

中国贵金属期货市场的波动率预测—— 基于梯度提升树模型和可解释性工具的 融合研究

冯凌秉^{1,2}, 黄达森³, 郑宇豪⁴

(1. 江西财经大学统计与数据科学学院, 南昌 330013; 2. 江西财经大学国势研究院, 南昌 330013;
3. 江西财经大学数字经济学院, 南昌 330013; 4. 江西财经大学金融学院, 南昌 330013)

摘要 黄金和白银由于其独特的金融属性, 成为投资和资产保值的首选, 其价格波动的定量描述和预测对投资者的风险管理决策至关重要. 本文引入了丰富的特征变量, 采用向前一步滚动算法来预测上海黄金和白银贵金属期货的已实现波动率 (RV), 并比较了多个机器学习模型在不同损失函数和评估方法下的预测效果. 结果表明, 梯度提升决策树 (GBDT) 模型在贵金属期货市场的预测中表现优异. 此外, 本研究融合 XGBoost 模型和可解释性工具, 分析了贵金属期货市场的特征变量对预测值的动态贡献, 并评估了重要变量对预测性能的异质性影响. 研究揭示了市场情绪变量的关键作用, 以及宏观经济变量与波动率分解变量在不同市场条件下的相对贡献. 本文的发现为贵金属期货市场波动预测的因子和模型选择提供了清晰证据, 为该市场的投资者和监管者提供了可信的投资和管理建议.

关键词 贵金属期货; 波动率预测; 梯度提升决策树; 模型可解释性

收稿日期: 2024-06-20

基金项目: 国家自然科学基金青年项目 (72001098); 中国博士后科学基金面上项目 (2024M761224); 江西省自然科学基金面上项目 (20242BAB25122)

Supported by National Natural Science Foundation of China (72001098); China Postdoctoral Science Foundation (2024M761224); Natural Science Foundation of Jiangxi Province (20242BAB25122)

作者简介: 冯凌秉, 副教授, 博士, 研究方向: 波动率建模与预测, 机器学习及其应用, E-mail: fenglb88@gmail.com; 黄达森, 硕士研究生, 研究方向: 金融工程与风险管理, E-mail: hdasen@126.com; 通信作者: 郑宇豪, 硕士研究生, 研究方向: 金融工程与风险管理, E-mail: ehoole1203@foxmail.com.

Forecasting Volatility in China's Precious Metals Futures: Integrating Gradient Boosting Decision Trees with Interpretability Tools

FENG Lingbing^{1,2}, HUANG Dasen³, ZHENG Yuhao⁴

- (1. School of Statistics and Data Science, Jiangxi University of Finance and Economics, Nanchang 330013, China; 2. Institute of Statistics, Jiangxi University of Finance and Economics, Nanchang 330013, China; 3. School of Digital Economics, Jiangxi University of Finance and Economics, Nanchang 330013, China; 4. School of Finance, Jiangxi University of Finance and Economics, Nanchang 330013, China)

Abstract Gold and silver, due to their unique financial properties, have become preferred choices for investment and asset preservation. Accurately quantifying and predicting their price fluctuations is crucial for investors' risk management decisions. This paper introduces a rich set of feature variables and employs a forward rolling algorithm to forecast the realized volatility (RV) of gold and silver futures in Shanghai. We compare the performance of various machine learning models under different loss functions and evaluation methods. The results indicate that the gradient boosting decision tree (GBDT) models demonstrate superior performance in forecasting the futures market for precious metals. Furthermore, this study integrates the XGBoost model with interpretability tools to analyze the dynamic contributions of feature variables to the predicted values in the precious metals futures market. It also assesses the heterogeneous impact of significant variables on predictive performance. Our findings reveal the critical role of market sentiment variables, as well as the relative contributions of macroeconomic variables and volatility decomposition variables under different market conditions. The research provides clear evidence for the selection of factors and models in forecasting precious metal futures market volatility, offering credible investment and management recommendations for investors and regulators in this market.

Keywords precious metal futures; volatility forecasting; gradient boosting decision tree; model interpretability

1 引言

黄金和白银作为重要贵金属, 兼具商品和金融资产的属性, 在投资和消费等领域用途广泛 (郭杨莉和马锋, 2024; Huang et al., 2014; Lanzano et al., 2006). 其期货价格波动与商品、股票和货币等多个市场密切相关, 受到市场供需、宏观经济、地缘政治、货币政策调整以及国际贸易环境等多重因素的影响, 呈现出高度波动性 (Batten et al., 2010; Dinh et al., 2022). 波动性作为衡量资产回报风险的关键指标, 其未来趋势的预判不仅对投资决策具有指导作用, 也是金融评估与风险管理的重要依据. 提高贵金属期货价格波动预测的准确性并解释相关变量的内在联系, 对投资者、交易员、套期保值者及其他市场参与者制定合理的投资策略和风险管理框架具有重要的现实意义.

早期的波动率预测主要依赖于 GARCH 族模型 (庞淑娟等, 2011; Bollerslev, 1986; Maree

et al., 2017), 大多数模型是广义自回归条件异方差 (GARCH) 模型的变体. 这些模型通常基于月度、周度和日度等低频数据, 但忽略了日内交易的宝贵信息, 可能导致预测偏差. 随着日内高频数据广泛可得, Andersen and Bollerslev (1998) 引入了已实现波动率 (realized volatility, RV) 的概念, 将其定义为日内高频收益率的平方和. RV 更能反映市场波动, 适合作为真实波动率的代理. Corsi (2009) 提出的异质性自回归已实现波动率 (HAR-RV) 模型成功捕捉了 RV 的长记忆性特征, 表现出良好的预测性能. 机器学习模型在波动率预测方面表现更优异, 能够更灵活地处理非线性数据, 深入挖掘预测因子中的信息 (Christensen, 2023; Rahimikia and Poon-H, 2020). 已有研究证明, LASSO (Zhou and Lu, 2023), Ridge (Wei et al., 2020), ElasticNet (Liang et al., 2022) 以及 SVM (Santamaria-Bonfil et al., 2013) 等机器学习模型均提升了金融市场波动率的预测效果.

现有金融预测研究逐步从传统机器学习模型扩展至深度学习与集成学习模型. 深度学习模型如卷积神经网络 (CNN), 长短期记忆网络 (LSTM), 在多项金融预测中表现出色 (Shu et al., 2023). 尤其是其中的集成学习模型, 在多项时间预测任务中展现出了稳健的预测性能 (Bâra and Oprea, 2024; Mohsin, 2023). 集成学习模型主要包括 Bagging、Boosting 和 Stacking 等算法. Elsayed et al. (2021) 发现在 9 个不同领域的数据集上, 简单的提升树集成算法的效果优于深度学习模型. Makridakis et al. (2022) 则指出, 基于 Boosting 的梯度提升决策树算法 (gradient boosting decision tree, 以下简称为 GBDT), 如 XGBoost、LightGBM 和 CatBoost 等, 在全球知名时间序列预测赛事中展现了巨大的预测潜力. Truong et al. (2024) 指出, GBDT 的并行算法和缓存优化技术显著提升了训练效率, 较深度学习模型耗时更短, 并展现了在金融预测中的优势.

最新的研究也指出, 虽然机器学习技术已经取得了显著进展, 但在面对复杂多变的金融市场, 仍需进一步研究以提高波动率预测的准确性和鲁棒性 (Feng et al., 2024). 目前, 在原油、农产品和股价预测中, GBDT 均实现了最佳的预测性能 (胡越等, 2023; Guliyev and Mustafayev, 2022; Yun et al., 2021). 在黄金月度价格预测中, GBDT 模型成效也十分显著 (Jabeur et al., 2024), 但对于颇为重要的中国贵金属期货市场, GBDT 模型的效果到底如何, 还鲜有研究进行过全面和系统的探索. 同时, 因为 GBDT 模型的“黑箱”特征, 大量的已有研究也较少有对其优异性能的驱动因素进行解释性的研究.

基于此研究方面的空白现实, 本文基于 GBDT 模型, 对上海贵金属期货波动率进行了预测与解释相结合的综合研究, 涵盖以下几个关键方面. 首先, 本文选择上海黄金和白银期货为研究标的, 基于夜盘交易实施近九年的高频数据, 采用向前一步滚动预测算法和多种损失函数与评估方法, 从不同角度评估各类 GBDT 模型在贵金属市场中预测效果的细微差异. 其次, 考虑到机器学习模型的性能受到超参数调优、损失函数的选择以及输入变量信噪比的影响 (Rahimikia and Poon, 2020), 本研究在模型构建阶段引入了多样化的预测变量, 并采用自动调参技术优化部分模型的性能. 同时, 运用稳健性检验和高低波动率分析, 验证 GBDT 模型预测精度的优越和稳健性及其异质性特征. 最后, 本研究融合模型预测性能的提升和模型可解释性工具, 采用滚动算法动态调整模型, 并生成有序的 SHAP 值 (SHapley Additive exPlanation) 输出, 以反映预测性能动态特征和趋势. 不同于 Jabeur et al. (2024) 和 Guliyev and Mustafayev (2022) 单步预测分析特征变量的 SHAP 值, 我们基于滚动预测

计算特征变量的 SHAP 值, 主要研究在滚动场景下特征变量 SHAP 值解释的稳定性. 通过 SHAP 值深入探讨 GBDT 模型在提升上海贵金属期货市场波动率预测效果中的驱动因素, 不仅揭示了不同预测变量对期货价格波动的动态影响, 还明确了提升模型预测性能的关键因素. 这些结论和发现进一步丰富了贵金属期货波动率的建模和解释性研究, 为该市场的参与者和监管者在制定投资组合和风险管理政策方面提供了可靠的风险建模支撑, 并从融合可解释性角度为波动率预测领域注入了新的活力.

余文的结构安排如下: 第二部分介绍研究方法; 第三部分主要列举预测使用的指标和数据; 第四部分呈现实证结果, 包含对各类模型预测性能的比较和稳健性检验等; 第五部分深入探究模型的可解释性分析结果; 最后, 第六部分对全文进行总结.

2 研究方法

2.1 预测模型

2.1.1 HAR-RV 模型

考虑到上海期货交易所 (Shanghai Futures Exchange, SHFE) 的交易时间, 黄金和白银期货的每个交易日可以划分为夜盘交易和日盘交易两个时段. 具体而言, 夜盘交易时间为 21:00–2:30, 而日盘交易时间分为 9:00–11:30 和 13:30–15:00 两个时段. 根据 SHFE 的交易规则, 从第 $t-1$ 天的 21:00 开始的夜盘交易时段被视为交易日 t . 参考 Yao et al. (2021) 的做法, 本文将 RV 定义为:

$$RV_t = \sqrt{\sum_{j=1}^{M_1} (r_{t-1,j}^n)^2 + (r_{t-1}^o)^2 + \sum_{j=1}^{M_2} (r_{t,j}^d)^2}. \quad (1)$$

式 (1) 中, $(r_{t-1,j}^n)^2$ 指的是在 $t-1$ 天中第 j 个夜间交易区间的对数收益率, $(r_{t-1}^o)^2$ 指的是 $t-1$ 日的隔夜收益, $(r_{t,j}^d)^2$ 指的是 t 天中第 j 个日间交易区间的对数收益率. M_1 是夜间交易的采样频率, M_2 是日内交易的采样频率. 通常, $r_{t-1,j}^n = \ln(p_{t-1,j} - \ln_{t-1,j-1}) \times 100$, $p_{t-1,j}$ 是资产价格.

Corsi (2009) 提出的 HAR-RV 模型被广泛应用于预测领域, 因此本文将 HAR-RV 模型作为本研究的基准模型 (Bench). HAR-RV 模型可以直接定义为:

$$RV_t = \beta_0 + \beta_d RV_{t-1} + \beta_w RV_{t-1}^w + \beta_m RV_{t-1}^m + \varepsilon_t. \quad (2)$$

式 (2) 中, RV_t 是当期日度 RV, RV_{t-1} 是滞后 1 期的日度 RV, RV_{t-1}^w 是周度 RV (5 日平均), RV_{t-1}^m 是月度 RV (22 日平均), ε_t 是误差项.

HAR-RV 模型的本质是使用日度、周度和月度的 RV 形成的多元线性回归, 进而允许使用最小二乘法估计系数.

2.1.2 变量选择模型

在传统机器学习中, 变量选择技术旨在通过降低模型复杂性来解决过拟合问题. Tibshirani (1996) 提出的 LASSO 模型通过在多元线性回归的损失函数中加入 L_1 范数正则项, 实现了变量选择和模型收缩. Hoerl and Kennard (1970) 提出的 Ridge 模型是一种在多元线

性回归的损失函数上通过添加 L_2 范式正则项的收缩方法. Zou and Hastie (2005) 提出的 ElasticNet 模型结合了 LASSO 和 Ridge 回归的优点, 利用 L_1 和 L_2 正则化共同实现对模型的正则化处理.

2.1.3 支持向量回归模型

支持向量回归 (support vector regression, SVR) 是一种用于回归任务的机器学习模型. 与传统回归算法相比, SVR 在目标函数两侧创建一个“间隔带”, 忽略落入该间隔带内的样本误差, 并通过最小化总损失和最大化间隔来得出优化后的模型. 因此, SVR 适用于数据量较小及含噪声数据的回归问题, 其核心思想是将问题转化为一个二次规划问题, 通过求解这一问题, 得到超平面的参数.

2.1.4 集成树模型

本文使用的集成树模型包含随机森林 (random forest, RF) 模型和 GBDT 类模型 (本文指 XGBoost、LightGBM 和 CatBoost 模型).

1) 随机森林模型

随机森林 (RF) 模型 (Breiman, 2001) 是一种集成学习算法, 由多个决策树组成. 它的核心思想是通过随机抽取数据和特征的子集来构建多个决策树, 然后利用这些决策树的结果在投票后用于预测.

2) XGBoost 模型

XGBoost (New et al., 2016) 是一种 GBDT 框架的集成学习算法. 与传统的 GBDT 方法仅利用一阶导数信息不同, XGBoost 对损失函数进行了二阶泰勒展开, 并在目标函数中加入了正则化项. 这种策略通过综合考虑目标函数的下降和模型复杂性的平衡来求解最优解, 有效避免了过拟合问题, 并提高了模型的求解效率. XGBoost 是由 k 个基模型组成的加法运算式:

$$\hat{y}_i^t = \sum_{t=1}^k f_t(x_i) = \hat{y}_i^{t-1} + f_t(x_i). \quad (3)$$

\hat{y}_i^t 是 t 次迭代后第 i 个样本的预测值, $f_t(x_i)$ 是第 t 棵树的预测值.

损失函数可表示为:

$$L = \sum_{i=1}^n l(y_i, \hat{y}_i). \quad (4)$$

目标函数由损失函数与降低复杂程度的正则项 Ω 构成:

$$\text{Obj}^t = \sum_{i=1}^n l(y_i, \hat{y}_i^{t-1} + f_t(x_i)) + \sum_{t=1}^k \Omega(f_t). \quad (5)$$

解 XGBoost 模型需要最优化目标函数, 即是求解 $f_t(x_i)$. 图 1 展示 XGBoost 算法计算预测结果的可视化过程.

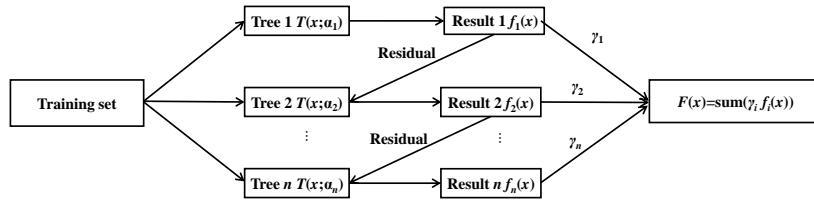


图 1 XGBoost 算法示意图

3) LightGBM 模型

LightGBM (Ke et al., 2017) 是由微软主持的机器学习项目, 在开源后在数据预测竞赛中表现卓越. 其设计目标是实现更快的运算速度、更小的内存占用和更优越的预测性能. LightGBM 采用了多种算法来提升预测性能, 其中包括以下关键技术: 首先, 采用直方图算法来有效处理连续特征. 其次, 基于梯度的单边采样 (GOSS) 是一种优化方法, 用于数据下采样, 以提高模型的效率. 最后, 互斥特征捆绑算法 (EFB) 用于处理高维稀疏数据. 通过这些算法的协同作用, LightGBM 实现了高效的梯度提升框架, 使其在大规模数据集上表现出色. 因此, LightGBM 在许多实际应用中得到了广泛的应用. LightGBM 算法见图 2.

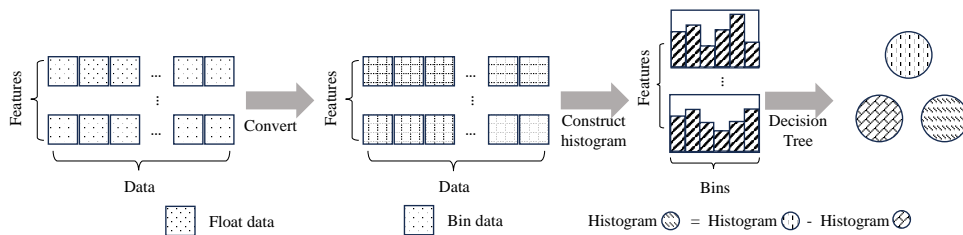


图 2 LightGBM 算法示意图

4) CatBoost 模型

CatBoost 是由俄罗斯公司 Yandex 发起的一个开源项目, 其主要优势在于高效处理分类特征. 相较于其他 Boosting 算法, CatBoost 模型通过以下方法提升了预测性能: 首先, 在处理类别特征时, CatBoost 会考虑特征组合, 并采用贪心策略构造新的分裂点. 在分裂过程中, CatBoost 将类别特征组合起来, 并转换为数值编码, 从而形成动态的数值特征. 其次, CatBoost 应用排序提升算法来处理梯度偏差和预测偏移问题. 在排序提升算法中, CatBoost 会针对不同的样本训练独立的模型, 这些模型不受之前样本的估计梯度影响. 模型使用单独训练的模型梯度来训练基学习器. 最后, 在基学习器的使用方面, CatBoost 采用对称树作为基学习器. 在对称树中, 相同的分割准则在树的整个层次上使用. 由于这种树是平衡的, 它能够有效减轻过拟合并提高运行速度.

2.2 损失函数和评估方法

2.2.1 损失函数

在评估模型预测精度时, 采用不同的损失函数可能会导致评价结果出现显著差异 (魏宇, 2010). 文献中学者常使用均方根误差 (RMSE)、平均绝对误差 (MAE) 和拟极大似然损失

(QLIKE) 中的一个或两个来对预测误差进行衡量. 而本研究引入了多个损失函数以更加全面和客观地衡量模型性能. 除了继承使用传统的 RMSE、MAE 和 QLIKE 损失函数, 本文还纳入了累积平方预测误差 (CumSFE), 以期更进一步捕捉模型间预测能力的细微差别. 以上四个损失函数的计算方式如下:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (\widehat{\text{RV}}_{i,t}^F - \text{RV}_t)^2}, \quad (6)$$

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n \left| \widehat{\text{RV}}_{i,t}^F - \text{RV}_t \right|, \quad (7)$$

$$\text{QLIKE} = \frac{1}{n} \sum_{t=1}^n \left(\ln(\widehat{\text{RV}}_{i,t}^F) + \frac{\text{RV}_t}{\widehat{\text{RV}}_{i,t}^F} \right), \quad (8)$$

$$\text{CumSFE} = \sum_{t=1}^n \left((\widehat{\text{RV}}_{i,t}^F - \text{RV}_t)^2 - (\widehat{\text{RV}}_t^B - \text{RV}_t)^2 \right), \quad (9)$$

其中 $\widehat{\text{RV}}_{i,t}^F$ 表示第 i 个模型在 t 日的预测 RV, RV_t 表示 t 日实际 RV 值, $\widehat{\text{RV}}_t^B$ 表示 HAR-RV 模型预测的 RV 值, 以 n 表示预测区间的长度, 即预测天数.

在以上损失函数中, RMSE 和 MAE 都是与原观测值量纲相同的精度描述, 数值越小, 表示模型的预测效果越好. RMSE 通过平方形式放大了较大预测误差 (指 $|\widehat{\text{RV}}_{i,t}^F - \text{RV}_t| > 1$) 的影响, 从而在模型间显著区分出预测误差较大的值; 而 MAE 对所有的预测误差赋予等同的权重, 不会因误差大小而有所偏重. 因此, 如果数据集中的误差相对均匀, RMSE 可能是一个更好的选择, 因为它能够提供关于误差值大小的更多信息, 更容易解释. QLIKE 是金融领域中评估波动率的常用损失函数, 它的误差惩罚特点是不对称和不成比例的. 当模型同时产生正负误差时, 相同幅度的负误差受到的惩罚会大于相同比例的正误差, 即它在低波动区间对点预测精确度的判断更加严格. 为检验机器学习模型预测在时间序列上预测效果的稳健性, 将使用 CumSFE 来绘制预测评估图, 以直观地展示模型在不同时间点的预测表现.

2.2.2 模型性能评估方法

1) 样本外 R^2

除了使用损失函数对模型性能评估外, 本研究还特别重视模型的样本外预测性能. 因此, 我们采用样本外 R^2 (R_{oos}^2) 来衡量其他模型相对于基准模型的均方预测误差 (MSFE) 降低的比例. 若 $R_{\text{oos}}^2 > 0$, 则表示在 MSFE 方面, 待评估模型的预测性能优于 HAR-RV 模型. 为验证结果的显著性, 本文使用了 CW 统计量 (Clark and West, 2007) 进行检验. 本文中的 CW 统计检验的原假设是基准模型 (即 HAR-RV 模型, 下文表格用 Bench 表示) 的 MSFE 小于或等于待评估模型.

$$R_{\text{oos}}^2 = 1 - \frac{\sum_{t=1}^n (\widehat{\text{RV}}_{i,t}^F - \text{RV}_t)^2}{\sum_{t=1}^n (\widehat{\text{RV}}_t^B - \text{RV}_t)^2}. \quad (10)$$

2) Doc 检验

方向性检验 (direction-of-change, Doc) 也常被用来评估样本外预测性能. 根据 Degiannakis and Filis (2017) 的研究, Doc 检验在市场参与者进行资产配置方面发挥着重要作用. 该检验衡量的是正确预测波动趋势的能力. 首先设定一个虚拟变量 p_t , 如果模型正确预测了市场在第 t 个交易日的波动趋势, 则取值为 1, 否则取值为 0, 虚拟变量 p_t 可以表示为:

$$p_t = \begin{cases} 1, & \text{RV}_t > \text{RV}_{t-1} \cap \widehat{\text{RV}}_t^F > \text{RV}_{t-1}, \\ 1, & \text{RV}_t < \text{RV}_{t-1} \cap \widehat{\text{RV}}_t^F < \text{RV}_{t-1}, \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

正确预测波动率变动方向的预测值占比可以定义为: $\text{Doc} = 1/n \sum_{t=m+1}^{m+n} p_t \times 100$. 我们利用 Pesaran and Timmermann (1992) 提出的非参数测试检验原假设, 即预测模型的 Doc 率小于或等于随机游走的 Doc 率.

3) MCS 检验

为了进一步考察不同波动率模型在预测性能上的优劣, 本文采用了 Hansen et al. (2011) 提出的模型置信集 (model confidence set, MCS) 检验方法. 在 MCS 检验时, 本文综合应用了 MSE、MAE 以及 R2LOG 三种损失函数作为模型性能的评价指标. MCS 检验基于等价性检验与逐步剔除原则, 其核心在于首先检验所有模型在预测性能上是否等价. 当拒绝原假设 (即所有模型性能等价) 时, MCS 检验将依据消除原则逐步排除性能最差的模型, 直至无法拒绝等价性假设为止. 这一过程最终产生一个包含指定置信水平的最优模型集合. 在本文中, 我们设置了两组不同的置信水平 Alpha, 分别为 0.1 和 0.2. 当某一模型的 P 值大于设定的置信水平 (0.1 或 0.2) 时, 意味着该模型在样本外预测性能上表现优异, 因而被纳入模型置信集. 值得注意的是, P 值的大小与模型预测精度成正比, 即 P 值越大, 对应模型的预测精度越高. 为了更全面地评估模型的预测性能, 本文采用了范围统计量 (range statistic, T_R) 和半二次统计量 (semi-quadratic statistic, T_{SQ}), 其定义如下:

$$T_R = \max_{u,v \in M} \frac{|\bar{d}_{i,uv}|}{\sqrt{\text{var}(d_{i,uv})}}, \quad T_{SQ} = \max_{u,v \in M} \frac{(\bar{d}_{i,uv})^2}{\sqrt{\text{var}(\bar{d}_{i,uv})}}$$

M 代表模型集合, $u, v \in M$, $\bar{d}_{i,uv}$ 为损失函数差值的均值, $\bar{d}_{i,uv} = m^{-1} \sum_{t=1}^m d_{i,uv,t}$.

通过以上评估方法全面验证各机器学习模型的运行效果, 本文将深入阐明模型的优劣, 提供稳健的解释结果, 为市场预测和模型选择提供科学依据. 具体来说, R_{OOS}^2 能够有效评价机器学习模型在点预测方面相较基准模型的预测表现, 着重于从性能提升的角度考察模型对预测误差的减少程度. Doc 率则主要是从趋势预测的角度评估模型, 关注模型在预测下一个交易日波动方向的准确性. MCS 主要的比较方式是模型竞争, 通过自举法 (bootstrap) 在多次运行后筛选出表现最优的模型, 并基于结果对模型进行排名和淘汰. 以上指标和方法在先前研究波动率的文献中已被广泛采用 (梁超等, 2022; 马锋等, 2015; Feng et al., 2024; Fu et al., 2024; Liang et al., 2022).

2.2.3 经济价值评估方法

除了统计评估,市场参与者和决策者同样重视预测模型在经济收益上的表现.为揭示模型应用价值,本文引入 Bollerslev et al. (2018) 提出的已实现效用框架进行经济价值分析.该框架仅依赖于 RV 预测,旨在量化具有平均方差偏好的投资者在投资于具有时变波动率和恒定夏普比率特征的资产时的效用收益.本文使用这个框架来预测贵金属期货的平均预期效用,具体衡量方法如下:

$$\bar{U}(\widehat{RV}_{i,t}^F) = \frac{1}{n+1} \sum_{t=1}^n \frac{SR^2}{\gamma} \left(\frac{\sqrt{RV_t}}{\sqrt{\widehat{RV}_{i,t}^F}} - \frac{1}{2} \frac{RV_t}{\widehat{RV}_{i,t}^F} \right), \quad (12)$$

其中, γ 和 SR 分别代表风险规避系数和夏普比.根据 Bollerslev et al. (2018) 和 Guo et al. (2023) 的研究,本文设定 SR 为 0.4,并考虑三个不同的 γ 值,即 2、3 和 6.当 SR 为 0.4 且 γ 分别为 2、3 和 6 时,如果模型能够完全预测 RV,则最大平均期望效用分别为 4.000%、2.667% 和 1.333%.

2.2.4 向前一步滚动预测算法

本研究对所有模型,包括 HAR-RV 和所有机器学习模型,均采用了向前一步滚动预测算法(魏宇, 2010),预测的工作分析流程见图 3.以下是预测过程的详细步骤:

1) **初始化:** 将样本的最后 20% 定义为测试集,时间跨度从 2020 年 5 月 27 日至 2022 年 2 月 11 日;设定滚动窗口长度为样本总量的 80%,即每个滚动窗口包含 1675 个样本,用于预测下一个交易日的 RV;

2) **首次预测:** 以 2013 年 7 月 8 日至 2020 年 5 月 26 日的数据,作为训练集 1,对模型进行训练.通过将预设的模型超参数输入到模型,并采用五折交叉验证方式,基于样本内 RMSE 最小化原则选择最优模型.利用选定的模型和已知的外生预测变量,估计测试集的第一个样本(初始预测点为样本 1676).根据预测结果和模型,生成对应的 VIP 和 SHAP 值(变量解释详见 3.2 节);

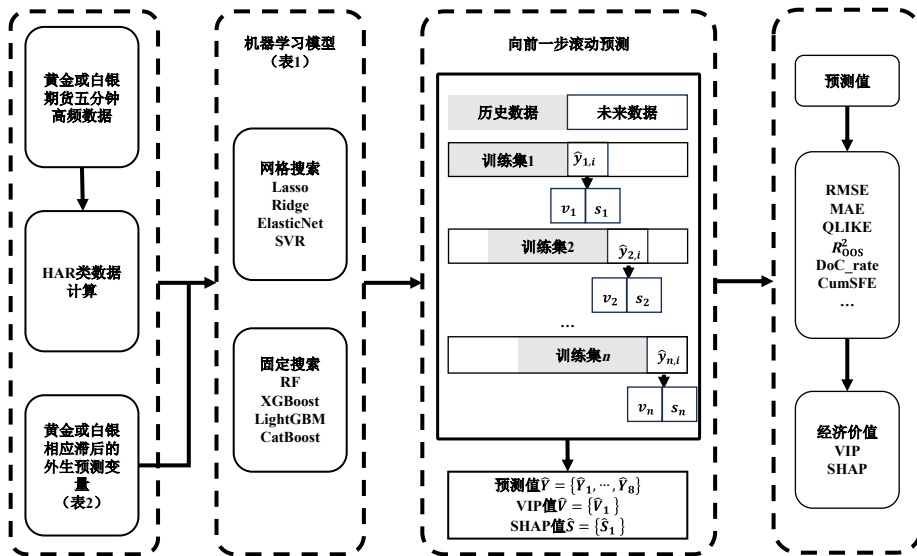
3) **更新:** 接着,将滚动窗口向前移动一天,得到训练集 2,然后重复训练步骤,对测试集的第二个样本进行预测(预测点为样本 1677);

4) **重复步骤:** 持续执行上述步骤,利用每一步选出的模型生成相应的预测值,直至完成对测试集中所有样本的预测.

本文通过这种滚动更新的方式,逐步推进预测流程,以适应数据的动态变化.最终,通过比较所有测试集的预测值与实际值,评估模型的样本外预测性能.

2.3 超参数调优

机器学习模型的性能受到超参数的调优的影响(Rahimikia and Poon, 2020).本研究采用网格搜索方法系统遍历参数空间的所有组合,通过评估每种组合下的 RMSE 来选取最佳参数配置,以优化模型.针对 LASSO、Ridge、ElasticNet 和 SVR 这四个模型,设定明确的参数搜索范围并实施超参数调优.对于集成树模型(RF、XGBoost、LightGBM 和 CatBoost),由于自动参数优化过程耗时较长,本研究选择了手动调参策略,以寻找能够显著提高模型预



注: \hat{V}_1 和 \hat{S}_1 分别代表模型预测结果生成的 VIP 和 SHAP 值, 本文主要生成 XGBoost 对应的解释值.

图 3 机器学习预测分析工作流程

测能力的参数组合, 并在后续的分析中, 固定了这四个集成树模型的参数. 这样, 如果集成树模型在固定参数下仍能保持稳定的预测性能, 甚至优于其他模型, 则更能凸显其优势. 本研究包含模型训练在内的所有程序均在 R 语言上实现, 表 1 给出了机器学习模型的参数设置.

2.4 模型可解释性方法

本文使用 Greenwell and Boehmke (2020) 提出的方法生成特征变量在预测中的变量重要性 (variable importance in projection, VIP), 以及通过 SHAP 值衡量特征变量在预测中的贡献程度. SHAP 值最早由 Shapley (1953) 提出, 是合作博弈论中的一个概念, 用来解释机器学习模型的输出 (lundberg and Lee, 2017). SHAP 值通过量化每个特征对模型预测的贡献, 提供了一种深入解释模型决策过程的方法. 其优势在于, 不仅揭示特征对模型预测的贡献程度, 还通过正负值反应特征变量对预测值的积极或消极影响. SHAP 值定义如下:

$$\phi_i = \sum_{s \subseteq N \setminus \{i\}} \frac{|S|!(M - |S|! - 1)}{M!} [f_x(S \cup \{i\}) - f_x(S)]. \quad (13)$$

ϕ_i 为 SHAP values, N 为输入的特征的集合, M 为输入特征的个数, S 为输入特征的子集 (lundberg and Lee, 2018).

本文将 VIP 和 SHAP 的输出与滚动预测相融合. 根据图 3 所示的滚动预测工作流程, 在模型输出每一步预测值 $\hat{y}_{n,t}$ 后, 继续计算得到了所有特征值的 VIP 和 SHAP 值. 这种融合方法不仅能够逐步跟踪预测值的变化, 还能动态评估各特征在每一步预测中的重要性和贡献, 提供了更为细致的模型解释和分析工具.

表 1 机器学习模型的参数选择

模型	待优化参数	参数选择
LASSO	惩罚系数 (penalty)	penalty = $10^{\text{seq}(-2, -0.1, \text{length.out} = 20)}$ ⁽¹⁾
Ridge	惩罚系数 (penalty)	penalty = seq(0.1, 0.6, length = 20)
ElasticNet	L1 比 (mixture), 惩罚系数 (penalty)	penalty(), mixture(), size = 30 ⁽²⁾
SVR	损失 (cost), 径向基函数 (rbf_sigma)	cost (), rbf_sigma(), size = 30
RF	随机采样的数量 (mtry), 树的棵树 (trees), 分裂节点中最小样本数 (min_n)	mtry = 11, min_n = 10, trees = 950
XGBoost	树的最大深度 (tree_depth), 学习率 (learn_rate), 分裂节点中最小样本数 (min_n), 随机采样的数量 (mtry) 等	mtry = 8, min_n = 8, trees = 1000, tree_depth = 9, learn_rate = 0.02
LightGBM	树的棵树 (trees), 树的最大深度 (tree_depth), 学习率 (learn_rate), 分裂节点中最小样本数 (min_n), 随机采样的数量 (mtry) 等	mtry = 7, min_n = 7, trees = 500, tree_depth = 7, learn_rate = 0.019
CatBoost	树的棵树 (trees), 树的最大深度 (tree_depth), 学习率 (learn_rate), 分裂节点中最小样本数 (min_n), 随机采样的数量 (mtry) 等	mtry = 7, min_n = 7, trees = 500, tree_depth = 7, learn_rate = 0.02

注: (1) seq(): 表示生成序列, 如 seq(-2, -0.1, length.out = 20), 生成 -2 到 -0.1 的等差数列, 总共有 20 个元素; (2) size=30: 表示生成 30 组超参数.

3 数据描述

3.1 贵金属期货数据

本研究基于 SHFE 黄金和白银期货的 5 分钟高频数据进行分析, 旨在通过此采样频率有效规避市场微观结构噪声的干扰, 其合理性在过往研究中已得到证实 (Liu et al., 2015). 研究时段覆盖了自夜盘交易实施以来, 从 2013 年 7 月 8 日至 2022 年 2 月 11 日的完整交易数据. 数据来源为 Wind 数据库. 通过长期的数据跨度, 本研究能够深入分析市场长期趋势和动态, 为理解市场行为提供了丰富的情境, 有助于更全面预测市场动向.

3.2 预测变量数据

预测变量的选择并非追求数量上的繁多, 而是要挑选并组合出最合适的变量集合, 这在机器学习变量选择中至关重要, 对提升模型预测性能起关键作用 (Fu et al., 2024). 因此, 借鉴已有研究 (梁超等, 2022; Batten et al., 2010; Dinh et al., 2022; Luo and Ye, 2015; Wei et al., 2020), 本文选取了以下对贵金属波动率有预测能力的变量作为预测变量. 这些变量包括: 全球经济政策不确定指数 (gepu)、中国经济政策不确定指数 (cnepu)、中国贸易不确定指数 (cntpu)、全球地缘政治风险指数 (gpr)、中国消费者物价指数 (cpi)、中国生产价格指数 (ppi)、中国消费者信心指数 (cci)、美元指数 (usdx)、美国联邦基金利率 (effr)、西得克

萨斯中间基原油价格 (wti)、布伦特原油价格 (brent)、中证 1000 指数 (zz1000)、黄金或白银期货日交易量 (vol5、vol10)¹、百度指数 (baidu)²、CBOE 黄金 ETF 波动率指数 (gvz)、CBOE 白银 ETF 波动率指数 (vxslv)³、伦敦金 (londongold) 和伦敦银 (londonsilver) 价格数据。另外, 为了捕捉市场动态和波动性的时间序列特征, 本文进一步引入滞后 1~5 阶的已实现波动率的跳跃成分 (Jt)、已实现正半差 (rsm) 和已实现负半差 (rsp), 以及滞后 1~22 阶的 RV, 作为 RV 的分解变量。上述变量的具体说明见表 2。

表 2 变量描述

预测变量	频率	处理	来源		
RV		滞后 1~22 阶			
rsm	Daily	滞后 1~5 阶	https://www.shfe.com.cn		
rsp					
Jt					
gepu					
cnepu	Monthly	滞后 1 阶	http://www.policyuncertainty.com		
cntpu					
gpr					
ppi					
cci				https://db.cei.cn/jsps/Home	
cpi					1 阶差分并滞后 1 阶
usdx				Daily	
brent					
wti					
effr					
gvz					
vxslv					
baidu	1 阶差分并滞后 1~5 阶	https://index.baidu.com			
londongold		https://www.wind.com.cn			
londonsilver					
vol5					
vol10			滞后 1 阶		

¹交易量指标 vol5 和 vol10 分别是根据前五日和前十日的日平均交易量为基础计算得出, 分别作为黄金和白银期货波动率的预测变量。

²baidu 对应的是“黄金”和“白银”两个关键词的百度搜索指数, 分别用作预测黄金和白银期货波动率的预测变量。

³gvz 和 vxslv, 衡量的分别是黄金和白银 ETF 期权的隐含波动率, 也被称作“恐慌指数”, 是由美国芝加哥期权交易所 (CBOE) 采用相应标的期权和期货交易价格计算, 表示投资者对该标的在未来 30 天的市场预期情况, 本文将它们分别用作预测黄金和白银期货波动率的预测变量。为提高阅读效率, 后文写作“隐含波动率”。

4 实证结果分析

4.1 描述性统计

表 3 汇总了所有变量的关键描述性统计量, 包括平均值 (Mean)、标准差 (SD)、偏度 (Skew)、峰度 (Kurt)、Jarque-Bera 统计量和 Augmented Dickey-Fuller (ADF) 单位根统计量. 结果显示, RV 呈现出右偏且有较高峰度的形态, 其他变量也大多以正偏度为主. Jarque-Bera 统计量表明, 所有变量均在 1% 的显著性水平下拒绝了正态性原假设. 此外, ADF 检验结果显示, 除 gepu 外, 其他预测变量的时间序列均为平稳⁴.

4.2 样本外预测结果

4.2.1 模型预测性能评价和比较

本文的研究对象包括黄金和白银期货. 相应地, 表 4 中 Panel A 和 Panel B 分别详细展示了两种贵金属的样本外预测结果, 且后续表格亦将遵循此格式.

观察表 4 可以发现, 所有机器学习模型的损失函数值几乎都小于基准模型 (HAR-RV), 同时, 这些模型均实现了正的 R_{OOS}^2 值和高于基准模型的 Doc 率. 此外, 样本外性能提升的 R_{OOS}^2 通过 CW 统计检验在 1% 水平上显著均显著, 这明确证实了机器学习模型在预测精度上优于基准模型.

具体来看, 在 Panel A 中, 黄金波动率的预测结果显示, GBDT 模型在所有机器学习模型中性能提升最为显著. 尽管 XGBoost 在黄金波动率预测方面表现最为突出, R_{OOS}^2 值达 29.210%, 超过了 LightGBM 和 CatBoost, 但在 MAE、QLIKE 和 Doc 率评估指标中, CatBoost 则展现出最优性能. 通过比较不同损失函数对误差值的统计方法, 可以看出, XGBoost 在较小误差值的预测点上表现更突出, 而 CatBoost 则在整体趋势判断上更具优势. 表 3 中 RV 的 ADF 检验结果显示, 时间序列数据表现出明显的平稳性特征, 表明历史数据中的极端波动情况较为罕见. 因此, 本文更倾向于将 RMSE 和 R_{OOS}^2 作为主要的性能评估标准. 在所有机器学习模型中, Doc 率这一方向评估指标的表现相近, 而集成树模型总体上优于其他模型, 这进一步证实了其预测效果的提升并非仅因特定点的误差减小, 而是由于预测误差的整体降低.

Panel B 中, 白银期货波动率的预测结果与黄金相似, 但白银期货波动率的损失函数值普遍高于黄金. 这主要是因为白银期货的波动率均值和方差都高于黄金, 即使白银期货通过了平稳性检验, 异常值的出现概率仍然较高. 此外, 黄金和白银期货的特征变量主要基于宏观经济变量, 在市场运作中, 投资者普遍对黄金表现出更高的关注度和信心, 这导致白银期货相对于黄金期货具有更高的信噪比, 从而增加了白银期货预测的难度. 在白银期货波动率预测中, LightGBM 展现了最优的性能, 而 XGBoost 的 R_{OOS}^2 性能提升与 LightGBM 非常接近, RMSE 和 QLIKE 损失函数评价也相差无几. 然而, 在 Doc 率指标上, 两种模型的预测结果存在显著差异. LightGBM 在模型优化中应用了互斥特征捆绑算法, 减少了高信噪比白银期货的数据维度, 通过考虑更少的特征来最小化冲突, 提高方向的准确性. XGBoost 则利用其泰勒展开二阶损失函数, 更精确地捕捉非线性特征关系, 降低特定点的预测误差. CatBoost

⁴ 本文的实证数据和代码发布在 ScienceData Bank, CSTR 编号为: 31253.11.sciencedb.20174.

表 3 描述性统计

	Factors	Mean	SD	Skew	Kurt	Jarque-Bera	ADF
Gold RV	RV	0.730	0.432	3.287	23.787	41451.379***	-6.618***
Silver RV	RV	1.182	0.782	2.864	19.303	26039.974***	-6.116***
Gold predictors	rsm	0.353	0.795	10.640	167.028	2385855.406***	-9.112***
	rsp	0.368	1.093	18.462	463.589	18619429.349***	-9.428***
	Jt	0.346	1.189	14.877	318.484	8757049.372***	-9.073***
	vol5	-12.915	21478.404	-0.417	13.661	9971.955***	-14.460***
	vol10	-14.346	11986.851	0.463	13.694	10048.042***	-12.859***
	londongold	0.288	13.184	-0.601	9.640	3970.776***	-12.425***
	baidu	0.107	0.248	1.316	4.195	728.418***	-10.500***
Silver predictors	rsm	1.022	3.192	20.551	612.321	32525351.634***	-9.927***
	rsp	0.986	2.173	10.124	159.669	2176311.592***	-8.463***
	Jt	0.861	2.688	10.234	141.266	1703729.271***	-9.606***
	vol5	-31730.480	149289.272	-0.238	10.108	4425.836***	-12.252***
	vol10	-5066.469	98455.504	-0.083	8.126	2294.115***	-11.645***
	londonsilver	0.002	0.340	-0.905	22.618	33847.681***	-11.728***
	baidu	0.161	0.575	0.992	3.273	349.949***	-12.005***
Common predictor	gepu	193.983	77.163	0.746	2.881	195.302***	-3.886
	cnepu	221.335	131.282	1.050	3.876	451.385***	-5.651***
	cntpu	221.335	131.282	1.050	3.876	451.385***	-5.651***
	cpi	-0.001	0.485	-0.405	4.329	211.268***	-8.339***
	gpr	94.434	20.026	0.731	3.101	187.248***	-5.588***
	ppi	0.128	0.877	0.534	4.580	317.357***	-4.261***
	cci	0.211	2.472	-0.449	3.263	76.453***	-10.369***
	gvz	-0.015	0.933	0.147	15.795	14285.538***	-13.378***
	effr	0.000	0.032	-9.904	282.643	6853919.244***	-10.149***
	usdx	0.008	0.331	0.176	8.525	2672.493***	-13.331***
	wti	-0.006	2.023	-4.251	394.475	13371217.633***	-12.416***
	brent	-0.003	1.268	-0.660	8.985	3275.873***	-10.869***
	vxslv	-0.011	1.941	2.612	49.909	194274.985***	-14.008***
	zz1000	1.634	138.865	-1.463	12.533	8671.807***	-11.492***

注: *** 为在 1% 的水平上显著, ** 为在 5% 的水平上显著, * 为在 10% 的水平上显著 (下同).

在黄金预测中表现更佳, 这得益于其在自动处理分类任务方面的专长, 但白银期货的高信噪比影响了其预测效果, 导致在白银期货预测上不如黄金. 本文从回归预测的视角发现 GBDT 模型在点预测结果中的差别, 表现为 LightGBM 对预测方向判断更准确, 而 XGBoost 有更小的预测误差. 这个发现与 Deng et al. (2022) 对股指分类预测的结果相似. 因此, 在 4.4 小节中, 本文将对模型在点预测中的效果分高低波动时期展开讨论.

表 4 样本外预测结果

Model	RMSE	MAE	QLIKE	R^2_{oos}	CWstat	CWp-value	Docrate	P'Tstat
Panel A: Gold								
Bench	0.358	0.227	0.787	-	-	-	73.684***	9.734
LASSO	0.331	0.216	0.779	14.844***	5.583	0.000	75.120***	10.282
Ridge	0.328	0.220	0.776	16.293***	4.326	0.000	73.445***	9.597
ElasticNet	0.341	0.224	0.784	9.615***	4.790	0.000	73.684***	9.704
SVR	0.331	0.211	0.775	14.360***	4.632	0.000	75.837***	10.588
RF	0.326	0.215	0.773	17.078***	5.439	0.000	76.555***	10.871
XGBoost	0.301	0.208	0.767	29.210***	4.465	0.000	75.598***	10.478
LightGBM	0.310	0.213	0.768	25.299***	4.327	0.000	76.555***	10.871
CatBoost	0.309	0.201	0.766	25.565***	6.241	0.000	77.033***	11.070
Panel B: Silver								
Bench	0.668	0.463	1.530	-	-	-	71.770***	8.944
LASSO	0.639	0.441	1.526	8.666***	4.249	0.000	74.402***	9.990
Ridge	0.613	0.426	1.519	15.903***	5.217	0.000	74.880***	10.188
ElasticNet	0.651	0.449	1.530	5.202***	3.994	0.000	74.641***	10.088
SVR	0.624	0.424	1.521	12.865***	5.064	0.000	76.794***	10.975
RF	0.651	0.457	1.524	5.155***	3.728	0.000	72.010***	9.010
XGBoost	0.586	0.412	1.515	23.201***	4.742	0.000	72.967***	9.408
LightGBM	0.585	0.406	1.515	23.350***	6.349	0.000	76.555***	10.879
CatBoost	0.619	0.430	1.518	14.225***	5.561	0.000	74.163***	9.903

4.2.2 CumSFE 曲线分析

图 4 和图 5 分别呈现了黄金期货和白银期货的预测模型 CumSFE 曲线,反映了在滚动预测中各机器学习模型相对于基准模型的误差平方差值的累积情况. 曲线的总体下降趋势表明,机器学习模型的预测性能优于基准模型. 进一步,通过观察 CumSFE 曲线的波动,可以判断模型的预测性能在各时间上是改善还是恶化,从而评估模型的预测性能在整个预测区间的稳定性和可靠性. 在黄金和白银期货波动率的预测中, XGBoost 和 LightGBM 模型在样本外预测区间内普遍显示出负的 CumSFE 值,并且这些值基本保持稳定下降,这表明这些模型在预测区间的任何时间点都能提供比基准模型更精确的预测,从而证实了它们出色的泛化能力. 相对地,其他一些机器学习模型在特定区间内 CumSFE 值的增加暗示了在这些区间内,这些模型的预测误差可能超过了基准模型,这使得它们的预测效果的持续性存疑.

4.3 稳健性检验

4.3.1 MCS 检验

表 5 展示了 MCS 检验的结果,揭示了在考察的所有损失函数和统计方法中,GBDT 模型在指定置信水平下都被纳入了最优模型集合内,这与第 4.2 节的发现相吻合. MCS 检验能够比较模型并识别最优模型,在黄金和白银期货市场中,GBDT 模型在不同损失函数条件下均展现出卓越性能,在黄金期货市场中, XGBoost 模型在 MSE 损失函数下的表现在所有统

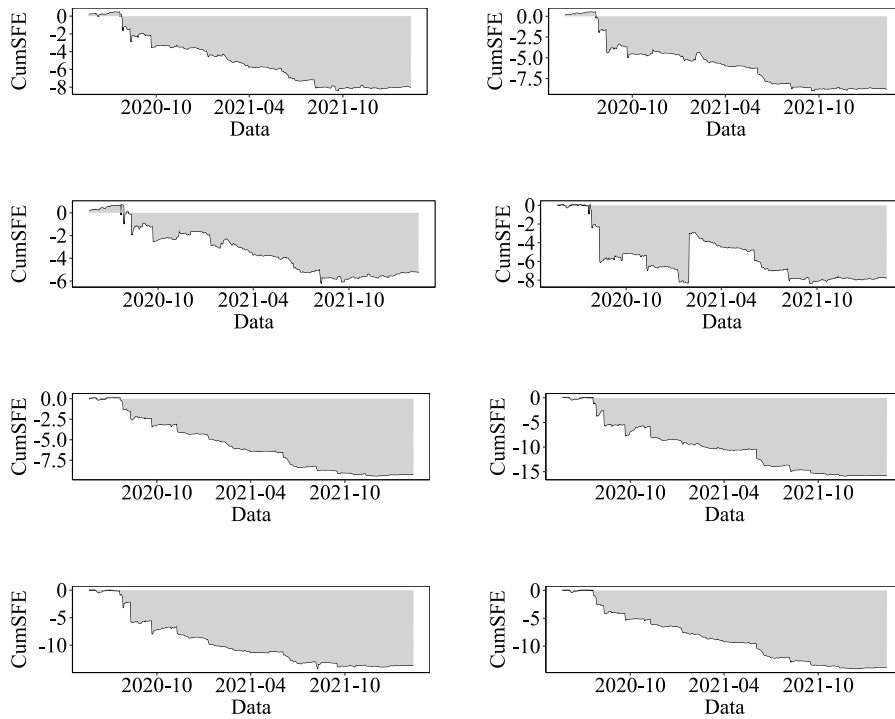


图 4 黄金期货市场机器学习模型的 CumSFE

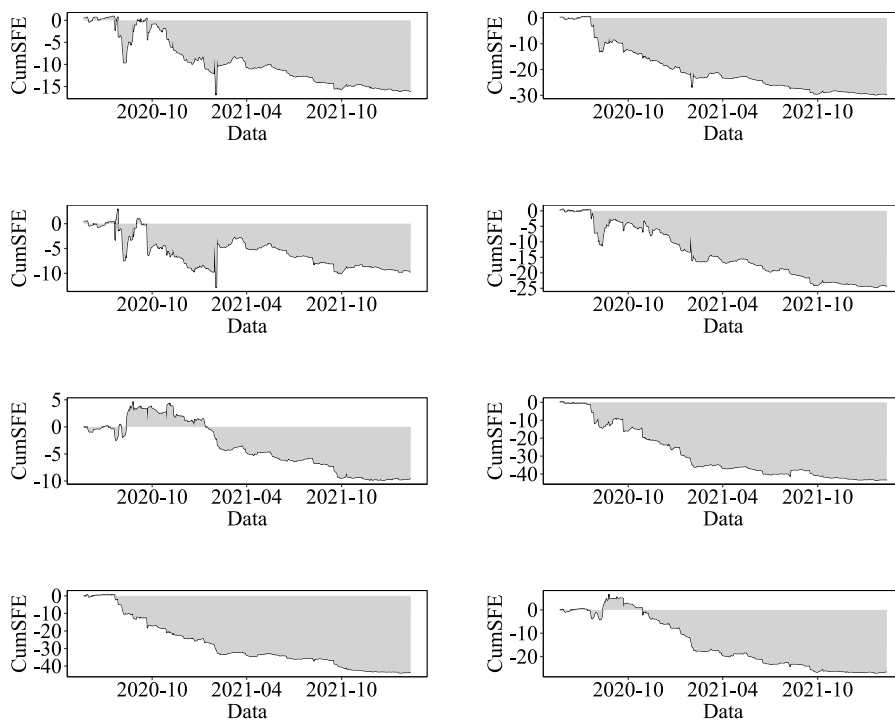


图 5 白银期货市场机器学习模型的 CumSFE

表 5 MCS 检验

Market	Model	MSE		MAE		R2LOG		MSE		MAE		R2LOG	
		T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}
Alpha = 0.1, B = 5000													
	Bench	-	-	-	-	-	-	-	-	-	-	-	-
	LASSO	0.0130	0.3700	0.0450	0.6700	0.0020	0.1510	0.0150	0.3620	0.0360	0.4480	-	-
	Ridge	0.1410	0.3730	0.0230	0.1340	0.0540	0.6340	0.1450	0.3970	-	-	0.0460	0.4510
	ElasticNet	-	-	-	-	-	-	-	-	-	-	-	-
Gold	SVR	0.7700	0.9200	0.8290	1.0000	0.4100	0.9960	0.7680	0.9150	0.7540	1.0000	0.3480	0.9220
	RF	0.0040	0.7310	0.0000	0.8770	0.0000	0.5930	0.0040	0.7210	0.0000	0.4910	0.0000	0.2820
	XGBoost	1.0000	1.0000	0.8890	1.0000	0.9260	1.0000	1.0000	1.0000	0.8130	1.0000	0.8870	1.0000
	LightGBM	1.0000	1.0000	0.2090	1.0000	0.4630	1.0000	1.0000	0.9700	1.0000	0.1580	0.3990	1.0000
	CatBoost	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9820	1.0000	1.0000	1.0000	1.0000	1.0000
Alpha = 0.2, B = 5000													
	Bench	-	-	-	-	-	-	-	-	-	-	-	-
	LASSO	0.1990	0.4170	0.2210	0.1360	0.3340	0.3250	-	-	-	-	0.3020	0.3020
	Ridge	0.5560	1.0000	0.5870	0.9860	0.8910	1.0000	0.3560	0.8400	0.5090	0.7670	0.8630	1.0000
	ElasticNet	0.1920	0.2380	-	-	0.3230	0.2800	-	-	-	-	0.3210	0.2680
Silver	SVR	0.4290	1.0000	0.8830	1.0000	0.9670	1.0000	0.2550	0.2340	0.8090	0.9270	0.9560	1.0000
	RF	0.1660	0.2010	-	-	0.0280	0.2160	-	-	-	-	0.0300	0.2140
	XGBoost	1.0000	1.0000	0.9970	1.0000	1.0000	1.0000	1.0000	1.0000	0.9890	1.0000	1.0000	1.0000
	LightGBM	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	CatBoost	0.8660	1.0000	0.6480	0.9130	0.9790	1.0000	0.6760	0.8430	0.5620	0.6180	0.9770	1.0000

注: 该表报告的是 MCS 检验的 p 值, 通过 5000 次 bootstrap 模拟得到. 带有下划线的加粗数字 1.000 表明该模型在对应的损失函数下预测精度最高.

计方法中最为优异, 与 XGBoost 的 RMSE 最小相对应; CatBoost 模型在其他损失函数下的预测效果更佳, 这一点在 MCS 检验中也得到了验证. 然而, MCS 检验并未揭示白银期货市场中 LightGBM 和 XGBoost 模型预测效果的独特特点, 但这两个模型的预测效果优于其他模型的结论与前文是一致的. 在白银期货预测中, Ridge 和 SVR 模型在某些损失函数和统计方法下的表现同样不容小觑, 在表 4 中的这两个模型的预测性能的效果与 CatBoost 接近, 因此在 MCS 的部分检验中不能被淘汰. 但这并不意味着这两个模型的预测效果接近 GBDT.

4.3.2 更换滚动窗口大小

选择不同的滚动窗口可能会影响预测结果 (Inoue et al., 2017). 因此, 本文设定滚动窗口 (训练集) 的大小分别为全样本的 70% 和 60%, 相应地, 将滚动预测窗口 (即测试集) 的比例调整为全样本的 30% 和 40%, 以检验模型的稳健性. 表 6 结果表明, 在更换滚动窗口大小后, 模型表现与前文结论基本一致, GBDT 模型的整体表现仍然较好. 而在 30% 的滚动预测窗口下, 白银期货市场中的 Ridge 和 SVR 模型展现出比 GBDT 模型更优的预测效果, 这与 MCS 检验中对白银市场的发现相一致.

表 6 不同预测窗口的表现

Model	RMSE	R^2_{Oos}	CW stat	Doc rate	RMSE	R^2_{Oos}	CW stat	Doc rate
	Gold				Silver			
Panel A: 30% of the entire sample set								
Bench	0.473	-	-	72.133***	0.779	-	-	69.745***
LASSO	0.436	15.077***	4.730	75.000***	0.749	7.507***	2.869	72.293***
Ridge	0.439	14.003***	4.689	74.363***	0.718	15.031***	4.011	74.363***
ElasticNet	0.431	17.098***	5.275	73.885***	0.749	7.563***	3.696	73.089***
SVR	0.429	17.524***	4.930	75.796***	0.708	17.286***	4.606	73.408***
RF	0.460	5.272***	3.265	74.363***	0.804	-6.442*	1.483	72.293***
XGBoost	0.424	19.716***	5.361	75.000***	0.743	9.061***	5.382	72.452***
LightGBM	0.420	21.111***	5.661	74.045***	0.739	10.041***	5.560	74.682***
CatBoost	0.432	16.409***	5.885	75.796***	0.767	2.914***	3.614	74.363***
Panel B: 40% of the entire sample set								
Bench	0.432	-	-	71.087***	0.687	-	-	68.698***
LASSO	0.412	9.018***	4.311	73.357***	0.666	5.993***	4.423	71.685***
Ridge	0.411	9.564***	3.859	73.955***	0.653	9.691***	3.963	71.326***
ElasticNet	0.403	12.875***	4.881	72.999***	0.670	4.833***	3.081	71.924***
SVR	0.398	14.904***	5.685	74.194***	0.680	2.073***	3.158	73.835***
RF	0.427	2.394***	2.719	73.565***	0.721	-10.166	0.972	70.574***
XGBoost	0.407	11.072***	5.475	73.716***	0.698	-3.285***	3.059	71.804***
LightGBM	0.389	18.962***	6.955	75.030***	0.641	12.736***	5.855	72.760***
CatBoost	0.402	13.264***	5.513	76.344***	0.680	2.004***	3.322	73.238***

4.3.3 更换滚动预测步长

对于投资者而言, 虽然向前一步的滚动预测算法有助于捕捉极短期的市场动向, 但中期预测能力同样至关重要. 为此, 本文设定了 3 天、6 天和 12 天的预测步长, 评估了模型在这些步长下的样本外性能, 即模型在每次滚动预测中预测未来 3、6 或 12 个交易日的数据. 表 7 结果显示, 在增加滚动预测步长后, 所有机器学习模型的预测性能均保持稳定. GBDT 模型在未自动优化参数的情况下, 其 R_{OOS}^2 和 Doc 率整体表现依然卓越, 进一步验证了其在预测贵金属期货波动率方面的稳定性和可靠性.

表 7 不同预测步长的表现

Model	RMSE	R_{OOS}^2	CW stat	Doc rate	RMSE	R_{OOS}^2	CW stat	Doc rate
	Gold				Silver			
Horizon = 3								
Bench	0.359	-	-	73.558***	0.669	-	-	71.635***
LASSO	0.327	16.923***	4.317	74.760***	0.644	7.349***	4.323	72.115***
Ridge	0.334	13.076***	4.946	74.038***	0.612	16.358***	5.217	75.000***
ElasticNet	0.323	19.023***	4.179	74.760***	0.627	12.080***	4.265	73.558***
SVR	0.304	28.387***	4.932	76.683***	0.586	23.140***	5.601	73.558***
RF	0.330	16.078***	5.708	75.545***	0.656	3.699***	3.427	72.771***
XGBoost	0.306	27.225***	4.306	77.404***	0.616	15.055***	5.160	72.837***
LightGBM	0.309	25.620***	4.920	76.442***	0.597	20.309***	5.754	72.596***
CatBoost	0.307	26.904***	5.668	77.885***	0.618	14.673***	5.230	73.317***
Horizon = 6								
Bench	0.360	-	-	73.850***	0.671	-	-	71.913***
LASSO	0.329	16.349***	4.867	75.787***	0.657	4.002***	4.006	74.092***
Ridge	0.327	17.304***	4.345	76.513***	0.615	16.046***	4.487	74.818***
ElasticNet	0.328	16.984***	5.046	72.397***	0.646	7.448***	3.899	72.639***
SVR	0.321	20.110***	3.984	77.240***	0.618	15.252***	5.100	74.818***
RF	0.329	16.324***	5.707	76.029***	0.654	5.038***	3.846	72.397***
XGBoost	0.322	19.699***	4.271	75.303***	0.627	12.821***	4.983	73.366***
LightGBM	0.324	18.783***	4.511	76.271***	0.598	20.628***	5.196	75.061***
CatBoost	0.312	24.571***	5.619	75.303***	0.630	11.956***	5.090	72.639***
Horizon = 12								
Bench	0.362	-	-	74.201***	0.675	-	-	71.744***
LASSO	0.325	19.295***	5.763	75.184***	0.645	8.569***	3.989	74.939***
Ridge	0.326	18.936***	4.913	75.921***	0.623	14.824***	4.759	73.464***
ElasticNet	0.325	19.447***	4.615	78.133***	0.661	3.926***	3.751	74.447***
SVR	0.335	14.386***	5.177	74.693***	0.670	1.246***	4.245	73.219***
RF	0.330	16.909***	5.289	74.939***	0.667	2.228***	3.081	72.482***
XGBoost	0.312	25.726***	4.055	75.430***	0.595	22.192***	5.466	73.464***
LightGBM	0.315	24.358***	4.307	75.676***	0.593	22.791***	5.767	74.447***
CatBoost	0.312	25.792***	5.614	77.887***	0.638	10.695***	4.706	73.956***

表 8 高波动期和低波动期的样本外预测性能

Market	Model	RMSE	MAE	QLIKE	R_{oos}^2	CW stat	CW p -value
Low volatility period							
Gold	Bench	0.224	0.189	0.476	-	-	-
	LASSO	0.192	0.154	0.466	26.276***	7.005	0.000
	Ridge	0.202	0.163	0.467	18.527***	5.524	0.000
	ElasticNet	0.203	0.159	0.469	17.553***	5.799	0.000
	SVR	0.182	0.136	0.461	33.682***	8.001	0.000
	RF	0.224	0.197	0.478	-0.237**	1.871	0.031
	XGBoost	0.221	0.174	0.471	2.775***	3.453	0.000
	LightGBM	0.222	0.178	0.472	2.021***	3.253	0.001
	CatBoost	0.208	0.177	0.471	13.390***	4.553	0.000
High volatility period							
Gold	Bench	0.455	0.266	1.099	-	-	-
	LASSO	0.427	0.279	1.092	12.093***	4.091	0.000
	Ridge	0.418	0.278	1.085	15.784***	3.444	0.000
	ElasticNet	0.437	0.290	1.100	7.697***	3.548	0.000
	SVR	0.432	0.286	1.089	9.686***	3.443	0.000
	RF	0.404	0.233	1.069	21.278***	5.297	0.000
	XGBoost	0.365	0.243	1.064	35.624***	4.067	0.000
	LightGBM	0.378	0.249	1.065	30.959***	3.924	0.000
	CatBoost	0.385	0.227	1.063	28.513***	5.570	0.000
Low volatility period							
Silver	Bench	0.437	0.333	1.185	-	-	-
	LASSO	0.393	0.303	1.181	18.866***	3.781	0.000
	Ridge	0.392	0.301	1.180	19.492***	4.408	0.000
	ElasticNet	0.405	0.306	1.181	14.051***	3.627	0.000
	SVR	0.332	0.250	1.173	42.111***	6.283	0.000
	RF	0.411	0.335	1.185	11.403***	2.586	0.005
	XGBoost	0.429	0.295	1.180	3.592***	2.861	0.002
	LightGBM	0.406	0.299	1.180	13.558***	3.444	0.000
	CatBoost	0.396	0.307	1.181	17.577***	4.000	0.000
High volatility period							
Silver	Bench	0.839	0.592	1.875	-	-	-
	LASSO	0.815	0.581	1.872	5.803***	3.527	0.000
	Ridge	0.774	0.554	1.859	14.839***	4.312	0.000
	ElasticNet	0.828	0.594	1.878	2.703***	3.363	0.000
	SVR	0.819	0.599	1.870	4.859***	3.435	0.000
	RF	0.825	0.581	1.863	3.357***	2.842	0.002
	XGBoost	0.710	0.531	1.851	28.415***	4.130	0.000
	LightGBM	0.722	0.514	1.851	25.921***	5.573	0.000
	CatBoost	0.782	0.555	1.856	13.214***	4.470	0.000

4.4 高、低波动时期分析

波动率预测模型在不同波动机制下的预测能力往往表现出差异 (Liang et al., 2019). 本研究将数据集划分为高波动与低波动两个子样本时期, 分别评估模型在这两个子样本的样本外预测表现. 表 8 结果表明, 在高波动和低波动时期, 机器学习模型几乎均展现出正向 R_{OOS}^2 的预测能力. 具体而言, 在低波动时期, 变量选择模型和 SVR 模型在预测精度上尤为突出. 这种优势主要得益于变量选择模型中 L_1 和 L_2 正则化项的引入, 有效降低了处理大量特征或特征共线性时的过拟合风险. SVR 模型则利用核技巧有效地捕捉了数据的非线性特征. 在低波动率区间, 市场的稳定性使得模型能更敏锐地识别线性和简单非线性关系, 从而增强了变量选择模型和 SVR 的适应性和预测准确性.

黄金和白银作为贵金属的代表, 尤其是黄金, 因其货币属性而成为衡量市场稳定性的关键指标. 因此, 对这些贵金属期货在高波动区间的预测价值应给予特别关注. 在黄金期货的波动性预测中, XGBoost 模型在所有损失函数上均表现最佳, 显示出较低的预测误差. 在白银期货方面, XGBoost 和 LightGBM 模型分别在 RMSE 和 MAE 上表现最优, 两者的 QLIKE 持平. 这可能是因为在高波动时期更有效地捕捉了数据的非线性特征, 尤其是在处理具有较大误差的极端数据点时. 与此同时, LightGBM 利用其 EFP 算法减少了特征间的相互作用, 这在波动性较高的市场条件下进一步突出了其优势. XGBoost 和 LightGBM 通过增强对市场高波动率环境的识别, 助力分析师和交易员更好地理解 and 预测市场行为, 从而在高波动性的市场环境中做出更加明智的决策.

5 模型可解释分析

5.1 经济价值分析

表 9 展示了不同风险规避系数下 (Guo et al., 2023), 各模型在波动率预测中所产生的平均预期已实现效用的结果. 对黄金和白银期货而言, GBDT 的已实现效用大幅超越了其他模型, 意味着不同风险偏好的市场参与者在 GBDT 模型的波动率预测中能够获得更大的已实现效用, 从而可能获得更高的投资收益. 以 $\gamma = 2$ 的黄金期货波动率为例, Bench 模型与

表 9 经济价值分析

Model	$\gamma = 2$	$\gamma = 3$	$\gamma = 6$	$\gamma = 2$	$\gamma = 3$	$\gamma = 6$
	Gold			Silver		
Bench	3.853	2.569	1.284	3.876	2.584	1.292
LASSO	3.872	2.581	1.291	3.885	2.590	1.295
Ridge	3.879	2.586	1.293	3.900	2.600	1.300
ElasticNet	3.860	2.574	1.287	3.875	2.583	1.292
SVR	3.882	2.588	1.294	3.894	2.596	1.298
RF	3.888	2.592	1.296	3.891	2.594	1.297
XGBoost	3.901	2.601	1.300	3.910	2.607	1.303
LightGBM	3.899	2.599	1.300	3.910	2.606	1.303
CatBoost	3.902	2.601	1.301	3.903	2.602	1.301

CatBoost 模型之间的已实现效用相差 0.049%, 说明在已知的风险厌恶系数和夏普比率条件下, 投资者需要多付出 0.049% 的成本才能将 Bench 模型的效用价值变成 CatBoost 的价值, 而这部分可能就是资产管理人的额外收入来源. 这一发现不仅证明了 GBDT 模型在预测上海贵金属期货波动性的准确性, 更凸显了其在实践中为投资者带来巨大经济收益的潜力. 因此, GBDT 模型在贵金属期货市场的波动率预测中表现出显著的优越性和实用性, 优化了投资组合配置的效率, 能为投资者提供了更可靠且高回报的策略选择.

5.2 VIP 模型可解释性工具分析

变量重要性衡量了在模型预测过程中各个变量对最终预测结果的影响力度, 其中每个变量在决策树节点分裂时所带来的信息增益或熵减被用来评估其对预测结果的影响. 这一指标主要用来研究模型预测准确性的提升及其解释能力. 本文选取了 GBDT 模型中的 XGBoost 作为代表性模型, 分析变量重要性. 图 6 和图 7 分别是黄金期货和白银期货 XGBoost 模型的变量重要性分析结果.

本研究计算了 XGBoost 模型的预测变量重要性, 并对这些重要性得分进行了平均, 以得出各变量的特征重要性. 图中展示了对预测效果影响最大的前十个预测变量. 在黄金期货的波动率预测中, `gvz_lag1`、`londongold_lag1` 和 `baidu_lag1`⁵ 分别是重要影响前三的变量, 并且影响程度远大于其它预测变量. 特别是 CBOE 黄金 ETF 波动率指数 (`gvz`), 其对中国黄金期货市场的波动特征有着密切的关联, 能够更精准地捕捉市场变化. 在白银期货波动率的预测中, `lag_1`、`londonsilver_lag1` 和 `vxslv_lag1` 是效果最佳的三个预测变量, 其中后两个

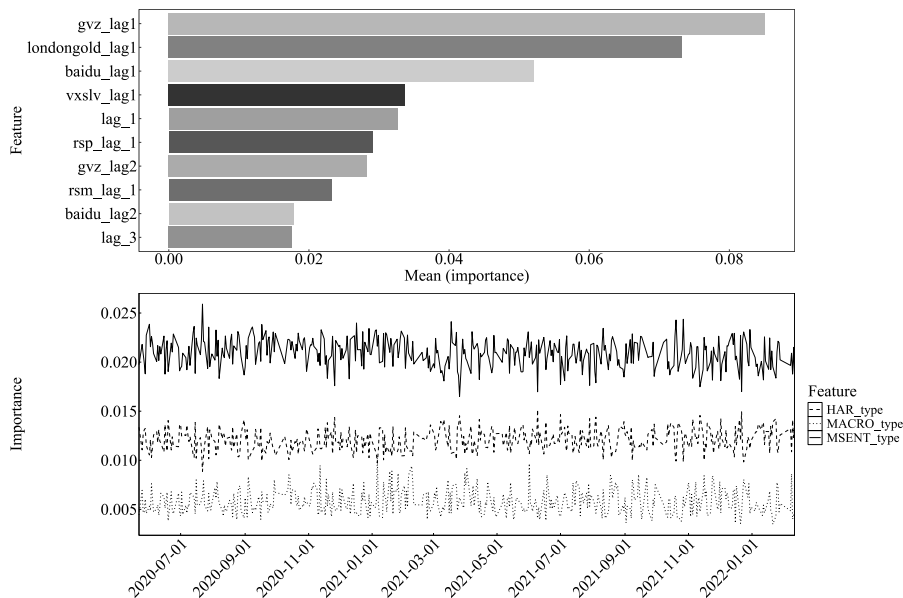


图 6 黄金期货 XGBoost 模型的 VIP 解释

⁵所有预测变量名称中, 带有“lag + 数字”的均表示预测变量滞后阶数, 为提高文章阅读效率, lag1 滞后一阶的变量, 后文均直接用变量名或者英文缩写代替, 不提及滞后阶数.

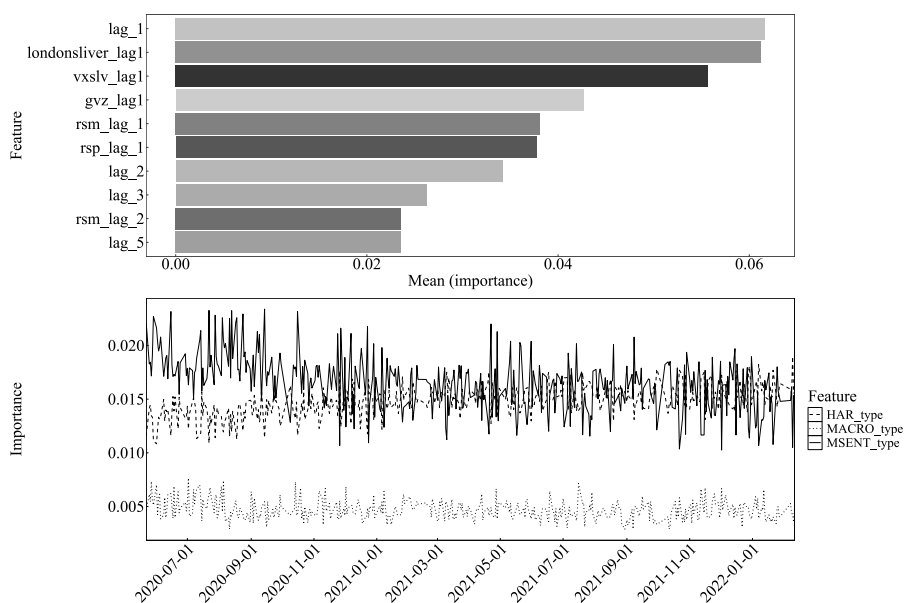


图7 白银期货 XGBoost 模型的 VIP 解释

与黄金期货的相关指标相呼应,表明 CBOE 隐含波动率指数及伦敦金或银的价格指数对中国贵金属期货波动率具有显著影响。值得注意的是,虽然百度搜索指数在黄金期货预测中具有重要影响,但在白银期货预测中并未显示出相同的重要性。这可能与黄金的货币属性和避险功能更为突出有关,导致市场参与者对黄金相关信息的关注度更高,进而在价格波动上得到反映。

本文将加入的变量分为三类,即 HAR 类、MSENT 类和 MACRO 类,然后在图 6 和图 1 呈现了各类变量影响程度随时间的变化趋势。其中, HAR 类主要就是 RV 及其分解信息,包含 RV 的跳跃成分 (Jt)、已实现正半差 (rsp) 和已实现负半差 (rsm),以及滞后 1 到 22 阶的 RV。MSENT 类变量反映了市场情绪,包括隐含波动率 (gvz、vxslv)、百度指数 (baidu)、交易量 (vol5, vol10) 和伦敦金/银 (londongold、londonsilver) 四种变量。MACRO 类包括剩余的所有经济变量。本文发现, MACRO 类变量对模型的影响程度大多处于较低水平。其次,在整个区间内,黄金期货波动率预测中 MSENT 类变量对模型预测的影响一直高于 HAR 类变量,并且两者总体上保持稳定,呈现在一定范围内波动。而白银期货波动率预测中 MSENT 类变量对模型预测的影响在 2020 年下半年高于 HAR 类变量,随后两者的影响程度相当。因此,对于不同的市场,变量对波动率预测的影响变化情况略有不同,因此对于预测因子的选择都应该考虑这一影响。总的来说,以上证据证明,反映市场情绪的 MSENT 类变量在预测贵金属期货波动性方面起着极其重要的作用。

5.3 SHAP 值分析

5.3.1 SHAP 主效应与交互效应分析

由于 XGBoost 模型在多数情况下都能显著提升预测性能并在预测准确性上表现卓越, 本节的 SHAP 值分析将以此模型进行全局解释性分析. SHAP 值的计算基于滚动预测的预测值, 这种滚动机制涉及数据的不断重构, 从而能够反映市场动态变化. 模型输出的有序的 SHAP 值用于量化每个预测变量对预测结果的贡献程度, 使市场参与者更好地理解 and 利用市场信息. 此外, 本研究还通过对 SHAP 值的分解, 得到 SHAP 主效应值和 SHAP 交互效应值, 即表示预测变量之间交互作用的 SHAP 交互值. 这一结果是通过矩阵输出的, 主效应在对角线上给出, 交互效应分布在矩阵中其它位置. 为避免正负 SHAP 交互值相互抵消, 本文将每个预测值对应的所有 SHAP 交互值取绝对值后, 按主效应和交互效应分类加总.

由于金融市场的信噪比非常低, 因此本文选用主效应和交互效应分别表示 XGBoost 预测提取的信号和噪音, 来研究模型在预测过程中对噪音和信号的识别能力. 图 8 和图 9 的上图分别是黄金期货和白银期货 XGBoost 模型的 SHAP 主效应、交互效应和总效应的时间趋势, 下图是对应市场的 RV 时间序列图. 从中可以得到两个关键信息: 第一, 预测变量对预测值的贡献程度基本保持稳定, 但随着时间的推移, 这种贡献仍然会逐渐减弱. 这种现象表明, 预测变量向预测值传递的信号在减弱. 贡献值的持续降低意味着需要考虑替换预测变量, 否则可能导致预测效果变差; 第二, 无论是黄金期货还是白银期货, 交互效应 (Interaction Effect) 在多数时间高于主效应 (Main Effect). 这意味着在预测过程中, 预测变量之间的相互影响比它们各自对预测值的影响更为显著. 换句话说, 预测变量之间存在的复杂非线性关系, 较预测值和预测变量之间的非线性关系更强. 这种数据的内在结构的复杂性是传统线性模型无法捕捉的, 而基于 GBDT 的 XGBoost 模型能够有效地应用于此类复杂情景的预测.

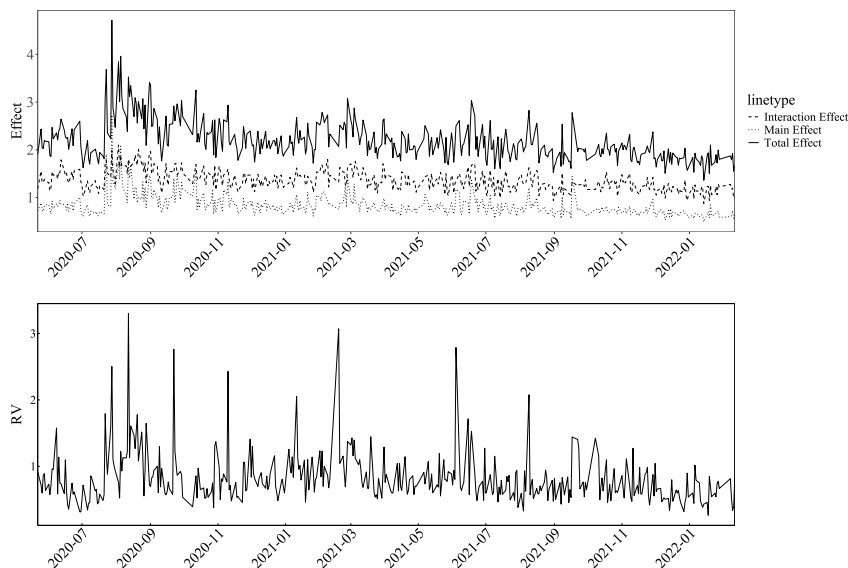


图 8 黄金期货 XGBoost 模型的 SHAP 主效应与交互效应

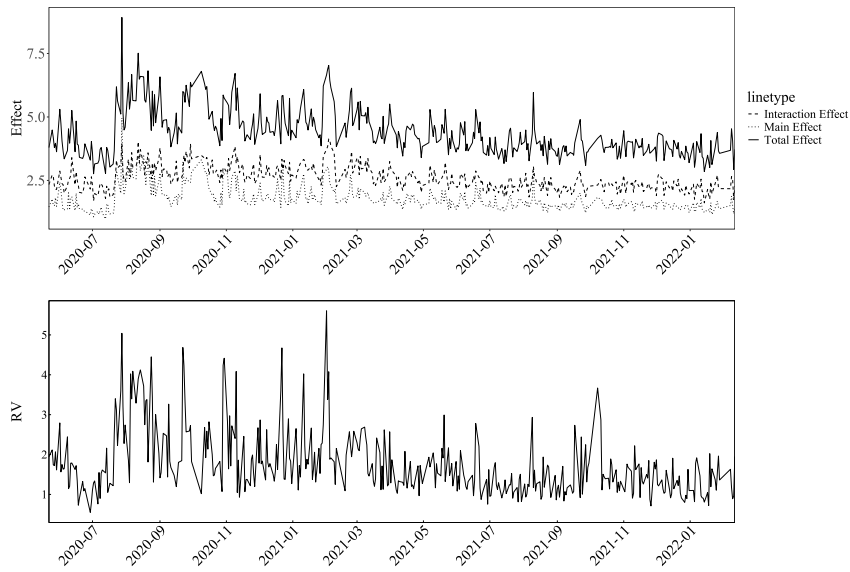


图 9 白银期货 XGBoost 模型的 SHAP 主效应与交互效应

5.3.2 黄金市场预测的 SHAP 值分析

主效应和交互效应之间的相互联系揭示了 XGBoost 能够从预测变量中提取信号并作用于预测值。因此, 本节将分析主效应与特征值之间的关系。图 10 和图 11 分别是黄金期货和白银期货市场的 SHAP 值的总结图 (summary plot) 和依赖图 (dependence plot)。其中

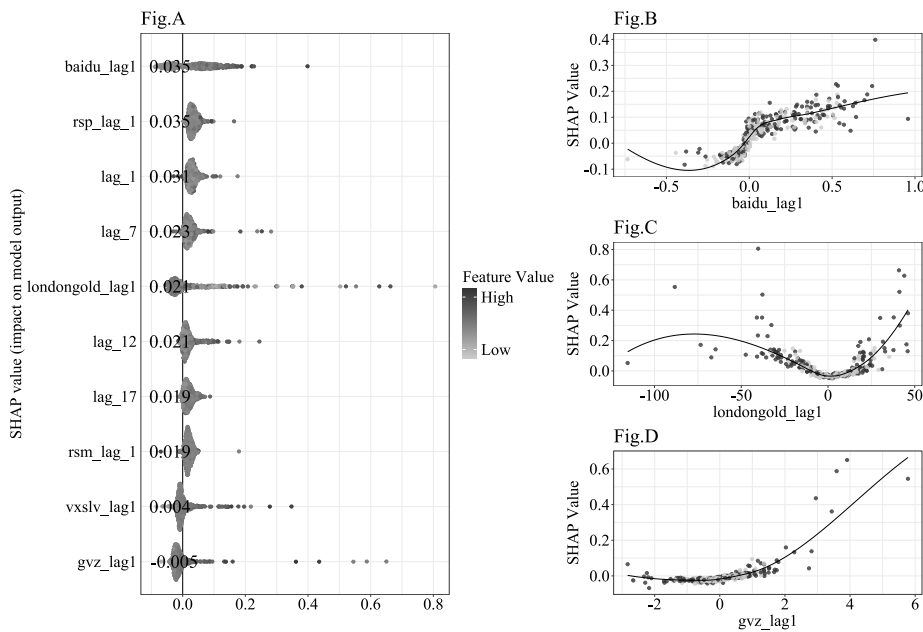


图 10 黄金期货 XGBoost 模型的 SHAP 值分析

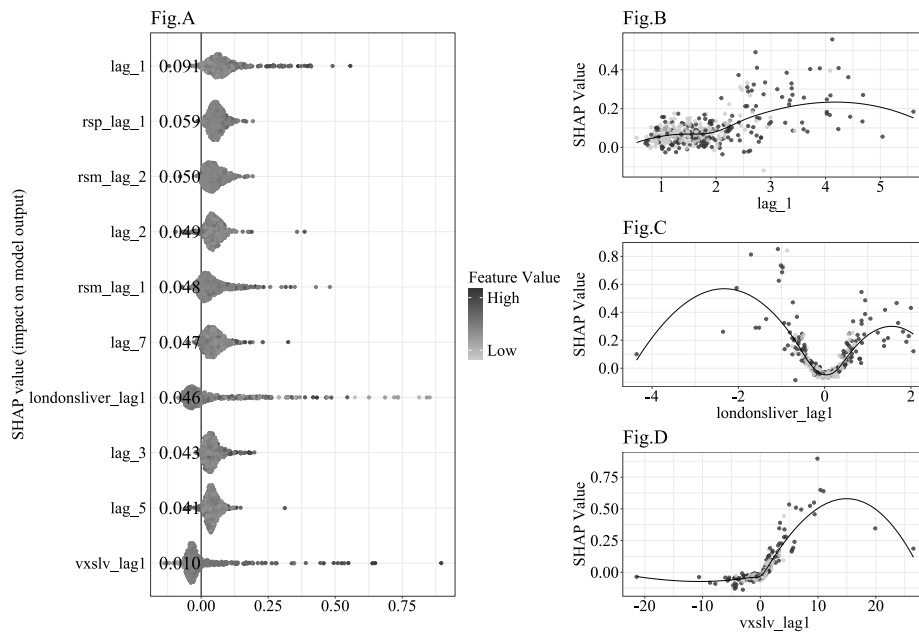


图 11 白银期货 XGBoost 模型的 SHAP 值分析

Fig.A 表示排名前十的特征变量的特征值与 SHAP 值之间的分布, 图中深灰色代表对应纵坐标的特征变量在整个样本外区间的值越高, 浅灰色代表越低. 横坐标表示特征变量的 SHAP 值, SHAP 值越大, 说明预测变量对预测值产生的正向影响越大; 而 SHAP 值越小, 表明预测变量对得出预测值产生的负向影响越大. Fig.B、Fig.C 和 Fig.D 反映 RV 真实值、特征值和 SHAP 值三者间的关系. 在这些图中, 横坐标表示特征值, 纵坐标表示对应的 SHAP 值, 黄点代表高波动的 RV, 紫点代表低波动的 RV (RV 真实值根据 4.4 节中高低波动率的划分方式).

由图 10 黄金市场 SHAP 值分析的 Fig.A 中, 本文发现 baidu_lag1 对总体预测效果的贡献最大, 同时 londongold_lag1 和 gvz_lag1 两个市场情绪指标贡献也突出. 因此, 本文挑选出这三个预测变量描绘特征值和 SHAP 值依赖关系, 描绘出 Fig.B, Fig.C 和 Fig.D. 在 Fig.B 和 Fig.D 中, 随着黄金关注度的提升或未来预期波动增加, 即黄金百度指数数值增大或隐含波动率指数提高, 市场参与者会更积极地调整黄金期货的投资头寸, 来应对未来黄金价格波动. 这种频繁交易行为导致黄金期货波动率增加, 如图表右尾密集的高波动. Fig. D 中, 当隐含波动率值大于 2 时, 所有点是黄点, 即真实值都属于高波动率类别. Fig.B 和 Fig.C 的右尾都表现出相似的特征: 随着特征值的增加, 它们在预测过程中的贡献也随之增强, 从而导致预测值上升. 当真实值处于高波动区间时, 这些特征变量通过提高预测值, 有助于使预测结果更贴近实际波动情况.

黄金百度指数通常反映市场参与者对“黄金”关键词的关注程度, 黄金隐含波动率则反映了未来一个月黄金市场预期波动. 随着黄金关注度的提升或者未来预期波动增加, 市场参与者会更积极地调整黄金期货的投资头寸, 来应对未来黄金价格波动. 这种频繁交易行为导致黄金期货波动率增加, 如 Fig.B 和 Fig.D 右尾密集的黄点所示. 在 Fig.C 中, 伦敦金和

SHAP 值之间存在“U”型关系,“U”型曲线的两侧高波动点更加密集.这暗示了前一天伦敦金的价格变化对黄金期货的波动性的增强作用,以及伦敦金价格的显著增加了对预测结果的贡献度的提升.综上所述,黄金百度指数、黄金隐含波动率和伦敦金共三个特征的 SHAP 依赖图揭示: XGBoost 不仅在预测性能表现上出色,而且通过细致地捕捉 SHAP 值与关键特征值之间的相互作用,进一步证明了其预测的精确性.这一发现为市场投资者提供了一个强有力的工具,使他们能够精准地识别并利用市场中的套利机会,进而实现利润最大化.这种基于数据驱动的分析方法,为投资者在复杂且动态的黄金市场中把握机遇提供了坚实的支持.

5.3.3 白银市场预测的 SHAP 值分析

图 11 的 Fig.A 揭示了 RV 的一阶滞后 (lag_1) 是预测中贡献最高的特征变量,这与 5.2 节 VIP 分析的结果相一致.在白银市场,白银百度指数的对预测的贡献较小,这可能反映了市场参与者对白银的关注度不如黄金.在预测贡献度排名前十的变量中, HAR 类变量占大多数,比如 rsp (正半差) 和 rsm (负半差) 的滞后变量,这表明 HAR 类变量在预测白银波动性方面施加重要的影响.在 Fig.B 中, lag_1 (前一天白银波动率) 的上升与其 SHAP 值的增大呈现正相关性.当 lag_1 大于 2 时,代表高波动区域的黄点更密集,表明前一日的高波动率对下一交易日有显著影响.这反映出白银期货波动具有明显的持续性,即如果前一日波动幅度较大,那么次日也倾向于出现较大波动.在图 11 的伦敦银 (Fig.C) 和白银隐含波动率 (Fig.D) 的依赖图中, SHAP 值与特征值的关系与黄金期货相似,表明市场参与者能迅速发现同类金融资产的套利机会,并对价格波动率做出预测.

在市场投资领域,当 CBOE 黄金和白银隐含波动率 (gvz 和 $vxslv$) 出现波动,或者伦敦金和伦敦银价格出现显著变动时,如果 XGBoost 模型预测的波动率处于历史波动率的高波动期,那么可以通过分析预测变量的 SHAP 值来确定波动的成因.通过对比历史 SHAP 值,可以判断剧烈波动是主要由关键预测变量引起,还是由其他特殊因素导致.如果预测结果显示高波动率的成因与历史类似,这通常表明市场波动属于正常现象.鉴于上海黄金和白银期权分别在 2019 年底和 2022 年底推出,在这种情况下,投资者可以采取做多波动率的期权策略,如购买跨式 (Straddle) 或宽跨式 (Strangle) 期权组合.通过细致的 SHAP 值分析,可以更精确地识别波动的驱动因素,从而帮助市场参与者做出更可靠的投资决策.

6 主要结论及启示

本文以上海贵金属期货为研究标的,重点研究其波动率的预测,首先应用多种机器学习模型,并在多种损失函数和检验方法下全面比较了这些模型的预测性能.实证结果显示,在样本外预测中,GBDT 模型展现出显著的优势.通过对预测点的波动期进行高低分类分析,发现 XGBoost 模型在贵金属期货的高波动率预测中误差较小.在此基础上,本文创新的融合了机器学习模型的可解释性工具,全面探究和深入解释了 XGBoost 模型在贵金属期货波动率预测上卓越表现的驱动因素.具体来说,本研究的发现主要有以下几个方面:

1) 在 XGBoost、LightGBM 和 CatBoost 三种 GBDT 算法的性能比较中, XGBoost 在部分预测点上展现出更小的预测误差, LightGBM 在预测值方向的判断上更为准确,而 CatBoost 则在特征分类方面表现突出,并对黄金市场的变化更为敏感.这些发现在经过

MCS 检验、调整滚动窗口大小和预测步长的策略以及经济价值分析后, 显示出良好的稳健性。

2) 在贵金属波动率的高低波动区间的预测中, GBDT 模型在高波动率的预测中表现更优, 这一结果与 GBDT 在原油波动率预测中的发现相符 (Fu et al., 2024). 对于黄金这类具有特殊货币属性的资产, 准确的高波动率预测能力, 反映了模型在识别市场异常波动方面的强大性能。

3) 在探究 XGBoost 模型在贵金属期货波动率预测中性能提升的原因时, VIP 分析揭示了一个关键发现: 市场情绪 (MSENT) 类变量在预测中扮演了至关重要的角色, 其次是 HAR 类变量的贡献, 而宏观指标 (MACRO) 类变量影响相对较小. 这一发现强调了市场情绪对贵金属期货价格波动的显著影响。

4) 深入分析 SHAP 主效应与特征值的关系, 发现百度指数和隐含波动率与黄金、白银期货波动预测呈正相关. 同时, 伦敦金和伦敦银价格与 SHAP 值之间存在 U 型关系. 这些因素能提高预测的精确度, 并对市场参与者制定精准投资决策具有重要意义。

本研究为贵金属期货波动率预测模型的选择提供了重要的实证支持, 并揭示了预测因子的重要性及其原因. 其应用价值和意义主要体现在以下几个方面: 首先, 通过融合机器学习预测性能和可解释性工具, 本文阐明了 XGBoost 模型预测性能提升的原因. 这为未来在其他市场中应用 GBDT 模型提供了一种有效的融合方法, 既能提高模型性能, 又能增强其解释性. 其次, 本研究将 VIP 和 SHAP 的解释方法与向前一步滚动预测算法相融合, 不仅揭示了变量与预测结果关系持续性的变动特征, 还提高了预测模型的精确性和可靠性. 最后, 通过对特征变量贡献度的分解分析, 本文强调了市场情绪变量在贵金属期货波动率预测中的重要性. 与此同时, 本文还探讨了特征值与 SHAP 值之间的关系对期货波动率投资的意义. 综上, 本研究为贵金属期货波动率预测的模型选择和驱动因子解释问题提供了崭新的融合解决方案并提供了全面的实证证据, 为相关市场参与者制定投资组合策略以及市场监管者制定风险管理策略提供了前沿且稳健的风险建模框架, 也为其他金融资产标的或金融市场的波动率预测领域的研究提供了有益借鉴和参考。

参 考 文 献

- 郭杨莉, 马锋, (2024). 基于马尔科夫和混频数据模型的黄金期货市场波动率预测研究 [J]. 中国管理科学, 32(1): 13-22.
- Guo Y L, Ma F, (2024). Forecasting the Chinese Gold Futures Market Volatility Using Markov-Switching Regime and Mixed Data Sampling Model[J]. Chinese Journal of Management Science, 32(1): 13-22.
- 洪永淼, 汪寿阳, (2021). 大数据、机器学习与统计学: 挑战与机遇 [J]. 计量经济学报, 1(1): 17-35.
- Hong Y M, Wang S Y, (2021). Big Data, Machine Learning and Statistics: Challenges and Opportunities[J]. China Journal of Econometrics, 1(1): 17-35.
- 胡越, 王桑原, 覃浩恒, 徐亮, 张一苇, (2023). 基于双重 XGBoost 模型的农产品期货波动率预测——以玉米期货为例 [J]. 系统管理学报, 32(2): 332-342.
- Hu Y, Wang S Y, Qin H H, Xu L, Zhang Y W, (2023). Volatility Prediction of Agricultural Products

- Based on Dual XGBoost Model: A Case Study of Corn Futures[J]. *Journal of Systems & Management*, 32(2): 332–342.
- 梁超, 魏宇, 马锋, 李薇, (2022). 投资者关注对中国黄金价格波动率的影响研究 [J]. *系统工程理论与实践*, 42(2): 320–332.
- Liang C, Wei Y, Ma F, Li W, (2022). A Study on the Impact of Investor Attention on Chinese Gold Volatility[J]. *Systems Engineering — Theory & Practice*, 42(2): 320–332.
- 梁超, 魏宇, 马锋, 李霞飞, (2022). 我国黄金期货价格波动率预测研究: 来自模型缩减方法的新证据 [J]. *中国管理科学*, 30(4): 30–41.
- Liang C, Wei Y, Ma F, Li X F, (2022). Forecasting Volatility of China Gold Futures Price: New Evidence from Model Shrinkage Methods[J]. *Chinese Journal of Management Science*, 30(4): 30–41.
- 马锋, 魏宇, 黄登仕, 张鹏云, (2015). 基于跳跃和符号跳跃变差的 HAR-RV 预测模型及其 MCS 检验 [J]. *系统管理学报*, 24(5): 700–710.
- Ma F, Wei Y, Huang D S, Zhang P Y, (2015). The HAR-RV Forecasting Model with the Jump and Signed Jump Variations and Its MCS Test[J]. *Journal of Systems & Management*, 24(5): 700–710.
- 庞淑娟, 刘向丽, 汪寿阳, (2011). 中国期货市场高频波动率的长记忆性 [J]. *系统工程理论与实践*, 31(6): 1039–1044.
- Pang S J, Liu X L, Wang S Y, (2011). Long Memory of China Futures Markets Volatility for High-frequency Time Series[J]. *Systems Engineering — Theory & Practice*, 31(6): 1039–1044.
- 魏宇, (2010). 沪深 300 股指期货的波动率预测模型研究 [J]. *管理科学学报*, 13(2): 66–76.
- Wei Y, (2010). Volatility Forecasting Models for CSI300 Index Futures[J]. *Journal of Management Sciences in China*, 13(2): 66–76.
- Andersen T G, Bollerslev T, (1998). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts[J]. *International Economic Review*, 39(4): 885–905.
- Bâra A, Oprea S V, (2024). An Ensemble Learning Method for Bitcoin Price Prediction Based on Volatility Indicators and Trend[J]. *Engineering Applications of Artificial Intelligence*, 133: 107991.
- Batten J A, Ciner C, Lucey B M, (2010). The Macroeconomic Determinants of Volatility in Precious Metals Markets[J]. *Resources Policy*, 35(2): 65–71.
- Bollerslev T, (1986). Generalized Autoregressive Conditional Heteroskedasticity[J]. *Journal of Econometrics*, 31(3): 307–327.
- Bollerslev T, Hood B, Huss J, Pedersen L H, (2018). Risk Everywhere: Modeling and Managing Volatility[J]. *The Review of Financial Studies*, 31(7): 2729–2773.
- Breiman L, (2001). Random Forests[J]. *Machine Learning*, 45(1): 5–32.
- Chen T, Guestrin C, (2016). Xgboost: A Scalable Tree Boosting System[C]// *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery: 785–794.
- Clark T E, West K D, (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models[J]. *Journal of Econometrics*, 138(1): 291–311.
- Corsi F, (2009). A Simple Approximate Long-Memory Model of Realized Volatility[J]. *Journal of Financial Econometrics*, 7(2): 174–196.
- Degiannakis S, Filis G, (2017). Forecasting Oil Price Realized Volatility Using Information Channels from Other Asset Classes[J]. *Journal of International Money and Finance*, 76: 28–49.
- Deng S K, Xiao C Y, Zhu Y K, Tian Y, Liu Z H, et al. (2022). Dynamic Forecasting of the Shanghai Stock

- Exchange Index Movement Using Multiple Types of Investor Sentiment[J]. *Applied Soft Computing*, 125: 109132.
- Dinh T, Goutte S, Nguyen D K, Walther T, (2022). Economic Drivers of Volatility and Correlation in Precious Metal Markets[J]. *Journal of Commodity Markets*, 28: 100242.
- Elsayed S, Thyssens D, Rashed A, Jomaa H S, Schmidt-Thieme L, (2021). Do We Really Need Deep Learning Models for Time Series Forecasting?[J]. *arXiv Preprint*: 2101.02118.
- Feng L B, Rao H C, Lucey B, Zhu Y Y, (2024). Volatility Forecasting on China's Oil Futures: New Evidence from Interpretable Ensemble Boosting Trees[J]. *International Review of Economics & Finance*, 92: 1595–1615.
- Fu T, Huang D S, Feng L B, Huang X P, (2024). More is Better? The Impact of Predictor Choice on the INE Oil Futures Volatility Forecasting[J]. *Energy Economics*, 134: 107540.
- Greenwell B M, Boehmke B C, (2020). Variable Importance Plots — An Introduction to the Vip Package[J]. *The R Journal*, 12(1): 343.
- Guliyev H, Mustafayev E, (2022). Predicting the Changes in the WTI Crude Oil Price Dynamics Using Machine Learning Models[J]. *Resources Policy*, 77: 102664.
- Guo X Z, Huang D S, Li X F, Liang C, (2023). Are Categorical EPU Indices Predictable for Carbon Futures Volatility? Evidence from the Machine Learning Method[J]. *International Review of Economics & Finance*, 83: 672–693.
- Hansen P R, Lunde A, Nason J M, (2011). The Model Confidence Set[J]. *Econometrica*, 79(2): 453–497.
- Hoerl A E, Kennard R W, (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems[J]. *Technometrics*, 12(1): 55–67.
- Huang J B, Tan N, Zhong M R, (2014). Incorporating Overconfidence into Real Option Decision-Making Model of Metal Mineral Resources Mining Project[J]. *Discrete Dynamics in Nature and Society*, 2014: e232516.
- Inoue A, Jin L, Rossi B, (2017). Rolling Window Selection for Out-of-sample Forecasting with Time-varying Parameters[J]. *Journal of Econometrics*, 196(1): 55–67.
- Jabeur S B, Mefteh-Wali S, Viviani J L, (2024). Forecasting Gold Price with the XGBoost Algorithm and SHAP Interaction Values[J]. *Annals of Operations Research*, 334(1): 679–699.
- Ke G L, Meng Q, Finley T, Wang T F, Chen W, et al. (2017). Lightgbm: A Highly Efficient Gradient Boosting Decision Tree[C]// *Advances in Neural Information Processing Systems*, 30: 3146–3154.
- Lanzano T, Bertram M, De Palo M, Wagner C, Zyla K, et al. (2006). The Contemporary European Silver Cycle[J]. *Resources, Conservation and Recycling*, 46(1): 27–43.
- Liang C, Umar M, Ma F, Huynh T, (2022). Climate Policy Uncertainty and World Renewable Energy Index Volatility Forecasting[J]. *Technological Forecasting and Social Change*, 182: 121810.
- Liang C, Wei Y, Li X F, Zhang Y H, Zhang Y F, (2019). Uncertainty and Crude Oil Market Volatility: New Evidence[J]. *Applied Economics*, 52: 1–15.
- Liang C, Zhang Y, Zhang Y J, (2022). Forecasting the Volatility of the German Stock Market: New Evidence[J]. *Applied Economics*, 54(9): 1055–1070.
- Liu L Y, Patton A J, Sheppard K, (2015). Does Anything Beat 5-minute RV? A Comparison of Realized Measures across Multiple Asset Classes[J]. *Journal of Econometrics*, 187(1): 293–311.
- Lundberg S M, Lee S I, (2017). A Unified Approach to Interpreting Model Predictions[J]. *arXiv Preprint*: 1705.07874.

- Lundberg S M, Lee S I, (2018). Consistent Feature Attribution for Tree Ensembles[J]. arXiv Preprint: 1706.06060.
- Luo X G, Ye Z N, (2015). Predicting Volatility of the Shanghai Silver Futures Market: What is the Role of the U.S. Options Market?[J]. *Finance Research Letters*, 15: 68–77.
- Makridakis S, Spiliotis E, Assimakopoulos V, (2022). M5 Accuracy Competition: Results, Findings, and Conclusions[J]. *International Journal of Forecasting*, 38(4): 1346–1364.
- Maree A, Card P, Murphy A, Kidman P, (2017). GARCH Models with a New Class of Heavy-tailed Distribution for the Silver Returns[J]. Available at SSRN: 3013693.
- Mohsin M, (2023). A Novel Deep-learning Technique for Forecasting Oil Price Volatility Using Historical Prices of Five Precious Metals in Context of Green Financing — A Comparison of Deep Learning, Machine Learning, and Statistical Models[J]. *Resources Policy*, 86: 104216.
- Pesaran M H, Timmermann A, (1992). A Simple Nonparametric Test of Predictive Performance[J]. *Journal of Business & Economic Statistics*, 10(4): 461–465.
- Rahimikia E, Poon S H, (2020). Machine Learning for Realised Volatility Forecasting[J]. Available at SSRN: 3707796.
- Santamaria-Bonfil G, Frausto-Solis J, Vázquez-Rodarte I, (2013). Volatility Forecasting Using Support Vector Regression and a Hybrid Genetic Algorithm[J]. *Computational Economics*, 45(2): 175–195.
- Shapley L S, (1953). A Value for N-person Games[C]// Kuhn H W, Tucker A W. *Contributions to the Theory of Games*. Princeton, NJ: Princeton University Press.
- Shu Q, Xiong H, Jiang W J, Mamon R, (2023). A Novel Perspective on Forecasting Non-ferrous Metals' Volatility: Integrating Deep Learning Techniques with Econometric Models[J]. *Finance Research Letters*, 58: 104482.
- Tibshirani R, (1996). Regression Shrinkage and Selection Via the Lasso[J]. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1): 267–288.
- Truong V H, Tangaramvong S, Papazafeiropoulos G, (2024). An Efficient Lightgbm-based Differential Evolution Method for Nonlinear Inelastic Truss Optimization[J]. *Expert Systems with Applications*, 237: 121530.
- Wei Y, Liang C, Li Y, Zhang X H, Wei G W, (2020). Can CBOE Gold and Silver Implied Volatility Help to Forecast Gold Futures Volatility in China? Evidence Based on HAR and Ridge Regression Models[J]. *Finance Research Letters*, 35: 101287.
- Yao X, Hui X F, Kang K C, (2021). Can Night Trading Sessions Improve Forecasting Performance of Gold Futures' Volatility in China?[J]. *Journal of Forecasting*, 40(5): 849–860.
- Yun K K, Yoon S W, Won D, (2021). Prediction of Stock Price Direction Using a Hybrid GA-XGBoost Algorithm with a Three-stage Feature Engineering Process[J]. *Expert Systems with Applications*, 186: 115716.
- Zhou H N, Lu X J, (2023). Investor Attention on the Russia-ukraine Conflict and Stock Market Volatility: Evidence from China[J]. *Finance Research Letters*, 52: 103526.
- Zou H, Hastie T, (2005). Regularization and Variable Selection via the Elastic Net[J]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2): 301–320.