学术前沿专栏:数智金融与资本市场高质量发展

可解释机器学习与股指收益率预测的经济价值

——来自中国市场的证据

李艳

(河南大学 商学院,河南 开封 475004)

摘要:本文针对上证综指和深证成指,搭建了一个同时具备预测能力和可解释性的多模态混合框架.输入端汇集了宏观经济、市场交易、企业盈利和投资者情绪 4 类信息.建模时,先用 CEEMDAN 把收益率序列按尺度分解,由 LightGBM 挑出各频段的关键特征,再分别用 Attention-LSTM、Attention-GRU 和 LSTM 拟合高、中、低频分量,最后由 SVR 做非线性集成.在样本外检验中,混合模型在两个市场都取得了正的调整后样本外 R^2——上证 0.48%,深证 1.23%,均优于线性基准和单一深度学习模型.SHAP 分析显示,真正驱动预测的主要是宏观货币与价格类变量.择时方面,在 long-cash 策略、0.1%交易成本的设定下,混合模型相对买入持有改善了风险调整后绩效.这些结果为多模态机器学习用于股指收益率预测、并兼顾可解释性与经济价值,提供了来自中国市场的经验证据.

关键词:股指收益率预测;多模态数据;可解释机器学习;SHAP;市场择时

中图分类号:F830.91;TP181 **文献标志码**:A **文章编号**:1000-2367(2026)05-0011-10

股指收益率的预测,关系到资产配置和风险控制.近些年,机器学习因为擅长捕捉变量间的非线性关系、又能消化高维信息,开始被用到收益预测和资产定价里[1].按传统资产定价的观点,公开信息会很快被价格吸收,可预测性因此受有效市场假说的约束[2];但自适应市场假说给出了另一种可能——市场效率本身并非恒定,会随投资者结构、制度环境和市场状态而变,这就为某些阶段、某些状态下的收益预测留出了余地[3].经验层面也有不少证据,把宏观基本面、经济不确定性、市场状态这些变量和资产收益及其波动联系在一起[2,4].

可是,机器学习相比线性基准到底有没有稳定的样本外增量,学界至今没有定论.指数的日度收益噪声大、信噪比低,复杂模型固然把拟合的自由度提了上去,可估计误差也随之被放大,并存在过拟合的风险[5].复杂模型能否在样本外优于那些结构清楚、含义明确的线性模型,还需进一步检验[6-7].Gu 等[1]的结果是,非线性机器学习在资产收益的样本外预测上确实优于多种线性基准;Chen 等[8]使用深度学习研究资产定价,说明复杂模型能挖出更丰富的状态依赖信息.关于中国市场的探讨,李斌等[9-10]、刘莉等[4]分别从可预测性、组合选择、时变稳健建模几个角度做过研究;相关综述也已经把机器学习和深度学习当作金融时序预测的主流工具[11].

除了传统金融与宏观变量,文本和情绪也成了收益预测的一类新信息.已有研究指出,媒体文本里的情绪、

收稿日期:2026-05-31;**修回日期**:2026-06-22.

基金项目:国家社会科学基金(23BJY014).

作者简介(通信作者):李艳(1996—),女,河南洛阳人,河南大学博士研究生,研究方向为金融工程,E-mail:104753190936@henu.edu.cn.

引用本文:李艳.可解释机器学习与股指收益率预测的经济价值——来自中国市场的证据[J].河南师范大学学报(自然科学版),2026,54(5):11-20.(Li Yan.Interpretable machine learning and the economic value of stock index return forecasts:Evidence from Chinese stock market[J].Journal of Henan Normal University(Natural Science Edition),2026,54(5):11-20.DOI:10.16366/j.cnki.1000-2367.2026.05.31.0001.)

以及来自不同渠道的金融文本,所含的信息量并不一样[12-13];张宗新等[14]、姚加权等[15]则从媒体情绪传染、金融情绪词典的角度,进一步检验了文本情绪对市场行为的作用.预训练模型出现后,BERT 这类工具为中文金融文本的情绪识别提供了新的支撑[16];Gan 等[17]还发现,新闻情绪和社交媒体情绪可能彼此互补,各自拥有对方没有的信息.

关于时间序列的预测,分解－集成的思路加上深度时序模型,已经被广泛用于金融序列.Lin 等[18]和 Cao 等[19]把 CEEMDAN 和 LSTM 结合起来进行股指预测;Li 等[20]则给 LSTM 增加注意力,以捕捉序列里的关键时点.国内的研究,从分解－集成、混频采样,到经验模态分解、谱分解和各种神经网络结构,同样遵循相似的研究逻辑[21-24].综合这些工作可以看到,多尺度分解配上非线性模型,有助于压住噪声、把不同频段里的可预测信息分开刻画,但引发的问题是模型变得更难解释.为了补上这块短板,事后解释方法被引了进来,其中基于博弈论的 SHAP 能算出每个变量对输出的边际贡献[25-26].

基于上述讨论,本文关注的核心问题是:在指数日度收益率预测中,非线性方法的增量价值究竟来自新的强预测信号,还是主要表现为在线性基础上的有限修正.围绕这一问题,本文主要从 3 个方面展开分析:第 1,考察宏观经济、交易、盈利和投资者情绪等多模态变量在不同频率成分中对收益率可预测部分的影响;第 2,检验带结构约束的混合模型是否能够在样本外表现上优于线性基准和单一非线性模型,并结合 SHAP 分析其预测优势来源;第 3,评估预测改进能否通过择时策略转化为具有经济意义的投资绩效.

本文的边际贡献体现在 4 个方面.第 1,将股指收益率预测置于统一的样本外比较框架中,识别非线性模型增量价值的性质和边界.第 2,整合宏观、交易、盈利和文本情绪等多模态信息,并结合特征选择与多尺度分解方法降低冗余和噪声.第 3,引入 SHAP 方法解释混合模型预测来源,增强复杂模型结果的透明度.第 4,从资产配置和择时角度检验预测信号的经济价值,说明即便统计意义上的改进有限,模型输出仍可能具有实践价值.

1 理论基础

在给定信息集 $\mathcal{F}$ 下,本文将市场指数超额收益率 y_{t+1} 的条件期望定义为:

$$E_{t(y_{t+1}|\mathcal{F}_t)}=\pi_t^{\text{risk}}+\alpha_t, \tag{1}$$

其中,π_t^{risk} 是由市场风险状态决定的条件风险补偿项,α_t 表示剩余条件可预测项.接下来,基于风险状态变量 Z_t 构建条件风险补偿项:

$$\pi_t^{\text{risk}}=f(Z_t), \tag{2}$$

其中,Z_t 包括市场波动率、流动性指标及宏观金融状态变量等. 进一步地,定义剩余预测对象为:

$$\tilde{y}_{t+1}=y_{t+1}-\hat{\pi}_t^{\text{risk}}. \tag{3}$$

在此基础上,本文将 $\tilde{y}_{t+1}$ 视为一个由多模态信息驱动的状态依赖函数, 即:

$$\tilde{y}_{t+1}=\mathcal{F}(X_t^{\text{macro}},X_t^{\text{trading}},X_t^{\text{profit}},X_t^{\text{sentiment}}), \tag{4}$$

其中,X_t^{macro}、X_t^{trading}、X_t^{profit} 和 $X_t^{\text{sentiment}}$ 分别表示宏观经济、交易行为、盈利指标与情绪指数等信息. 选择多种预测指标的原因是不同指标能够从不同维度反映研究对象的变化特征,有助于提升模型的预测性能.

基于上述理论设定,本文遵循风险补偿基准识别、多尺度剩余项预测及非线性集成的脉络,构建了 CEEMDAN-Attention-LSTM/GRU-SVR 混合预测模型,模型共包含 3 个阶段:特征分解与重构、特征选择、分量预测和非线性集成.本文预测模型总体框架见附录图 S1 所示.

1.1 混合预测模型框架

本文首先利用风险状态变量 Z_t 估计市场指数收益率中的条件风险补偿项 $\hat{\pi}_t^{\text{risk}}$,并据此构造剩余预测对象 $\tilde{y}_{t+1}$,由此区分风险补偿驱动的可预测部分与由多模态状态变量驱动的剩余预测信号.

考虑到 $\tilde{y}_{t+1}$ 具有非平稳、高噪声和多尺度波动等特征,采用自适应噪声的完全集成经验模态分解(CEEMDAN)对其进行分解.CEEMDAN 通过引入自适应噪声并进行集成处理,能够在一定程度上缓解模态混叠问题,从而更稳定地提取不同时间尺度下的动态成分.经 CEEMDAN 分解后,剩余收益率序列被表示为 M 个模态分量,各分量具有不同的波动频率和信息特征.

各模态分量的信息频率并不相同,若用同一个模型、同一套特征统一建模,很容易把不同频段的结构差异抹平.为此,本文借助排列熵(permutation entropy,PE)识别各模态分量的复杂程度,并据此将其重构为高频、中频和低频 3 类预测分量.具体而言,模态分量 k 的排列熵定义为:$H_{pe}(m)=-\sum_{j=1}^{K}P_j\ln(P_j)$,其中,$P_j$ 表示第 j 种排列模式出现的概率.H_{pe} 值越大,表明时间序列的波动频率越强;H_{pe} 值越小,表明时间序列的波动频率越弱.

借鉴已有研究[20,27],本文设定阈值 $\lambda_0=0.5$,$\lambda_1=0.01$.当 $H_{pe}\geq\lambda_0$,则模态分量 k 被识别为高频分量;若 $\lambda_1<H_{pe}<\lambda_0$,则模态分量 k 被识别为中频分量;若 $H_{pe}\leq\lambda_1$,则模态分量 k 被识别为低频分量.根据各模态分量识别结果,将被识别高频、中频、低频的模态分量进行加和得到收益率序列的高频分量、中频分量和低频分量.

不同频段背后的驱动力本就不同:高频成分主要来自短期交易和情绪冲击,低频成分更多承载宏观和基本面趋势.如果不加区分地给所有分量输入同一套变量,冗余信息会拖累样本外表现.所以本文对高、中、低频分别用 LightGBM 做特征选择,因其能处理变量间的非线性关系,再按重要性排名挑出更有预测价值的那些,从而提升输入质量与预测稳定性.

在分量预测时,基于其特征使用不同的预测模型.高频分量噪声多、变化快,使用 Attention-LSTM,通过注意力提升对关键信息的捕捉;中频分量带一些趋势和周期性,用结构更轻、参数更少的 Attention-GRU 进行预测;低频分量主要是长期趋势,使用长记忆刻画能力较强的 LSTM 模型.

3 个分量各自预测完之后,再用支持向量回归(support vector regression,SVR)把它们集成最终结果.相较于简单的线性加和,SVR 借助核函数能刻画各频段预测值与真实收益之间的非线性关系,其基本形式为:

$$f(x)=\sum_{j=1}^{N}\alpha_i K(x,x_i)+b, \tag{5}$$

其中,α_i 是拉格朗日乘子,x_i 是本文通过 Attention-LSTM/GRU 拟合的高、中、低频分量结果,$K(x,x_i)$ 是用于将数据映射到高维空间的核函数,b 是偏置项.

1.2　样本外预测结果评价

为评价不同模型的样本外预测表现,本文构建了样本外 R^2、平均绝对误差 MAE、中位绝对误差 $MeDAE$、均方误差 MSE 和均方根误差 $RMSE$ 作为评价指标.具体定义如下:

$$R^2=1-\frac{\sum_{t=1}^{N}(y_t-\hat{y}_t)^2}{\sum_{t=1}^{N}(y_t-\bar{y})^2}, \tag{6}$$

$$MAE=\frac{\sum_{t=1}^{N}|\hat{y}_t-y_t|}{N}, \tag{7}$$

$$MeDAE=\text{median}(|y_1-\hat{y}_1|,\cdots,|y_n-\hat{y}_n|), \tag{8}$$

$$RMSE=\sqrt{\frac{\sum_{t=1}^{N}(\hat{y}_t-y_t)^2}{N}}, \tag{9}$$

其中,y_t 表示上证综指收益率的真实值,$\hat{y}_t$ 表示上证综指收益率的预测值,$\bar{y}_t$ 表示上证综指收益率真实值的平均值,$\bar{y}$ 表示上证综指收益率预测值的平均值,N 是观测次数.

R^2 越高,说明模型相对于基准预测具有更强的解释和预测能力;MAE、$MeDAE$、MSE 和 $RMSE$ 均用于衡量预测误差,数值越小表示预测结果越接近真实收益率.

2　收益率预测实证研究

2.1　样本选择

2.1.1　结构化数据选取

本文选择我国股票市场代表性指数作为研究对象,数据来源于 Wind 终端,结构化数据涵盖 2011 年1 月

1 日至 2025 年 12 月 31 日.样本区间覆盖 2011－2017 年的经济扩张期、2018－2019 年的经济震荡期、2020－2022 年的经济衰退期以及 2023 年后的经济复苏期,其中训练集为 2011－2022 年、测试集为 2023－2025 年,能够检验模型在不同市场状态下的样本外表现,$Return_t = \ln(P_t/P_{t-1})$.

第二类结构化交易数据是宏观经济变量,我们关注了多个反映经济健康状况的重要指标.具体而言,选取了货币政策指数(MIP)、经济政策不确定性指数(EPUI)、居民消费价格指数增长率(CPIG)、货币供应量 M0、M1、M2 的增长率(M0G、M1G、M2G)、定期存款利率(TDR)、短期利率(SRa)、汇率(Exchange)和新增就业人数增长率(NERG)等变量,分别从经济运行状况、物价水平变化、货币流通量、市场资金供求变动及劳动力市场活跃度等角度对整体经济状况进行衡量.

最后,我们使用了企业估值数据,包括市盈率(PER)、市净率(PB)、股息率(DR)、市销率(PS)、实现率(PCF),评估市场的盈利能力.

2.1.2 投资者情绪构建

投资者情绪的测度,文献里大体有两类研究:基于市场交易指标或基于文本数据.交易类指标虽能反映市场情绪动态,但在准确性方面存在不足,因为它们容易受到市场其他因素的影响.文本数据能够及时提供投资者丰富的情绪信息,其测度可以根据信息来源分为新闻和社交媒体两类,Gan 等[17]的研究表明社交媒体信息比新闻信息对情绪的刻画更有竞争力.因此,本文选择社交媒体平台作为文本情绪信息来源.

众多平台里,东方财富网股吧的讨论量和覆盖面都比较突出.本文从中抓取与两市指数相关的发帖与回复,构建 2011 年 1 月 1 日至 2025 年 12 月 31 日的情绪文本样本,用来刻画投资者的预期和分歧.

为便于后续分析,本文对投资者情绪文本数据集进行如下预处理与文本清洗.首先,去除网址链接、乱码等各类非中文字符;其次,为了减少噪声文本的影响,本文删除了内容为单个字符的短文本数据,因为通常来说过短的文本几乎无法表达有效信息;随后,删除重复文本字段、无效新闻(如转发、表情符号等)和回顾性历史新闻(如月度前瞻回顾等)等无效样本.经过上述预处理过程后,投资者情绪文本数据集共计 211 万条样本.

由于本文使用的投资者情绪文本以中文表述为主,为此,本文应用了谷歌开源的“进一步预训练”(further pre-training)中文 BERT 框架(PyTorch 版本),该框架在大规模的中文文本上进行了预训练,可用于各种中文文本任务(如文本分类等).为了进一步提高该模型在投资者情绪文本上的表现,本文借助人工标注方法,通过微调将中文大语言模型针对投资者情绪文本分类任务进行优化.在微调过程中,本文运用五折交叉验证方法划分训练集、验证集,实现模型参数的微调优化.最后,本文利用训练完成的投资者情绪文本分类大语言模型,输出对样本中投资者情绪文本的分类结果,并在此基础上进一步构建投资者情绪指数.

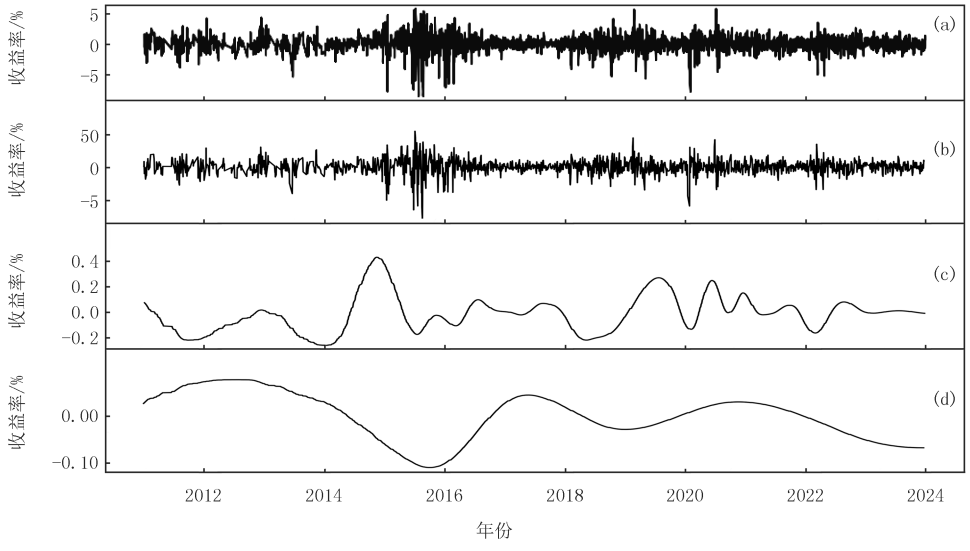
得到投资者情绪分类结果后,根据每日投资者发布帖子主题的句子极性以及发帖时间计算每日的投资者情绪指数:通过发帖时间计算每日消极帖子数 I^{positive} 和积极帖子数 I^{negative},每日投资者情绪指数计算方式为:$PNI = \dfrac{I^{\text{positive}} - I^{\text{negative}}}{I^{\text{positive}} + I^{\text{negative}}}$,其 $I^{\text{positive}} - I^{\text{negative}}$ 中代表着评论主题中积极情绪和消极情绪的差,取值越大表明情感倾向差异性越大,投资者对市场的预期越乐观;$I^{\text{positive}} + I^{\text{negative}}$ 是看涨帖子和看跌帖子总数.PNI 的取值范围为$[-1,1]$,PNI 取值越接近于 1,表明看涨帖子数量越多,投资者对市场越乐观;PNI 取值越接近于 -1,表明看跌帖子数量越多,即投资者对市场预期越悲观.

综上,本文确定了 24 个股票市场收益率的预测变量.考虑到不同变量间的数量级存在差异,本文对各变量日度序列数据进行了标准化处理,即 $z_{\text{scaled}} = \dfrac{z - \bar{z}}{\sigma_z}$,其中 $\bar{z}$ 和 σ_z 分别表示各变量的均值和标准差.

2.2 特征选择

本文使用 CEEMDAN 分解得到 11 个模态分量,并依据排列熵将其重构为高频、中频和低频收益分量.图 1 是收益率序列的 PE 重构结果,其中图 1(a)是原始的收益率时间序列图;图 1(b)是收益率序列经 CEEMDAN 分解后通过 PE 重构的高频分量,可以看出其波动趋势大致与原始的收益率时间序列相同,波动较为剧烈;图 1(c)是收益率序列经 CEEMDAN 分解后通过排列熵(PE)重构的中频分量,从图中可以看出其中频分量波动与原始收益率序列的波动趋势大致相同,且波动性相较于高频分量较低;图 1(d)是低频分

量,波动频率低,反映收益率序列的长期趋势.



(a)原始收益率;(b)高频分量;(c)中频分量;(d)低频分量.

图1　上证综指收益率序列PE重构结果

Fig. 1　PE reconstruction results of the Shanghai composite index return series

分解到不同时间尺度之后,市场在各个频段上的动态就分得更清楚了,接下来用 LightGBM 逐个频段评估变量进行重要性评估,分析宏观经济、交易和投资者情绪等不同类型的变量在不同频段对上证综指股指收益的贡献.

附录图 S2 的结果表明各类变量贡献在各频率上并不均衡:原始、高频和中频分量上各变量得分普遍偏低,真正的重要度集中在低频分量的宏观与估值变量上.这可以理解为,指数收益中那部分相对稳定、可被预测的成分,主要来自货币、价格和估值环境,而不是高频的交易扰动.

2.3　股指收益率预测精度

表 1 报告了上证综指的各模型调整 R^2,同时给出 *MAE*、*RMSE* 与方向准确率.从模型拟合效果的角度,混合模型的调整 R^2 为 0.48%,高于基准模型,说明本文提出的混合模型在样本外仍保留相对预测优势.从预测误差看,混合模型的 *MAE* 和 *RMSE* 分别为 65.10 bp 和 97.54 bp,均低于 6 类基准模型;其方向准确率为 51.88%.这些结果说明,混合预测框架能有效降低预测误差.

表 1　样本外预测精度比较

Tab. 1　Comparison of out-of-sample forecasting accuracy

模型	调整后 R^2/%	MAE/bp	RMSE/bp	方向准确率/%	模型	调整后 R^2/%	MAE/bp	RMSE/bp	方向准确率/%
HA	−0.01	65.14	97.78	52.02	LSTM	−19.30	74.91	106.80	**53.28**
AR(1)	+0.15	65.12	97.71	49.51	LightGBM	−21.75	79.33	107.89	48.95
Lasso	+0.37	66.15	97.60	51.19	Mix Model	**+0.48**	**65.10**	**97.54**	51.88
ER	+0.23	66.26	97.67	50.91					

注:加粗下划线为该评价指标下表现最优的模型,下同.

为了更直观地比较不同方法在时间序列上的预测效果,本文计算并绘制了基于历史均值估计的累积平方误差序列(*CSED*,以衡量预测模型在长期尺度上是否优于简单的均值预测. $CSED_t = \sum_{j=1}^{t}[(y_j - \bar{y}_{j,\mathrm{mean}})^2 - (y_j - \hat{y}_{j,M})^2]$,其中,$y_j$ 表示第 j 期的上证综指真实收益率值;$\bar{y}_{j,\mathrm{mean}}$ 表示对第 j 期的历史均值预测值,即仅利用前$(j-1)$期的历史数据均值来预测当期值;$\hat{y}_{j,M}$ 表示模型 M 对第 j 期的预测值.

图 2 是上证市场各模型相对历史均值基准的累积平方误差差异(*CSED*),橙色虚线是其线性趋势线.

Mix Model 的趋势线向上,累积差整体保持在正区间;Lasso 和 ER 的向上的趋势线则要弱得多;LSTM 的趋势线向下.

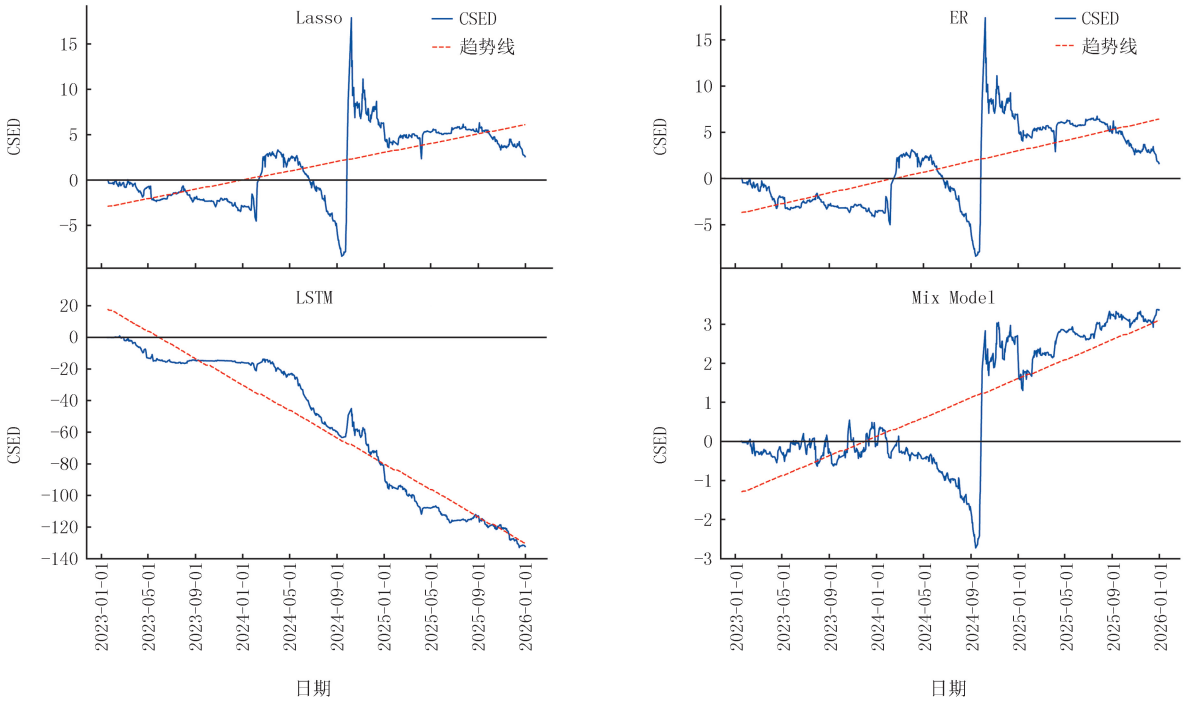


图2 各模型的样本外累积平方误差

Fig. 2 Out-of-sample cumulative squared errors of various models

分开看,Lasso 的 *CSED* 曲线早期也有过正向累积,但中段以后一路回落,到后期已经跌到 0 轴下方.也就是说,Lasso 并不能稳定地比历史均值基准做得更好.ER 的轨迹和它几乎一样,前期为正、之后向下,在长区间里同样不具备显著的优势.

2.4 稳健性检验

接下来,本文以深证成分指数为对象进行稳健性检验,以考察上述结论是否依赖于单一市场样本.结果表明,CEEMDAN-Attention-LSTM/GRU-SVR 混合预测模型在深证市场中仍保持了相对较好的预测表现,整体优于所有基准模型.从表 2 可以看出,混合模型的调整 R^2 为 1.23%,明显高于 Lasso 的−46.77%、ER 模型的−57.50%、LightGBM 的−12.6%和 LSTM 模型的−2.56%.在预测误差方面,混合模型的 RMSE 为 140.17 bp,均为 7 类模型中最低;不过,其方向准确率为 50.91%,提升幅度并不十分突出,表明该模型的优势主要体现在误差控制和整体拟合改善方面,这与上证指数表现一致.

结合 CSED 曲线来看,图 3 中 Mix Model 在深证市场中相对于历史均值基准的累积误差改进同样呈现逐步扩大趋势,趋势线斜率为正.这一结果进一步支持了在上海市场中得出的结论,即混合模型在复杂市场环境下能够更好地利用多模态信息.

表 2 深证样本外预测精度比较

Tab. 2 Comparison of out-of-sample forecasting accuracy for SZ

模型	调整后 R^2/%	*MAE*/bp	*RMSE*/bp	方向准确率/%	模型	调整后 R^2/%	*MAE*/bp	*RMSE*/bp	方向准确率/%
HA	−0.00	**96.06**	141.04	48.81	LSTM	−2.56	98.10	142.83	49.37
AR(1)	+0.48	96.50	140.70	46.03	LightGBM	−12.60	105.88	149.66	46.72
Lasso	−46.77	123.36	170.87	**51.19**	Mix Model	**+1.23**	96.26	**140.17**	50.91
ER	−57.50	128.98	177.00	**51.19**					

2.5 混合预测模型预测优势分析

为进一步解释混合模型相对其他方法的预测优势,本文引入 SHAP(SHapley Additiveex Planations)方法对模型预测结果进行解释.SHAP 方法基于博弈论中的 Shapley 值思想,能够为每个特征分配边际贡献

值,从而从全局重要性和局部影响分布两个层面揭示模型的预测机制.

附录图 S3 把混合模型和传统 LSTM 在上证综指上的特征贡献进行了对照.混合模型中贡献明显向少数变量集中,M2 增长率的平均绝对 SHAP 值居首,与其余变量拉开了距离,紧随其后的是定期存款利率、M0 增长率和 CPI 增长率,均为宏观货币与价格类指标.换句话说,模型主要是依据货币环境和价格水平来判断指数收益,这和前面 LightGBM 给出的重要性排序表现一致.

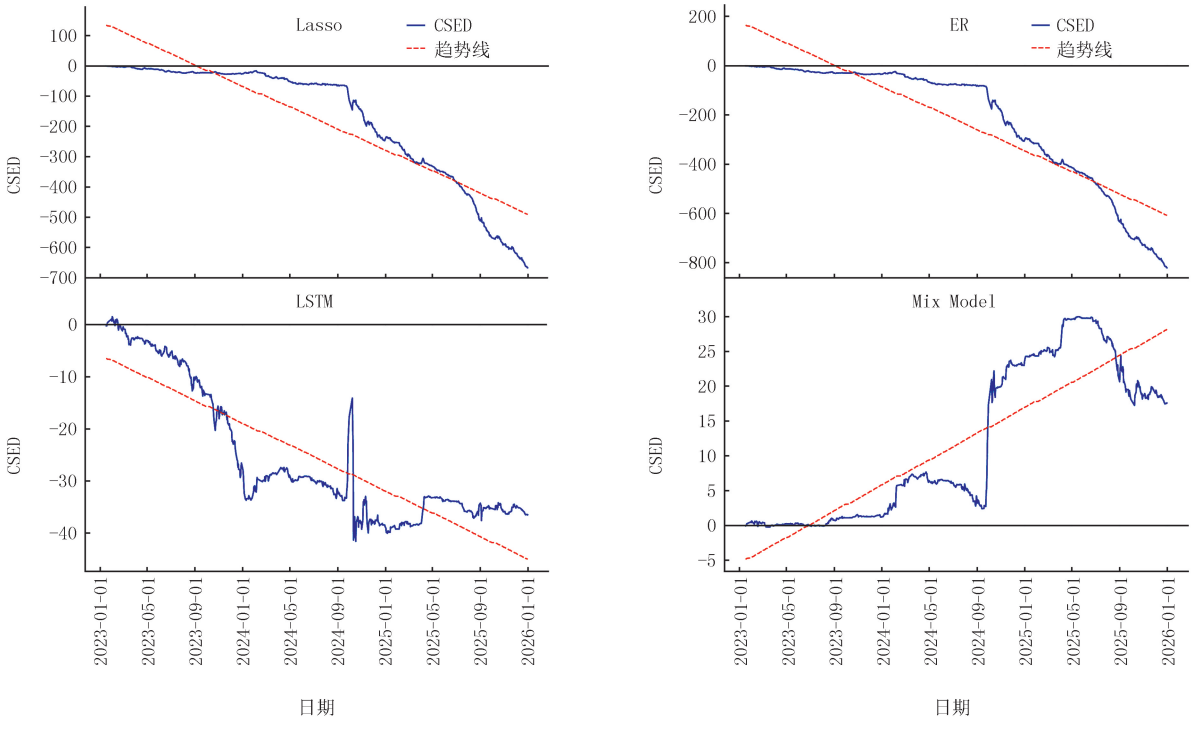


图3　各深证综指预测模型的样本外累积平方误差

Fig.3　Out-of-sample cumulative squared errors of various models for SZ

从局部 SHAP 分布看,混合模型的取值更集中在零轴附近,对特征水平变化的响应较为平稳,在不同市场区间里给出的边际贡献也比较一致.这种平稳性和模型前端的处理有关:原始序列经分解后噪声已被剥离,注意力层只需在信息量较高的局部上分配权重,SVR 则吸收各频段之间残余的非线性关系,模型因而不必在高频扰动上反复调整.传统 LSTM 则不同:贡献分布更散,没有真正突出的核心变量,局部 SHAP 值正负摆动的幅度也更大.这意味着它对变量变化的敏感程度并不稳定,遇到高频波动或噪声较重的区间,预测误差容易被放大,样本外表现随之打折.

进一步来看,交易类变量如成交额、收盘价和开盘价等 SHAP 贡献位于中游且整体数值较小,说明在本文设定下,日内交易特征对上证综指收益率的边际解释力相对有限.估值类指标和汇率变量的全局 SHAP 贡献更小,整体排序靠后,表明混合模型的预测优势主要来自低频宏观变量,而非高频交易变量或估值类变量.

投资者情绪指数的全局 SHAP 贡献也相对靠后,说明就日度指数收益率预测而言,情绪变量的边际解释力有限,模型的主要可解释结构仍由宏观货币和价格类变量承载.不过,情绪变量在极端市场环境下仍可能对收益率产生阶段性影响.混合模型通过多频成分提取和注意力权重分配,能够在一定程度上捕捉这类非线性和阶段性信息,从而为预测结果提供补充.

可见,混合模型的优势主要源于:分解先剥离了序列里的高频噪声,LightGBM 和注意力权重把建模重心放到信息量更高的特征上,SVR 最后集成各分量之间的非线性关系.

3　股指收益率预测的经济价值

本文基于样本外预测信号构建 long-cash 择时策略:若模型预测下一期收益率为正,则持有股指资产;若

预测为负,则持有现金.考虑 A 股市场做空限制,本文不设置 short 策略,并将单边交易成本设为 0.1%.买入持有策略作为比较基准.

从表 3 的结果可以发现,在上海和深圳两个市场中,基于调优后混合模型预测信号构建的择时策略均改善了风险调整后绩效.以上海市场为例,混合模型的 long-cash 策略 Sharpe 比率为 0.32,高于买入持有策略的 0.29,最大回撤从−20.41%收窄至−14.73%,CER 增益为 39.42 bp.

表 3 市场择时策略的经济价值

Tab. 3 Economic value of market-timing strategies

市场	策略	年化收益率/%	年化波动率/%	Sharpe	最大回撤/%	CER 增益/bp
上海市场	Buy-and-hold	**7.54**	15.52	0.29	−20.41	0.00
上海市场	HA	7.50	15.52	0.29	−20.41	−3.53
上海市场	LSTM	7.45	**14.43**	**0.31**	**−17.89**	**24.28**
上海市场	Mix Model	7.24	13.18	0.32	−14.73	39.42
深圳市场	Buy-and-hold	8.03	22.38	0.22	−29.84	0.00
深圳市场	HA	7.99	22.38	0.22	−29.84	−3.53
深圳市场	LSTM	8.61	19.84	0.28	−24.80	160.39
深圳市场	Mix Model	**13.54**	**16.49**	**0.64**	**−20.11**	**725.92**

在深证市场样本中,混合模型对应策略的经济价值表现更为突出.具体来看,基于混合模型预测信号构建的 long-cash 策略年化收益率为 13.54%,Sharpe 比率为 0.64,明显高于买入持有策略的 0.22.同时,该策略的最大回撤由买入持有策略的−29.84%降低至−20.11%,CER 增益达到 725.92 bp.由此可见,混合模型的预测信号虽然在统计精度上的提升幅度有限,但仍能转化为具有经济意义的市场择时价值.

4 结束语

股指收益率的准确预测一直是学术界和业界关注的重点.本文针对上证综指与深证成指的收益率预测问题,提出了一种基于 CEEMDAN 分解、Attention-LSTM/GRU 模型以及 SVR 集成的可解释混合预测框架.通过对高、中、低不同频段的收益率序列分别建模,再借助 SVR 对各频段预测结果进行非线性集成,实证结果显示,调优后的混合模型在两个市场均取得正的调整后 R^2(上证 0.48%、深证 1.23%),在预测精度与稳定性上相对线性基准与单一深度学习模型有所改进,其中深证市场的改进更为明显.基于 SHAP 的解释分析表明,模型的预测主要由宏观货币与价格类变量驱动:上证市场以 M2 增长率、定期存款利率、M0 与 CPI 增长率的贡献居前,深证市场则以 CPI 增长率与股息率居前;交易类与情绪类变量(含投资者情绪指数 PNI)的边际解释力相对有限.进一步地,市场择时分析显示,混合模型的预测信号可转化为正的经济价值:在 long-cash 策略与 0.1%交易成本下,混合模型相对买入持有获得更高的夏普比率与更小的最大回撤,深证市场的 CER 增益尤为可观.

就应用而言,本文的框架在市场择时和资产配置中具有一定参考价值,尤其是在控制回撤、改善风险调整后收益方面.但研究本身也有边界:择时检验只覆盖了受 A 股做空限制约束的 long-cash 一种策略,样本也仅限两个宽基指数.后续可以把框架推广到允许做空的品种或行业指数上,单独刻画深证的特征贡献,并进一步检验情绪指数在极端行情中的增量价值,以及更细的可解释性工具能否让信号在实盘风控中更好落地.

附录见电子版(DOI:10.16366/j.cnki.1000-2367.2026.05.31.0001).

参 考 文 献

[1] Gu S H,Kelly B,Xiu D C.Empirical asset pricing via machine learning[J].The Review of Financial Studies,2020,33(5):2223-2273.

[2] Fama E F.Efficient capital markets:II[J].The Journal of Finance,1991,46(5):1575-1617.

[3] Lo A W. The adaptive markets hypothesis: market efficiency from an evolutionary perspective[J]. Journal of Portfolio Management, 2004,30(5):15-29.

[4] 刘莉,郝显峰,王玉东.基于时变稳健加权最小二乘法的股市收益率预测[J].管理科学学报,2024,27(1):141-158.
Liu L,Hao X F,Wang Y D.Forecasting stock returns:a time-varying robust weighted least squares ap-proach[J].Journal of Management Sciences in China,2024,27(1):141-158.

[5] Masini R P,Medeiros M C,Mendes E F.Machine learning advances for time series forecasting[J].Journal of Economic Surveys,2023,37(1):76-111.

[6] Engle R F,Ghysels E,Sohn B.Stock market volatility and macroeconomic fundamentals[J].Review of Economics and Statistics,2013,95(3):776-797.

[7] Sezer O B,Gudelek M U,Ozbayoglu A M.Financial time series forecasting with deep learning:a systematic literature review:2005-2019[J].Applied Soft Computing,2020,90:106181.

[8] Chen L Y,Pelger M,Zhu J.Deep learning in asset pricing[J].Management Science,2024,70(2):714-750.

[9] 李斌,龙真.中国股票市场可预测性研究:基于机器学习的视角[J].管理科学学报,2023,26(10):138-158.
Li B,Long Z.Return predictability in the Chinese stock markets:a machine learning per-spective[J].Journal of Management Sciences in China,2023,26(10):138-158.

[10] 李斌,屠雪永.基于机器学习和资产特征的投资组合选择研究[J].系统工程理论与实践,2024,44(1):338-359.
Li B,Tu X Y.Portfolio selection based on machine learning and asset characteristics[J].Systems Engineering-Theory & Practice,2024,44(1):338-359.

[11] Tang Y J,Song Z Y,Zhu Y L,et al.A survey on machine learning models for financial time series forecasting[J].Neurocomputing,2022,512:363-380.

[12] 范小云,王业东,王道平,等.不同来源金融文本信息含量的异质性分析:基于混合式文本情绪测度方法[J].管理世界,2022,38(10):78-95.
Fan X Y,Wang Y D,Wang D P,et al.Heterogeneity analysis of information content for financial text from different sources:a hybrid text sentiment measurement method[J].Journal of Management World,2022,38(10):78-95.

[13] 姜富伟,刘雨旻,孟令超.大语言模型、文本情绪与金融市场[J].管理世界,2024,40(8):42-59.
Jiang F W,Liu Y M,Meng L C.Large language model and textual sentiment analysis in Chinese stock markets[J].Management World,2024,40(8):42-59.

[14] 张宗新,吴钊颖.媒体情绪传染与分析师乐观偏差:基于机器学习文本分析方法的经验证据[J].管理世界,2021,37(1):170-185.
Zhang Z X,Wu Z Y. Media sentiment contagion and analysts' optimistic bias: empirical evidence based on machine learning text analysis[J].Management World,2021,37(1):170-185.

[15] 姚加权,冯绪,王赞钧,等.语调、情绪及市场影响:基于金融情绪词典[J].管理科学学报,2021,24(5):26-46.
Yao J Q,Feng X,Wang Z J,et al.Tone,sentiment and market impacts:The construction of Chinese sentiment dictionary in finance[J].Journal of Management Sciences in China,2021,24(5):26-46.

[16] Devlin J,Chang M W,Lee K,et al.BERT:pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics:Human Language Technologies,Volume 1(long and Short Papers).[S.l.:s.n.],2019:4171-4186.

[17] Gan B Q,Alexeev V,Bird R,et al.Sensitivity to sentiment:News vs social media[J].International Review of Financial Analysis,2020,67:101390.

[18] Lin Y,Yan Y,Xu J L,et al.Forecasting stock index price using the CEEMDAN-LSTM model[J].The North American Journal of Economics and Finance,2021,57:101421.

[19] Cao J,Li Z,Li J.Financial time series forecasting model based on CEEMDAN and LSTM[J].Physica A:Statistical Mechanics and its Applications,2019,519:127-139.

[20] Li Y R,Zhu Z F,Kong D Q,et al.EA-LSTM:Evolutionary attention-based LSTM for time series prediction[J].Knowledge-Based Systems,2019,181:104785.

[21] 胡青渝,陈其安.基于先验前馈神经网络的股票市场收益预测[J].系统工程理论与实践,2025,45(10):3287-3303.
Hu Q Y, Chen Q A.Stock market return prediction based on prior feedforward neural networks[J].Systems Engineering-Theory & Practice,2025,45(10):3287-3303.

[22] 戴星宇,王群伟.融合谱分解和傅里叶系数的多元区间值时序模型及其预测应用[J].系统工程理论与实践,2025,45(7):2385-2404.
Dai X Y,Wang Q W.A multi-variate ITS model with spectral decomposition and Fourier-coefficient[J].Systems Engineering-Theory & Practice,2025,45(7):2385-2404.

[23] 陈凯杰,唐振鹏,吴俊传,等.基于分解-集成和混频数据采样的中国股票市场预测研究[J].系统工程理论与实践,2022,42(11):3105-3120.
Chen K J,Tang Z P,Wu J C,et al.Forecasting China's stock index:a hybrid method based on decomposition-integrated and mixed-frequency data[J].Systems Engineering-Theory & Practice,2022,42(11):3105-3120.

[24] 林昱,常晋源,黄雁勇.融合经验模态分解与深度时序模型的股价预测[J].系统工程理论与实践,2022,42(6):1663-1677.
Lin Y,Chang J Y,Huang Y Y.On the prediction of the stock price based on empirical mode decomposition and deep time series model[J].Systems Engineering-Theory & Practice,2022,42(6):1663-1677.

[25] Lundberg S M,Lee S I.A unified approach to interpreting model predictions[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems.[S.l.]:ACM,2017:4768-4777.

[26] Hassija V,Chamola V,Mahapatra A,et al.Interpreting black-box models:a review on explainable artificial intelligence[J].Cognitive Computation,2024,16(1):45-74.

[27] Bandt C,Pompe B.Permutation entropy:a natural complexity measure for time series[J].Physical Review Letters,2002,88(17):174102.

Interpretable machine learning and the economic value of stock index return forecasts:Evidence from Chinese stock market

Li Yan

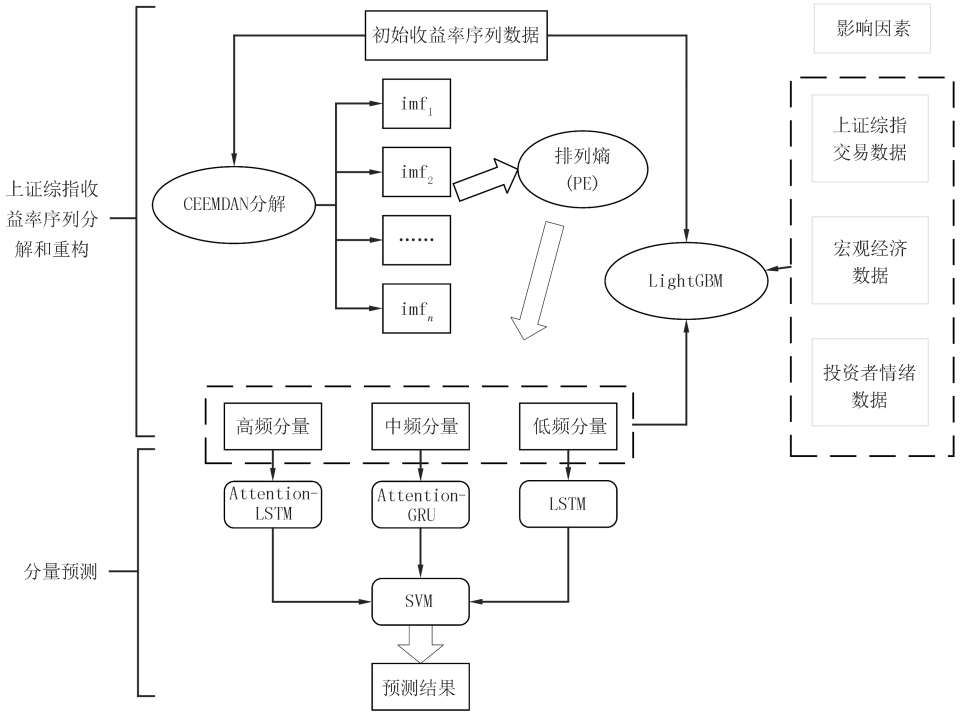
(School of Business, Henan University, Kaifeng 475004, China)

Abstract: This paper builds a multimodal hybrid framework for the Shanghai Composite index and the Shenzhen Component index that aims to be both predictive and interpretable. On the input side it draws on four kinds of information—macroeconomic conditions, market trading, corporate earnings, and investor sentiment. The modeling proceeds in stages: CEEMDAN first decomposes the return series by scale, LightGBM selects the key features within each frequency band, Attention-LSTM, Attention-GRU, and LSTM then fit the high-, medium-, and low-frequency components respectively, and SVR carries out the nonlinear ensemble. The hybrid model attains a positive adjusted external examination R^2 in both markets-0.48% for Shanghai and 1.23% for Shenzhen-outperforming linear benchmarks and a single deep-learning model. SHAP analysis shows that the predictions are driven mainly by monetary and price-related macroeconomic variables. On the timing side, under a long-cash strategy with a 0.1% transaction cost, the hybrid model improves risk-adjusted performance relative to buy-and-hold. Taken together, these results offer evidence from the Chinese market on the interpretability and economic value of multimodal machine learning for stock index return forecasting.

Keywords: stock index return forecasting; multimodal data; interpretable machine learning; SHAP; market timing

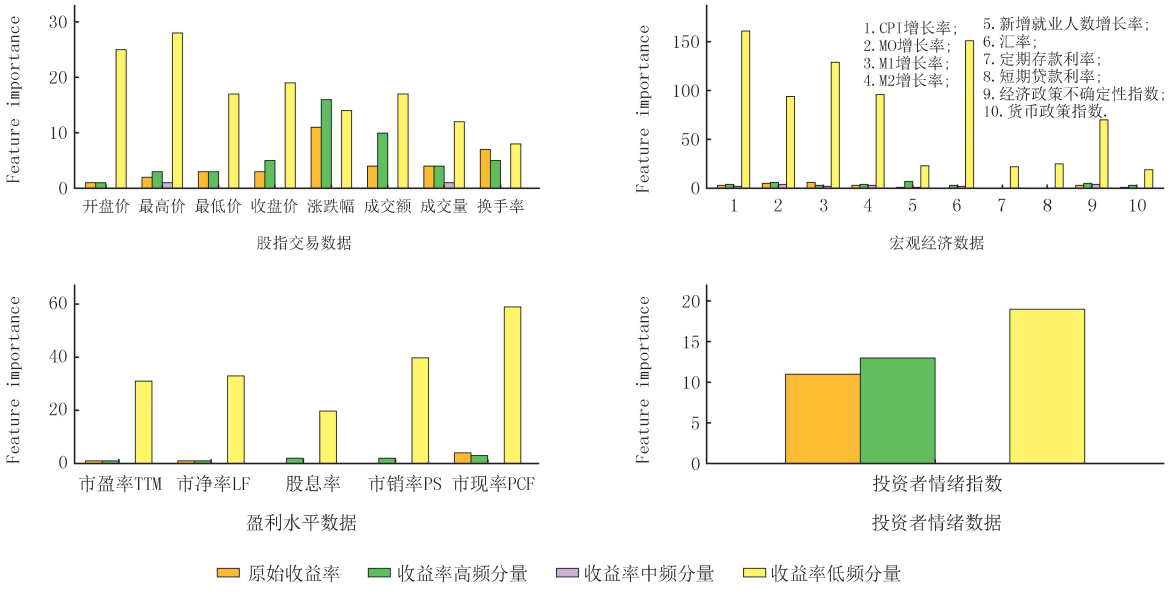
[责任编校 陈留院 杨浦]

附　录



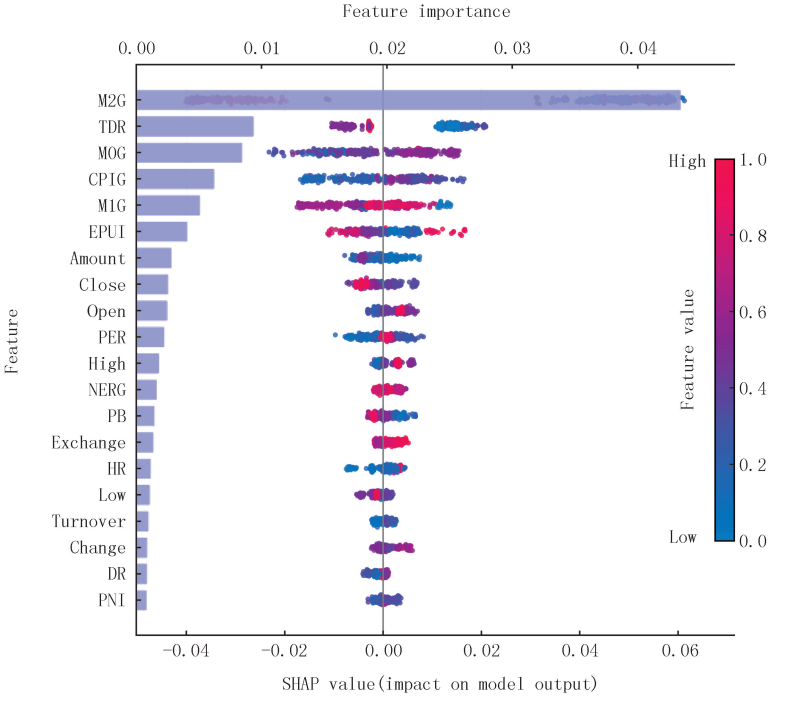
图S1　CEEMDAN-Attention-LSTM/GRU-SVR模型结构图

Fig.S1　CEEMDAN-Attention-LSTM/GRU-SVR hybrid forecasting framework

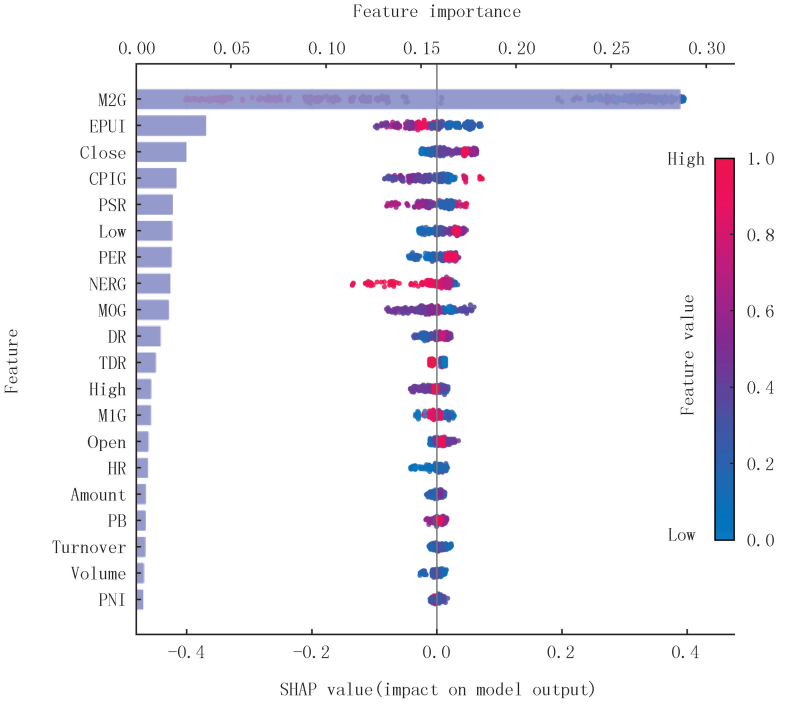


图S2　LightGBM 特征重要度结果(上证市场)

Fig.S2　LightGBM feature importance results(SH)



(a)Mix model



(b)LSTM

图S3　混合模型的SHAP值

Fig.S3　SHAP value impact on model output