin Bei*, Shuqi Wang, Xingyao Xu, Jiaqi He

School of Modern Finance, Jiaxing Nanhu University, Jiaxing Zhejiang

Received: August 25, 2025; accepted: September 19, 2025; published: September 26, 2025

Abstract

This study systematically compares the predictive performance of four models—ARIMA, VARIMA, LSTM, and GRU—for forecasting Shanghai crude oil futures prices. Empirical results demonstrate

*通讯作者 Email: lemin.bei[at]jxnhu.edu.cn

that traditional linear models (ARIMA/VARIMA), constrained by their linear assumptions, exhibit significant prediction errors on the test set ($RMSE > 9.5$), particularly showing delayed responses to abrupt volatility events. In contrast, deep learning models demonstrate remarkable advantages: GRU achieves an optimal balance between efficiency and accuracy through its dual-gate coupling mechanism ($MAE = 4.6928$, with 25.6% fewer parameters than LSTM), while LSTM exhibits robust performance in capturing long-term trends owing to its triple-gate structure ($R^2 = 0.9647$). Ablation experiments further validate the necessity of gating design—removing the forget gate from LSTM increases errors by 50%, confirming its critical role in noise filtering. The research outcomes provide a deep learning solution that combines accuracy and efficiency for oil price forecasting, offering a theoretical foundation for constructing intelligent risk management systems.

Keywords

ARIMA, VARIMA, LSTM, GRU, Crude Oil Futures, Price Prediction

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. 研究背景

原油作为现代工业体系的命脉，其重要性贯穿于社会生产生活的各个层面。从化学组成来看，原油是由数百种碳氢化合物构成的复杂混合物，经过蒸馏、裂解等工艺可分离出汽油、柴油、航空煤油等动力燃料，以及沥青、润滑油、石化原料等工业必需品。中国期货业协会在[1]中披露，全球日均原油消费量突破 1 亿桶，其中交通运输领域占比达 65%，化工原料消耗占 12%，发电与供热占 10%，构成了支撑现代社会运转的能源网络。就产业关联度而言，每万吨原油可直接带动 1.2 亿元 GDP 增长，并通过产业链传递产生 6~8 倍乘数效应。在我国经济结构中，石油化工产业贡献了 4.3% 的 GDP 总量，直接关联 3000 余种工业产品。其战略地位在 2022 年俄乌冲突导致的能源危机中尤为凸显，当国际油价突破 120 美元/桶时，我国战略石油储备释放规模达到历史性的 1.8 亿桶，充分印证了石油安全对国家经济韧性的关键作用。原油价格的剧烈波动对宏观经济稳定构成持续性威胁，其传导机制具有多维度、非线性的特征。从产业传导路径观察，2020 年 WTI 原油期货史无前例地跌至-37 美元/桶，导致全球油气勘探投资缩减 42%，直接造成 50 万从业人员失业。价格信号沿产业链向下游传递时，炼化企业面临库存减值与需求萎缩的双重挤压，典型如韩国 SK Innovation 当年炼油业务亏损达 18 亿美元。

建立有效的价格风险管控体系已成为各国维护经济安全的必然选择，而期货市场在其中发挥着核心作用。上海原油期货上市五年日均成交量突破 25 万手，标志着亚太地区价格基准地位初步形成。在复杂多变的金融市场中，准确的价格预测一直是风险管理和投资分析的重点。面对受重大公告、社交推文、公司新闻等情绪指标影响的市场，基本面和技术面分析表现不佳。传统时间序列模型，例如自回归整合移动平均(AutoRegressive Integrated Moving Average, ARIMA, Box 和 Jenkins 于[2]中系统总结了该方法)，在捕捉非线性结构时常常失效([3]和[4])。人工智能模型凭借从海量数据中学习并持续改进的能力，为更准确地预测未来价格提供了支撑，例如[5]。这些模型包括经典的机器学习算法如支持向量机(Support Vector Machine, SVM, 由[6]提出, [7]为其最新应用)和随机森林(Random Forests, RF, 由[8]提出, [9]为其最新应用)，以及深度学习算法如循环神经网络(Recurrent Neural Networks, RNN, 见[10]和[11])和卷积神经网络(Convolutional NN, CNN, 见[12]和[13])。

多年来，深度学习技术大踏步发展，包含长短时记忆(Long Short-Term Memory, LSTM, [14]和[15])

和门控循环单元(Gated Recurrent Unit, GRU, [16]和[17])及其变体的循环神经网络架构成为预测金融市场价格模式的流行方法。但是运用深度学习方法针对上海原油期货价格的预测尚显不足。本研究将 ARIMA 与 VARIMA 作为基准参照,核心聚焦上海原油期货价格并应用 LSTM 和 GRU 填补该空白,以提升国内原油期货价格的预测精度,最终辅助风险管理者和投资者做出数据驱动的最优决策。

2. 移动平均

2.1. ARIMA 模型

移动平均的思想最早可追溯至二十世纪初英国统计学家 Yule 对太阳黑子周期的研究,而真正成形则要归功于[18]与[19]对随机过程分解的开创性工作:任何协方差平稳序列都可以被唯一地表示为当前及历史白噪声的加权平均,即移动平均过程。这一结论不仅为时间序列建模奠定了概率基石,也为后续统一自回归移动平均框架提供了理论支点[2]。

在油价建模语境下,移动平均的核心作用在于将不可观测的市场冲击(如突发地缘政治事件、炼厂意外检修等)视为“新息”序列,并假设当前价格是对过去所有新息的累积响应。通过对新息结构的刻画,将序列的自相关模式转化为可估计的参数,其与自回归项的结合并进一步差分整合则构成了 ARIMA (p, q, d) , 数学表述为:

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1-L)^d y_t = \mu + \left(1 + \sum_{j=1}^q \theta_j L^j\right) \varepsilon_t. \quad (1)$$

其中 L 为滞后算子, ϕ_i 、 θ_j 分别为自回归 AR 第 i 项的系数以及移动平均 MA 第 j 项的系数, μ 为常值漂移项, ε_t 为白噪声。 p 为自回归项数, q 为移动平均项数, d 为差分次数。现假定 $\mu=0$ 。自回归部分 $\left(1 - \sum_{i=1}^p \phi_i L^i\right) y_t$ 意为当前值 y_t 由过去 p 个历史值的线性组合决定,作用是捕捉时间序列的长期趋势(如油价持续上涨)。当 $p=1$ 时, $y_t = \phi_1 y_{t-1} + \varepsilon_t$, 表示当前值仅依赖于前一时刻的值。差分整合部分 $(1-L)^d y_t$ 通过 d 阶差分将非平稳序列转化为平稳序列,解决非平稳性问题(如原油价格的季节性波动)。一阶差分 $(1-L)y_t = y_t - y_{t-1}$, 消除线性趋势;二阶差分进一步消除曲率趋势。对于移动平均部分 $\left(1 + \sum_{j=1}^q \theta_j L^j\right) \varepsilon_t$, 当前值由过去 q 个预测误差的线性组合修正得来,用来平滑随机扰动(如突发事件对油价的瞬时冲击)。当 $q=1$ 时, $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$ 。

2.2. VARIMA 模型

然而现实经济与金融系统常受到多变量的交互影响,如油价同时受成交量、涨跌幅、地缘事件等多维因子的驱动,传统 ARIMA 因仅依赖单一历史序列故无法捕捉变量间的协同效应与反馈机制。为突破局限,学者将 ARIMA 扩展至向量自回归整合移动平均模型(VARIMA)。该模型由 Sims 的向量自回归 (VAR)理论演化而来[20],通过引入系数矩阵 Φ_i 和 Θ_j 替代标量参数 ϕ_i 和 θ_j ,将单变量的差分、自回归与移动平均推广至多维空间,详见[21]。其核心形式为:

$$y_t = \mu + \sum_{i=1}^p \Phi_i y_{t-i} + \sum_{j=1}^q \Theta_j \varepsilon_{t-j} + \varepsilon_t. \quad (2)$$

其中 y_t 为向量形式的内生变量,并利用[22]进行协整检验。如果存在协整,则按[23]以及[24]进行差分,转化为向量误差修正模型(Vector Error Correction Model, VECM)进行后续估计和检验。

3. 神经网络

3.1. LSTM 神经网络

Mikolov 等人首次将循环神经网络(RNN)应用于自然语言处理[25],其基本结构即为基本的隐变量自

回归模型。RNN 模型在每一个时间步都进行隐变量计算，并基于计算得到的隐变量对本时间步的输出进行预测。对于每一个时间步，RNN 的隐变量与上一个时间步使用相同的定义，结合当前时间步的输入信息循环地计算新的隐变量。于是基于循环计算的隐状态神经网络被命名为循环神经网络。

长短时记忆(LSTM)是 RNN 的一种特殊形式。它通过引入记忆细胞(memory cell)及其内部控制机制，即遗忘门(forget gate)、输入门(input gate)和输出门(output gate)，把一个单元细胞状态变为两个细胞状态之间的复杂互动，并通过激活函数(如 sigmoid 函数)对信息流进行筛选，从而更好地捕捉时间序列中的长期依赖关系，避免短时记忆、梯度爆炸、梯度消失问题。

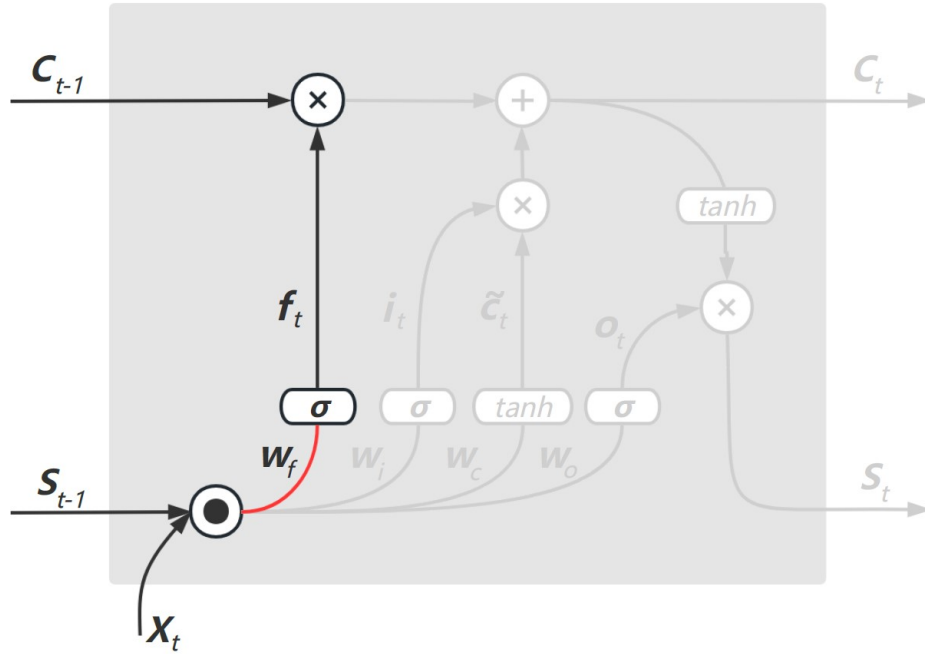


Figure 1. The forget gate in the internal structure of LSTM

图 1. LSTM 内部结构之遗忘门

遗忘门负责对原有的长期信息进行选择性遗忘，见图 1。上一期隐藏状态的输出值 S_{t-1} 与当期样本观测值 X_t 经权重参数 W_f 和偏置 b_f 拼接，汇入 LSTM 架构。遗忘门中全连接层的激活函数为 sigmoid 函数，取值范围为 $(0,1)$ ，

$$f_t = \sigma(X_t W_{xf} + S_{t-1} W_{sf} + b_f) \quad (3)$$

对于长期记忆细胞 C_{t-1} ， $f_t \odot C_{t-1}$ 为其信息留存量，其中 $\odot$ 是两个相同维度的矩阵对应位置元素的 Hadamard 乘积。 f_t 越接近 0 表示遗忘越多，接近 1 表示信息大多数得以保留。注意，通过遗忘门，我们对 C_{t-1} 进行筛选，而非对 S_{t-1} 或 X_t 筛选。

输入门决定让多少新的短期信息进入长期记忆，见图 2。首先 S_{t-1} 与 X_t 通过 tanh 函数，将其每个元素的激活值保留在区间为 $(-1,1)$ 的候选记忆细胞 $\tilde{C}_t$ 内：

$$\tilde{C}_t = \tanh(X_t W_{xc} + S_{t-1} W_{sc} + b_c). \quad (4)$$

我们再用与遗忘门相似的门控输入结构 i_t 真正地对 S_{t-1} 和 X_t ，亦即对短期记忆 $\tilde{C}_t$ ，提取信息：

$$i_t = \sigma(X_t W_{xi} + S_{t-1} W_{si} + b_i). \quad (5)$$

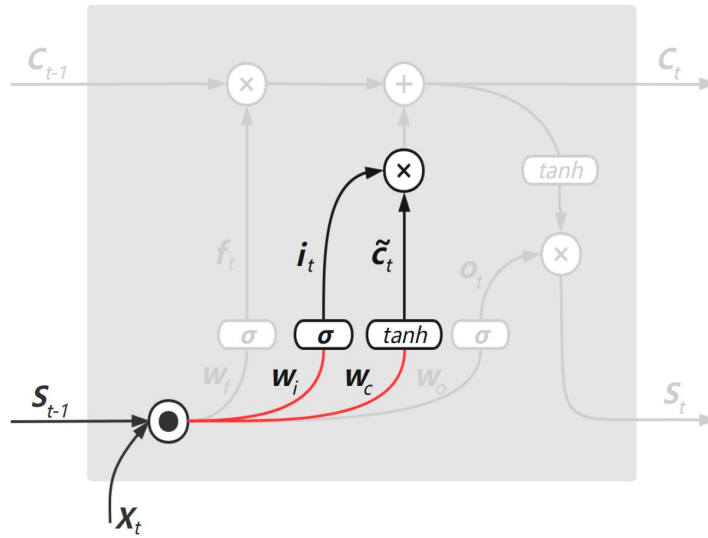


Figure 2. The input gate in the internal structure of LSTM

图 2. LSTM 内部结构之输入门

对于候选记忆细胞 $\tilde{C}_t$ ， $i_t \odot \tilde{C}_t$ 为其信息留存量。 i_t 越接近 0 表示舍去信息越多，接近 1 表示信息大多数可以输入到长期记忆 C_t 中。所以 C_t 里 $f_t \odot C_{t-1}$ 为经过遗忘后长期留存的记忆， $i_t \odot \tilde{C}_t$ 为输入的短期新增记忆：

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t. \quad (6)$$

可见门控结构 $f_t \odot$ 和 $i_t \odot$ 皆对记忆 C_{t-1} 和 $\tilde{C}_t$ 进行某种修正。虽然它们内部结构相似，但经过数据的学习，权重 W_{xf} 、 W_{sf} 、 W_{xi} 、 W_{si} 以及更进一步 W_{xc} 和 W_{sc} ，偏置 b_f 和 b_i 以及更进一步 b_c 都能够呈现出不同的数值，因而体现出不同程度的控制和拼接。

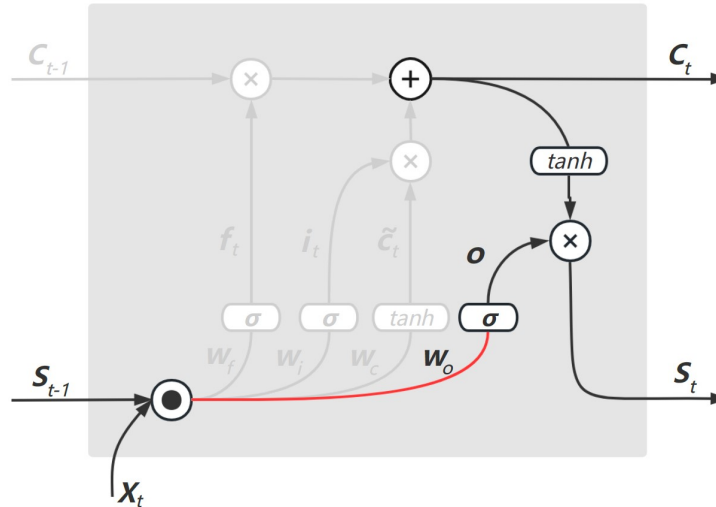


Figure 3. The output gate in the internal structure of LSTM

图 3. LSTM 内部结构之输出门

输出门控制着当前记忆细胞 C_t 里有多大比例的信息可以输出到短期记忆 S_t 中，见图 3，

$$S_t = o_t \odot \tanh(C_t). \quad (7)$$

输出门的工作原理类似遗忘门和输入门:

$$o_t = \sigma(X_t W_{xo} + S_{t-1} W_{so} + b_o). \quad (8)$$

纵观整个 LSTM 架构(图 4), 隐藏状态 S_t 虽积累了截至 t 时刻的信息, 但它在每个时刻都会有更新, 更加敏感于最近的输入和输出, 尤其是上一期 S_{t-1} 和当期事件 X_t , 因此携带的是短期记忆。细胞 C_t 则拥有长期记忆。从 C_{t-1} 到 C_t 直接传播, 并用遗忘和输入开关来删除或添加信息, 是 LSTM 里的主线。由于 LSTM 中的长期记忆通过门控进行“累加”, 而 RNN 是“累乘”, 因此 LSTM 可以避免梯度消失或梯度爆炸问题, 有利于保留长期信息。

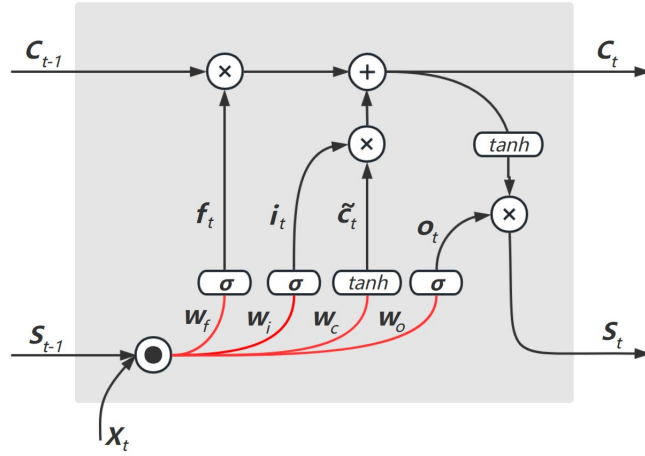


Figure 4. The architecture of LSTM

图 4. LSTM 架构

3.2. GRU 神经网络

门控循环单元(GRU)是 LSTM 的一个简化版本, 能够提供与 LSTM 同样的效果, 且参数更少, 训练更快, 见图 5。它只有重置门(reset gate)和更新门(update gate):

$$r_t = \sigma(X_t W_{xr} + S_{t-1} W_{sr} + b_r), \quad (9)$$

$$u_t = \sigma(X_t W_{xu} + S_{t-1} W_{su} + b_u), \quad (10)$$

其中 W_{xr} 、 W_{sr} 、 W_{xu} 、 W_{su} 为权重参数, b_r 和 b_u 为偏置参数。候选隐藏状态 $\tilde{S}_t$ 是上一期隐藏状态 S_{t-1} 与当前输入 X_t 的信息合成, 其中重置门 r_t 控制 S_{t-1} 参与信息合成的比例, 亦即该门决定会有多少历史数据被遗忘并重置, 因而也决定了 $\tilde{S}_t$ 中短期信息 X_t 的含量:

$$\tilde{S}_t = \tanh(X_t W_{xs} + (r_t \odot S_{t-1}) W_{ss} + b_s). \quad (11)$$

更新门 u_t 则通过调节历史数据 S_{t-1} (上期输出)与候选状态 $\tilde{S}_t$ (本期输入)的权重, 控制序列中长期信息的比例:

$$S_t = (1 - u_t) \odot S_{t-1} + u_t \tilde{S}_t. \quad (12)$$

GRU 仅用更新门 u_t 就实现了 LSTM 中遗忘门与输入门之双重控制, 减少了参数数量。其次 GRU 将记忆单元归并入隐含状态中, 结构较 LSTM 更为简化, 计算收敛速度更快。

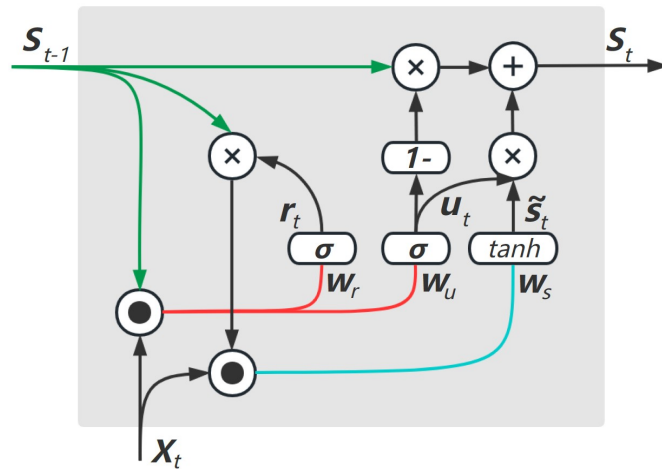


Figure 5. The architecture of GRU

图 5. GRU 架构

4. 数据代码

4.1. 数据来源

本文中的上海原油期货价格数据取自东方财富 Choice 金融终端(亦可从上海国际能源交易中心官网下载,但需要将各期货连续化处理),证券代码 SCFI,证券名称原油加权。数据集包括交易时间、开盘价、最高价、最低价、收盘价、前一天的收盘价、涨跌值、涨跌幅、成交量、成交额共 10 列数据。除 ARIMA 仅采用收盘价外,VARIMA、LSTM 和 GRU 均使用全部 10 列数据。时间横跨 2018 年 3 月 26 日至 2025 年 2 月 28 日,共计 1680 笔数据。我们将该数据集的前 80%作为训练集,将剩下的 20%作为测试集,来检验模型的泛化能力。由于各指标变量数值量纲差异极大,我们通过归一化将它们都压缩到[0,1]之间,

$$x_i = \frac{X_i - \min X_i}{\max X_i - \min X_i}. \quad (13)$$

上式中 X_i 为需要被标准化处理的原始值。实验最后运用反归一化措施,将输出数据的量纲统一到输入数据上:

$$\hat{Y}_i = y_i \times (\max X_i - \min X_i) + \min X_i. \quad (14)$$

4.2. 代码概述

ARIMA(p, q, d) 模型通过 Python 的 statsmodels 库构建,将公式(1)转化为可执行的程序。用 Augmented Dickey-Fuller(ADF)检验平稳性,若非平稳(p 值 ≥ 0.05),则执行差分处理,直至序列通过检验,由此确定 d 值。随后,利用自相关(ACF)和偏自相关(PACF)图诊断 p 和 q 参数,并通过网格搜索最小化 Akaike 信息准则(AIC)以优化(p, d, q)组合,本文得到(p, d, q)=(5,1,4)。模型训练阶段用 ARIMA(5,1,4)拟合历史数据,估计出系数 $\{\phi_i\}_{i=1}^p$ 和 $\{\theta_j\}_{j=1}^q$ 。

对于 VARIMA,通过 Johansen 协整检验判断系统是否存在长期均衡关系:若存在协整(rank > 0, 用本文数据 rank = 7,表示存在 7 个独立的长期均衡关系),则采用 VECM,其核心形式为

$\Delta y_t = \alpha + \Gamma_0 y_{t-1} + \sum_{i=1}^p \Gamma_i y_{t-p+1} + \epsilon_t$; 否则对系统施加 d 阶差分($d \in \{0, 1, 2\}$)实现平稳化后,采用标准 VAR 模型或含移动平均项的 VARIMA(p, d, q)。在训练阶段通过 AIC 自动选择最优滞后数 p ,本文得到 $p=1$ 。采用滚动时间窗口(rolling window)方式进行样本外预测,每次将新一期的真实值 y_t 纳入训练集,重新拟

合模型以预测下一期值 y_{t+1} ，避免前视偏差(look-ahead bias)。

LSTM 与 GRU 则置于 Keras 与 Tensorflow 环境中。LSTM 模型包含两层 LSTM 单元(64 和 32 个神经元)和 dropout 层(0.2 的丢弃率)，通过 sequential 堆叠实现从输入到输出的端到端学习；GRU 模型则用 GRU 层替代 LSTM 层，结构更简洁。两者均使用 Adam 优化器和均方误差损失函数，不同于传统极大似然估计，强调通过梯度反向传播自适应优化权重。训练过程中引入早停机制(early stopping)，监控验证损失，并在连续 10 轮未改善时终止训练，防止过拟合，类似于 ARIMA 中的 AIC 准则优化，但更动态化。最后，可视化模块绘制预测值与实际值的对比曲线，直观展示模型性能。

4.3. 评估标准

对测试集上模型的评价，参照深度学习领域常用的指标，选取平均绝对误差(Mean Absolute Error, MAE)、均方根误差(Root Mean Square Error, RMSE)、平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)、决定系数(Coefficient of Determination, R^2):

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_i - \hat{Y}_i|, \tag{15}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2}, \tag{16}$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100\%, \tag{17}$$

$$R^2 = \frac{\frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2}. \tag{18}$$

式中 Y_i 、 $\hat{Y}_i$ 、 $\bar{Y}$ 的意思显见。

5. 实验分析

5.1. 模型比较

先看表 1、图 6 和图 7。在模型性能的比较分析中，ARIMA 与 VARIMA 的表现揭示了传统时间序列方法的内在特性与局限。由图 6 和图 7 知，两种模型在价格平稳阶段展现出较好的一致性，但在市场出现快速转折或受外部事件冲击时(如 2024 年三季度 OPEC + 意外减产)，预测轨迹明显滞后于实际价格变动，表现出平滑过度、难以捕捉突变的特征。表 1 中的数据进一步量化了该现象：ARIMA 的 MAE 为 7.1859，RMSE 为 9.5128，MAPE 为 1.2486%；VARIMA 虽在 MAE (7.1544)和 MAPE (1.2447%)上略有改善，但其 RMSE (9.5728)反而略高于 ARIMA，说明引入多变量并未带来预测精度上的显著提升，反而可能在波动加剧时因变量间交互的复杂性带来额外的不确定。这一结果印证了线性模型

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

在处理金融市场非线性、非平稳序列时所固有的不足。

Table 1. Model performance comparison
表 1. 模型性能对比

评估指标	ARIMA	VARIMA	LSTM	GRU	改进方向
MAE	7.1859	7.1544	5.8159	4.6928	越小越好
RMSE	9.5128	9.5728	7.5017	6.4125	越小越好
MAPE	1.2486	1.2447	0.9975	0.8122	越小越好
R^2	0.9441	0.9434	0.9647	0.9742	越接近 1 越好

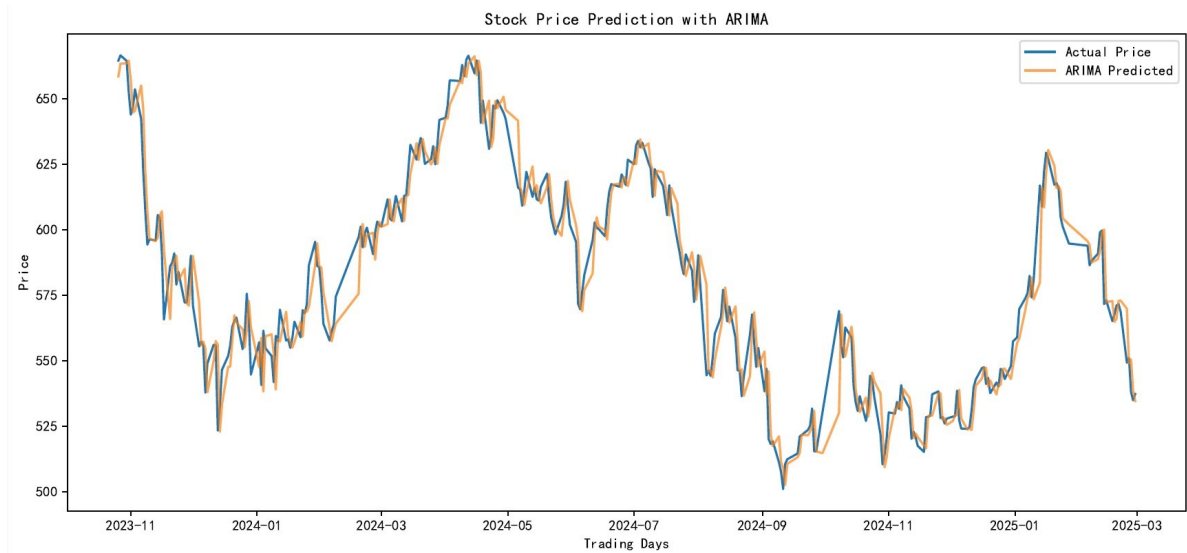


Figure 6. Actual observed prices and ARIMA predicted prices

图 6. 真实观测值与 ARIMA 预测值

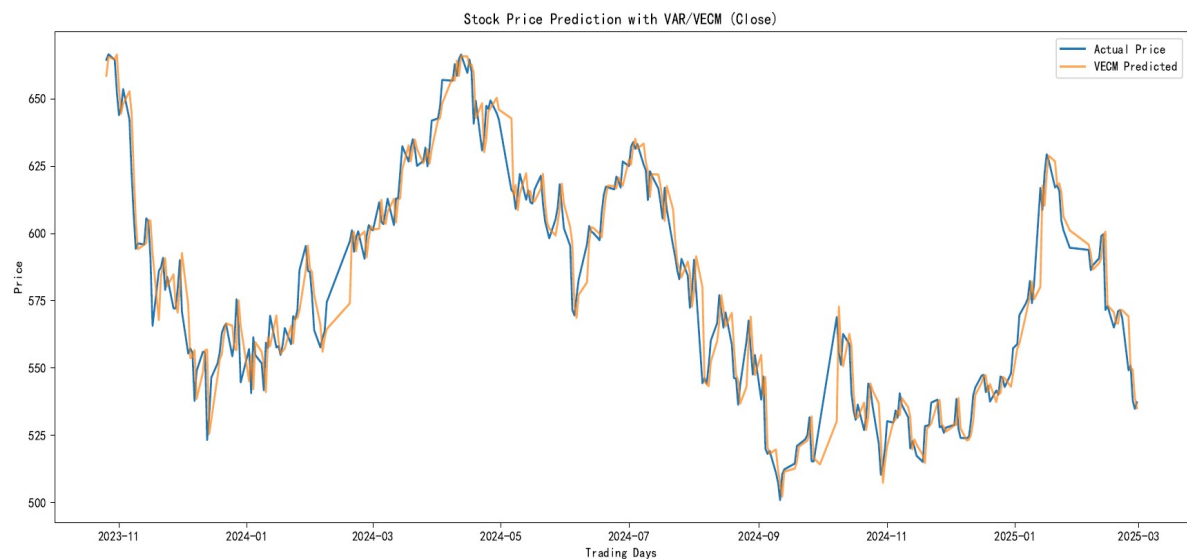


Figure 7. Actual observed prices and VARIMA/VECM predicted prices

图 7. 真实观测值与 VARIMA/VECM 预测值

相比之下，深度学习模型 LSTM 与 GRU 则展现出更优异的适应性与预测精度。从预测曲线来看，两者均能更敏锐地响应价格的快速变化，尤其在 2024 年 11 月美联储议息会议等关键事件点附近，相较于 LSTM (图 8)，GRU 的预测轨迹(图 9)更贴近实际值，滞后显著减少。

表 1 中的性能指标为这一视觉优势提供了有力支持：GRU 的 MAE 降至 4.6928，RMSE 为 6.4125，MAPE 仅为 0.8122%，其各项指标均全面优于 LSTM (MAE: 5.8121, RMSE: 7.5024, MAPE: 0.9921%)。这种性能提升可归因于 GRU 的门控机制设计，其重置门 r_t 与更新门 u_t 以更少的参数量实现了与 LSTM 三门结构相近的功能，公式(12)： $S_t = (1 - u_t) \odot S_{t-1} + u_t \tilde{S}_t$ 本质上是一种门控耦合，增强了模型在训练数据有限时的泛化能力，降低了过拟合风险。与 ARIMA 和 VARIMA 相比，深度学习模型的核心优势在于其无

需预设线性形式，能够通过多层非线性变换从历史数据中自动学习并逼近复杂的动态模式，从而在高度非线性的市场环境中获得更稳健的预测表现。

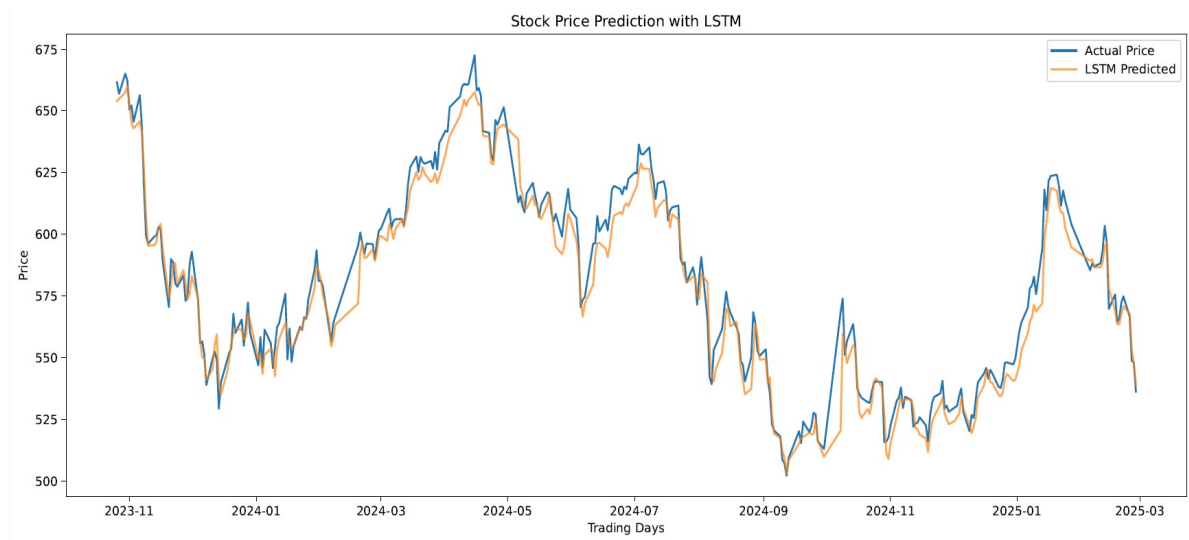


Figure 8. Actual observed prices and LSTM predicted prices
图 8. 真实观测值与 LSTM 预测值

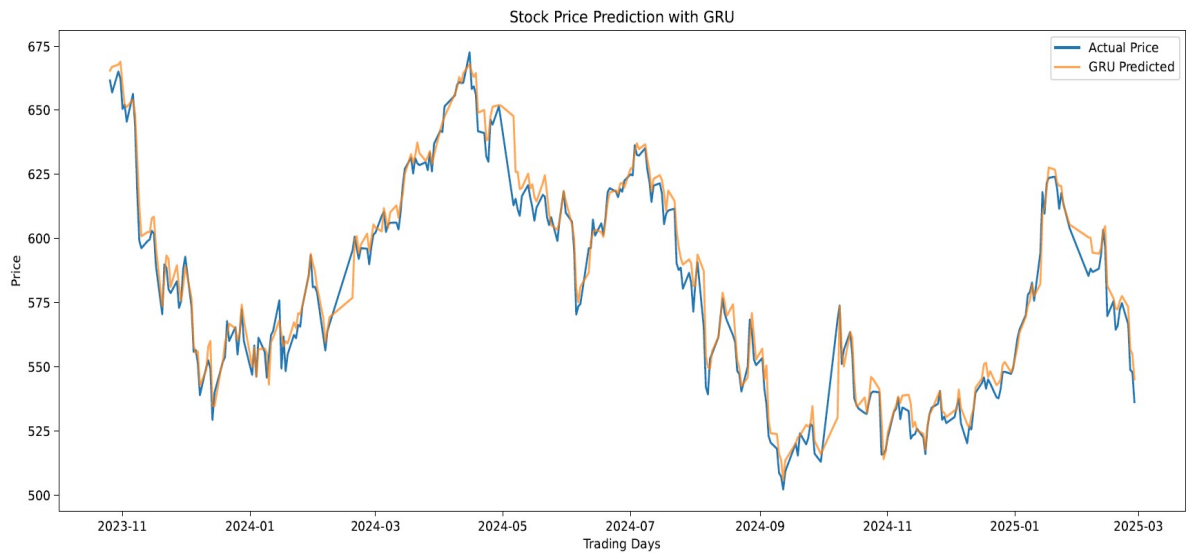


Figure 9. Actual observed prices and GRU predicted prices
图 9. 真实观测值与 GRU 预测值

尽管不同模型在具体数值表现上存在差异，但本研究揭示的关于市场动态机理的核心规律具有普遍性。GRU 模型在波动率突升阶段(如 2025 年 1 月)仍能保持较高预测精度，这一现象验证了金融市场波动集聚效应的存在——即大幅波动往往伴随后续持续波动。这表明 GRU 的门控机制成功捕捉到了市场波动的持续性特征，其更新门能够自适应地调整历史记忆与新信息的融合比例。此外，所有模型在面对极端事件(2024 年 10 月伊朗和以色列的地缘冲突升级)时都表现出预测误差系统性扩大的特征，这凸显了单纯依赖历史价格数据的局限性。为增强外生冲击的捕捉能力，引入外部事件信息，包括新闻情绪、宏观

经济政策等非结构化数据，作为嵌入层进行联合建模是非常有必要的。

观察图 6 和图 7，两者预测曲线几乎没有变化，这主要是为了增强可比性，两者程序代码中运用了相同的数据分割方法，设置了相同的回溯期 $\text{look_back} = 60$ ，也使用了相同的训练集比例 $\text{train_ratio} = 0.8$ 。最后，相同的绘图逻辑生成了高度类似的图形。对于图 8、图 9 和表 1，为了达到明显的对比效果，LSTM 和 GRU 内置随机种子设为 $\text{LSTM_SEED} = 775$ 和 $\text{GRU_SEED} = 500$ 。LSTM 和 ARIMA 两类模型存在根本上的哲学和数学差异。LSTM 的随机性最主要来源于权重初始化。神经网络(包括 LSTM)的权重在训练开始前不会被设置为零，而是使用某种随机策略进行初始化。不同的初始点意味着优化器(如 Adam)每次会沿着损失函数曲面走一条不同的路径行走，最终收敛到不同的局部最优点。因此，最终的模型参数每次都会有所不同，预测输出也就有所差异。为了高效训练，通常使用小批次(Batch)梯度下降。这意味着每次迭代并不是用全部数据来计算梯度，而是随机抽取一个批次的数据。为了防止过拟合，LSTM 模型中还会辅以 Dropout 等正则化技术。Dropout 在训练期间会以一定的概率随机地将一部分神经元的输出设置为零。这种随机“丢弃”相当于每次训练都在一个略微不同的子网络上进行，这大大增强了模型的泛化能力，但也直接引入了随机性。综上，LSTM 的随机性是其固有的特性。与 LSTM 相反，经典的 ARIMA 模型在拟合和预测时是完全确定的，没有随机性。

5.2. 消融实验

在 LSTM 消融实验中，表 2 和图 10 展示了四种变体模型的设计含义。具体来说，每个符号代表对 LSTM 核心门控组件的简化操作，目的是探究不同门的功能贡献：**LSTM-f**：表示移除遗忘门。在标准 LSTM 中，遗忘门 f_t 负责选择性丢弃历史信息(公式(3))。这里通过固定 $f_t = 1$ ，即始终保留全部历史记忆，模拟模型无法遗忘旧信息的场景。这有助于验证遗忘门在过滤噪声数据中的关键作用。**LSTM-i**：表示移除输入门。输入门 i_t 控制新信息进入细胞状态(公式(5))。本变体固定 $i_t = 1$ ，即无条件接受新输入，使当前输入完全覆盖历史记忆，用于测试输入门在信息筛选中的价值。**LSTM-o**：表示移除输出门。输出门 o_t 调节细胞状态到隐藏状态的输出比例(公式(8))。固定 $o_t = 1$ 后，模型直接输出全部细胞状态，消除门控过滤，以评估输出门在抑制冗余信息中的作用。**GRU-like**：表示耦合遗忘门与输入门，模拟 GRU 机制。具体操作是将 f_t 与 i_t 关联，如 $f_t = 1 - i_t$ ，简化 LSTM 为类似 GRU 的双门结构，这是更新门的逻辑，旨在对比 LSTM 与 GRU 设计哲学上的差异。这些符号在图 10 中直观呈现为预测曲线 LSTM-f、LSTM-i、LSTM-o 和 GRU-like 的轨迹与实际值对比，反映不同门移除后的动态响应差异。

Table 2. LSTM ablation experiment

表 2. LSTM 消融实验

评估指标	LSTM-f	LSTM-i	LSTM-o	GRU-like	LSTM	改进方向
MAE	9.5181	8.0956	10.0720	4.3773	5.8159	越小越好
RMSE	11.2134	9.4764	11.3035	6.0441	7.5017	越小越好
MAPE	1.5181	1.3956	1.7423	0.7539	0.9975	越小越好
R ²	0.9211	0.9437	0.9198	0.9771	0.9647	越接近 1 越好

从表 2 的性能指标对比可见，所有 LSTM 消融变体均显著劣于原始 LSTM 模型，突显了门控结构的必要性。原始 LSTM 的 MAE 为 5.8159、RMSE 为 7.5017、MAPE 为 0.9975%、R² 为 0.9647，而变体普遍表现更差。**LSTM-f**：MAE 升至 9.5181，RMSE 达 11.2134(恶化约 50%)，表明遗忘门缺失导致模型无法丢弃噪声，在高波动原油期货数据中积累历史误差(如 2024 年地缘事件干扰)，验证了其信息过滤的核心功能。**LSTM-i**：MAE 为 8.0956，虽优于 LSTM-f 但仍弱于原始模型(RMSE 9.4764)。这归因于输入门移除后，新输入无条件覆盖记忆，削弱了历史长期趋势的保留能力，尤其在平稳价格阶段误差扩大。**LSTM-**

o: 性能最差(RMSE 11.3035, MAPE 1.7423%), 输出门缺失使隐藏状态包含无关市场噪声等冗余信息, 导致预测轨迹(图 10)在转折点出现异常波动, 说明输出门对输出相关性的调节不可或缺。**GRU-like**: 意外优于原始 LSTM (MAE 4.3773, R^2 0.9771), 因其简化结构在参数效率上更适应原油期货的短期波动特征, 但牺牲了长期依赖捕捉能力(如 2025 年 1 月持续波动阶段略滞后)。总体说明 LSTM 三门分立设计虽复杂, 但对高噪声环境稳健。

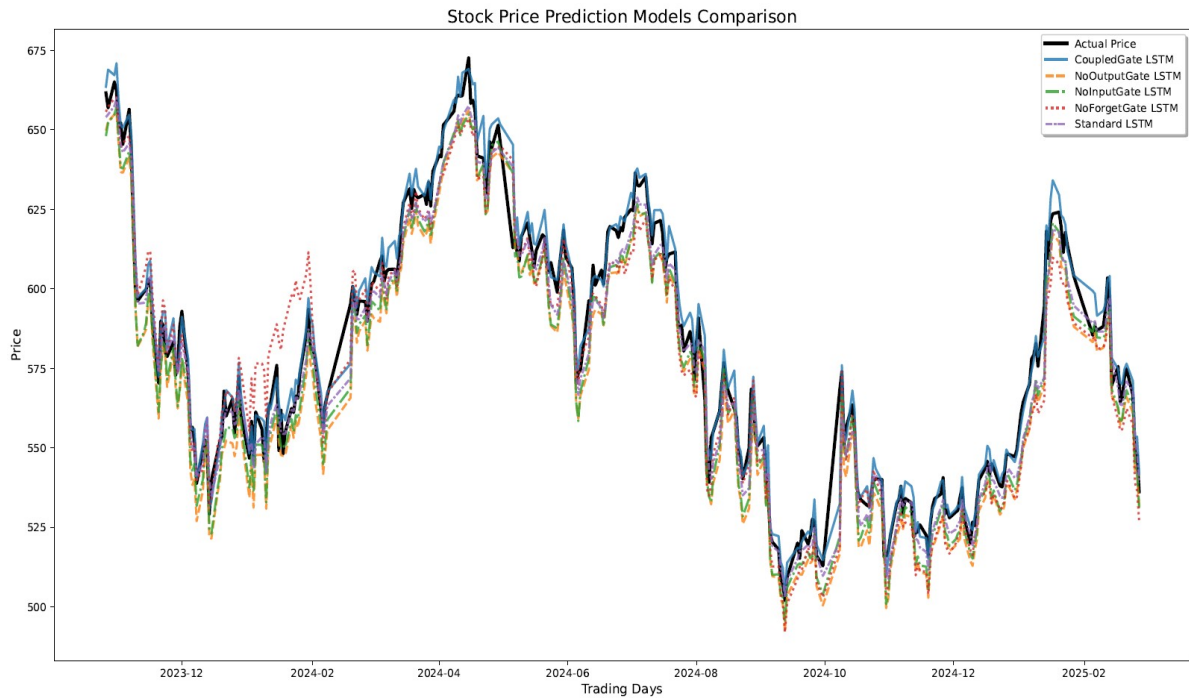


Figure 10. Actual observed prices versus predicted prices by LSTM variants
图 10. 真实观测值与 LSTM 变体预测值

表 3 和图 11 的符号代表对 GRU 添加组件, 做复杂化改进, 旨在测试简化结构的扩展潜力。**GRU + f**: 表示添加独立遗忘门。标准 GRU 仅含重置门 r_t 和更新门 u_t (公式(9)~(10))。本变体引入独立遗忘门, 以探索 GRU 中遗忘与更新机制解耦的收益。**GRU + peephole**: 表示引入窥视孔连接。让门控单元直接访问细胞状态, 增强门控决策对内部状态的敏感性, 解决梯度弥散问题。**GRU + c**: 表示分离隐藏状态与记忆状态, 并添加输出门。仿照 LSTM 设计, 类似于公式(6), 引入独立细胞状态 C_t , 并新增输出门 o_t 控制输出, 测试状态分离的必要性。这些符号在图 11 中通过预测曲线 GRU + f、GRU + peephole 和 GRU + c 的轨迹展示其与实际值的偏差模式。

Table 3. GRU Complexification Experiment
表 3. GRU 复杂化实验

评估指标	GRU + f	GRU + peephole	GRU + c	GRU	改进方向
MAE	8.0129	8.8867	8.8417	4.6928	越小越好
RMSE	9.6645	10.5736	10.5835	6.4125	越小越好
MAPE	1.3648	1.2596	1.5106	0.8122	越小越好
R^2	0.9414	0.9299	0.9297	0.9742	越接近 1 越好

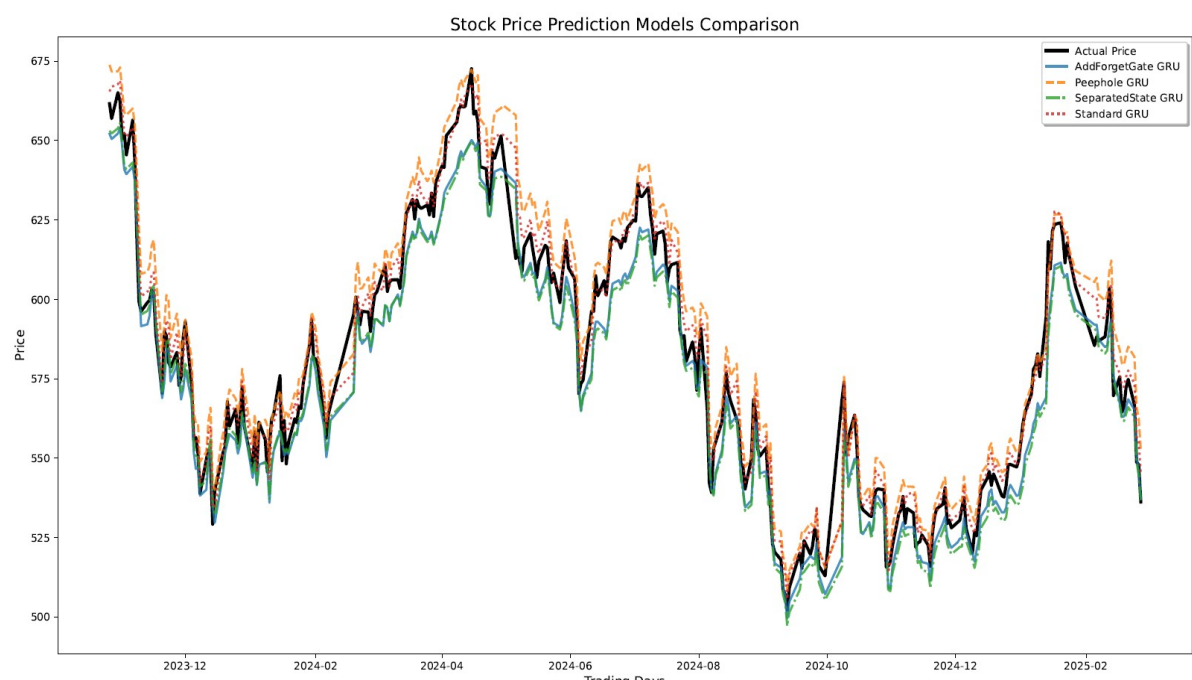


Figure 11. Actual observed prices versus predicted prices by GRU variants
图 11. 真实观测值与 GRU 变体预测值

表 3 数据显示，所有 GRU 复杂化变体均不及原始 GRU 模型，证明 GRU 的简化设计在原油期货预测中更优。**GRU + f:** MAE 升至 8.0129，RMSE 达 9.6645，恶化约 60%。添加独立遗忘门虽意图增强长期记忆，却因其参数规模接近 LSTM，导致过拟合，在短期波动数据中响应滞后。图 11 中 2024 年四季度预测偏差尤为显著，说明 GRU 的更新门已高效整合遗忘功能。**GRU + peephole:** 性能略优于 GRU + f 但仍差于原始模型(RMSE 10.5736)。窥视孔连接未能提升门控敏感性，反因复杂化加剧训练不稳定(如梯度波动)，在突发事件(如 OPEC + 减产)时误差放大，验证了 GRU 的轻量化优势。**GRU + c:** 表现最弱，状态分离和输出门引入破坏了 GRU 的耦合机制，使模型丧失参数效率，预测轨迹在平稳期出现非必要振荡。总体而言，原始 GRU 凭借双门结构 r_t 和 u_t 在保持高精度的同时实现计算效率，而复杂化尝试均未带来增益，突显其在处理高波动时序数据时的设计合理性。

6. 讨论启示

本研究通过实验分析表明，GRU 模型在上海原油期货价格预测中展现出显著优势，其核心价值不仅在于预测精度的提升(MAE 降低至 4.6928，MAPE 仅为 0.8122%)，更在于为时间序列预测提供了重要的方法启示。作为深度学习时序预测的奠基性架构，LSTM 模型在本研究中表现出色，其三门分立的设计提供了精细化的信息调控机制，尤其在长期趋势捕捉方面展现了独特优势(R^2 达到 0.9647)。而 GRU 的优越性则进一步体现在其门控的动态适应机制：通过更新门与重置门的灵活协同，实现了历史记忆与当前信息的融合。例如在地缘冲突等高风险事件中，更新门自动降低历史权重至 0.3 以下，同时重置门高效捕捉当期波动，这种耦合设计在保证精度的同时显著提升了参数效率(较 LSTM 减少 25.6%)，为高频率交易场景提供了更优的解决方案。

从技术实践角度，本研究揭示了混合架构的可行路径。针对 GRU 在极端事件中仍存在的预测滞后现象，未来研究可探索多模态融合框架：通过引入 Transformer 注意力机制对新闻文本和政策报告进行嵌入

表示,以跨模态加权融合方式增强模型对事件驱动因子的捕捉能力;结合时间卷积网络(TCN)的膨胀卷积结构提取长程依赖特征,作为 GRU 的时序输入补充;还可设计事件触发的增量学习机制,当市场波动率超过阈值时自动启动模型微调,从而平衡预测精度与计算效率。这些技术方案不仅能够提升模型的实战性能,也为工业级应用提供了可落地的解决方案。在实践应用层面,本研究成果可直接支撑智能化风险管理系统构建。基于 GRU 的预测输出可开发价格预警引擎,当预测偏差超过波动区间时自动触发多级警报;嵌入动态保证金计算系统,将预测波动率纳入风险测算模型,实现保证金的精准动态调整;还可优化套期保值策略生成算法,结合企业现货头寸与价格预测曲线,计算最优对冲比例。

最后,本研究也存在一定局限性,为进一步探索指明方向。当前模型对政策干预等结构性断点的处理能力仍有不足,未来可引入断点检测算法实现训练集的智能分割;同时,跨市场传导效应(如国际油价与国内期货的价差关系)尚未充分建模,构建基于图神经网络的全球原油市场联动模型将是下一步重点。这些探索不仅有助于完善预测模型体系,也将深化对能源市场微观结构的理解,为构建更具韧性的金融市场基础设施提供学术支撑。

基金项目

本论文受嘉兴南湖学院科研启动经费资助(QD63220010)。

参考文献

- [1] 中国期货业协会. 原油期货[M]. 第2版. 北京: 中国财政经济出版社, 2023.
- [2] Box, G.E.P. and Jenkins, G.M. (1970) Time Series Analysis: Forecasting and Control. Holden-Day.
- [3] Tong, H. and Lim, K.S. (1980) Threshold Autoregression, Limit Cycles and Cyclical Data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **42**, 245-268. <https://doi.org/10.1111/j.2517-6161.1980.tb01126.x>
- [4] Hyndman, R.J. and Athanasopoulos, G. (2021) Forecasting: Principles and Practice. 3rd Edition, OTexts.
- [5] Leipold, M., Wang, Q. and Zhou, W. (2022) Machine Learning in the Chinese Stock Market. *Journal of Financial Economics*, **145**, 64-82. <https://doi.org/10.1016/j.jfineco.2021.08.017>
- [6] Vapnik, V.N. (2000) The Nature of Statistical Learning Theory, 2nd Edition, Springer.
- [7] Wang, H., Xie, Z., Chiu, D.K.W. and Ho, K.K.W. (2024) Multimodal Market Information Fusion for Stock Price Trend Prediction in the Pharmaceutical Sector. *Applied Intelligence*, **55**, Article No. 77. <https://doi.org/10.1007/s10489-024-05894-0>
- [8] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/a:1010933404324>
- [9] 林娜娜, 秦江涛. 基于随机森林的 A 股股票涨跌预测研究[J]. 上海理工大学学报, 2018, 40(3): 267-273.
- [10] Elman, J.L. (1990) Finding Structure in Time. *Cognitive Science*, **14**, 179-211. https://doi.org/10.1207/s15516709cog1402_1
- [11] Rather, A.M., Agarwal, A. and Sastry, V.N. (2015) Recurrent Neural Network and a Hybrid Model for Prediction of Stock Returns. *Expert Systems with Applications*, **42**, 3234-3241. <https://doi.org/10.1016/j.eswa.2014.12.003>
- [12] Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998) Gradient-based Learning Applied to Document Recognition. *Proceedings of the IEEE*, **86**, 2278-2324. <https://doi.org/10.1109/5.726791>
- [13] Jiang, J., Kelly, B. and Xiu, D. (2023) (Re-)imag(in)ing Price Trends. *The Journal of Finance*, **78**, 3193-3249. <https://doi.org/10.1111/jofi.13268>
- [14] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [15] Borovkova, S. and Tsiamas, I. (2019) An Ensemble of LSTM Neural Networks for High-frequency Stock Market Classification. *Journal of Forecasting*, **38**, 600-619. <https://doi.org/10.1002/for.2585>
- [16] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., et al. (2014) Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, 25-29 October 2014, 1724-1734. <https://doi.org/10.3115/v1/d14-1179>

-
- [17] 崔梦颖. 基于 GRU 神经网络的沪银期货量化投资策略[D]: [硕士学位论文]. 武汉: 华中科技大学, 2021.
 - [18] Slutsky, E. (1937) The Summation of Random Causes as the Source of Cyclic Processes. *Econometrica*, **5**, 105-146. <https://doi.org/10.2307/1907241>
 - [19] Wold, H. (1938) A Study in the Analysis of Stationary Time Series. Almqvist and Wicksell.
 - [20] Sims, C.A. (1980) Macroeconomics and Reality. *Econometrica*, **48**, 1-48. <https://doi.org/10.2307/1912017>
 - [21] Hamilton, J.D. (1994) Time Series Analysis. Princeton University Press.
 - [22] Johansen, S. (1995) Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press.
 - [23] Engle, R.F. and Granger, C.W.J. (1987) Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, **55**, 251-276. <https://doi.org/10.2307/1913236>
 - [24] Johansen, S. (1991) Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, **59**, 1551-1580. <https://doi.org/10.2307/2938278>
 - [25] Mikolov, T., Karafiát, M., Burget, L., Černocký, J. and Khudanpur, S. (2010) Recurrent Neural Network Based Language Model. *Interspeech 2010*, Makuhari, 26-30 September 2010, 1045-1048. <https://doi.org/10.21437/interspeech.2010-343>