

双融合模型在香港股市的波动率量化交易策略研究

颜轲越¹, 王宁², 李莹^{3*}

¹澳门大学, 蔡继有书院, 澳门

²德州学院, 体育学院, 山东 德州

³北京理工大学珠海学院, 中美国际学院, 广东 珠海

收稿日期: 2024年4月9日; 录用日期: 2024年4月17日; 发布日期: 2024年5月17日

摘要

在金融市场中, 波动率反映了资产价格的不稳定性和风险程度。波动率量化交易策略是一种利用波动率来指引投资者进行交易的方法, 并能够带来一定的收益。本文的目的是探讨时间序列模型和机器学习模型相结合的双融合模型在波动率量化交易策略中的应用, 以提升策略的效果和稳健性。我们采用了3种时间序列模型Generalized Autoregressive Conditional Heteroskedasticity (GARCH)、Glosten Jagannathan Runkle GARCH (GJR-GARCH)和Exponential GARCH (EGARCH), 以及4种传统机器学习模型Random Forest (RF)、Adaboost (ADA)、Gradient Boosting Decision Tree (GBDT)和Histogram Based Gradient Boosting (Hist-GB), 组合成了12种双融合模型对波动率进行预测。同时, 我们进行了模拟交易的回测实验, 以评估该波动率量化交易策略的性能。该策略在香港市场中的股票都获得了较好的收益, 为投资者提供了有价值的参考。

关键词

时间序列模型, 机器学习模型, 波动率预测, 量化交易

Research on Double Fusion Modeling for Volatility Quantitative Trading Strategies in Hong Kong Stock Market

Keyue Yan¹, Ning Wang², Ying Li^{3*}

¹Choi Kai Yau College, University of Macau, Macau

²School of Physical Education, Dezhou University, Dezhou Shandong

³College of Global Talents, Beijing Institute of Technology (Zhuhai), Zhuhai Guangdong

*通讯作者。

文章引用: 颜轲越, 王宁, 李莹. 双融合模型在香港股市的波动率量化交易策略研究[J]. 金融, 2024, 14(3): 844-855.

DOI: 10.12677/fin.2024.143090

Received: Apr. 9th, 2024; accepted: Apr. 17th, 2024; published: May 17th, 2024

Abstract

In the financial markets, volatility is indicative of the fluctuation in asset prices and the associated level of risk. The volatility quantitative trading strategy is an approach that leverages volatility to make trading decisions and generate specific returns. This paper aims to explore the application of double fusion modeling that combines time series models and machine learning models within the volatility quantitative trading strategy to augment its efficiency and stability. We employ 3 time series models: Generalized Autoregressive Conditional Heteroskedasticity (GARCH), Glostn Jagannathan Runkle GARCH (GJR-GARCH), and Exponential GARCH (EGARCH); alongside 4 traditional machine learning models: Random Forest (RF), Adaboost (ADA), Gradient Boosting Decision Tree (GBDT), and HistogramBased Gradient Boosting (Hist-GB). These models are amalgamated into 12 double fusion models for prediction analysis. Meanwhile, we execute backtesting trading experiments to assess the strategy's performance. The strategy has demonstrated superior returns in Hong Kong market, and offers a valuable trading reference for investors.

Keywords

Time Series Models, Machine Learning Models, Volatility Prediction, Quantitative Trading

Copyright © 2024 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

在如今的金融、经济等商业场景中，计算机技术有着广泛的应用[1] [2]。而在股票市场中，计算机技术主要用于金融数据分析与挖掘、算法交易、风险管理和量化交易策略的开发[3]。其中量化交易策略是利用统计模型进行自动投资交易策略的过程。在此过程中，量化交易策略会严格按照最初设计好的交易逻辑进行交易，不会因为投资者的主观情绪做出买入和卖出股票的决定[4]。现如今有不少金融领域的研究着重于如何使用计算机技术构造量化策略。Yan 等人使用了 RF 模型和传统布林带策略，构造出了利用模型预测结果和布林带进行买卖操作的策略[5]。而在股票交易的涨跌中，波动率经常用于衡量股票价格在某个时间范围内上下波动的幅度。波动率在金融工程领域比较常见，通常被定义为股票日收益率的标准差。被用于衡量股票收益数据分散程度的统计量和衡量金融市场中的期权价值[6]。波动率越大，则表示股票收益率的分散程度越大；股票价格的波动程度越大。因此，波动率可以用来度量股票的风险和偏离程度。在此研究中，我们尝试使用不同的特征变量构造方法和机器学习模型组成的双融合模型预测股票未来的波动率，利用预测的波动率和历史波动率线构造量化交易策略。为了方便观察收益或损失，我们为每只股票建立了一个模拟交易环境。通过香港股市的股票实验结果，我们发现采用波动率量化交易策略可以获得正向的收益。

2. 文献综述

目前许多研究人员和投资人希望可以借助数学统计模型或者计算机技术算法在投资理财中获得超额收益。在机器学习和深度学习被广泛应用之前，传统的时间序列模型在金融预测中占有很重要的地位。

Kumbure 等人对近些年的金融股票价格预测做了很详细的描述总结以及附上了相关的研究报告[7]。ARIMA 和 GARCH 等传统统计时间序列模型被广泛用于预测股价或者风险预测。目前基于传统统计时间序列模型的使用流程都已经趋于成熟，并在预测和风险管理领域均有广泛的应用。近些年随着人工智能和大数据等概念的热度持续提高，很多研究都致力于使用机器学习分类模型或者回归模型预测股票价格的涨跌或者股票未来的价值[8] [9]。在传统的机器学习应用领域，Basak 等人的研究使用了 Xgboost 和 RF 对不同交易期下的美国股票价格进行预测，得到了在不同标签的设定下股票未来的涨跌会有显著的不同[10]。类似的，有研究人员使用集成学习模型来预测经过滑动平均法处理后的股票价格，发现使用滑动平均线作为标签的预测误差较小[11]。一些学者认为机器学习模型训练前需要构造出与研究问题相关的特征，模型才能表现出比较好的拟合效果。Alsulmi 等人构造了指数移动平均线、三重指数移动平均线、能量潮等 31 个股票技术特征来预测沙特的股票[12]。

随着计算机技术的发展，深度学习算法和更加丰富的数据源也被用于研究股票涨跌。Vijh 等人使用了神经网络模型对在美国上市的四间公司股票价格进行预测，通过与机器模型对比，发现神经网络模型具有更小的误差[13]。除了单纯使用历史股价数据进行预测，有的研究会考虑使用更加丰富的数据源进行分析，Khan 等人通过收集社交媒体和财经新闻数据，经过筛选分析得到投资者对金融市场乐观或者悲观的情绪分数，从而方便对未来股票价格进行更加准确的预测，研究中还指出使用文本数据构造的变量对不同类型的股票预测具有不同的效果[14]。

至今为止，大部分的研究都侧重于股票的涨跌和未来价格预测，关注股票的波动率预测的研究却相对较少。然而现实交易中波动率在金融市场中也占有很重要的地位。有研究尝试结合 GARCH 模型和 Support Vector Machine (SVM)，构造出了 GARCH-SVR 模型对市场风险进行预测[15]。然而这些研究都只是停留在预测和误差对比的阶段，几乎没有研究尝试使用预测的波动率作为量化策略中的交易信号。在本文中，我们提出了一种使用预测的波动率来进行量化交易策略的新想法。我们的目标是通过双融合模型预测第二天的波动率走势，结合历史的波动率趋势线来做出买入和卖出操作的决策。通过一套完整的模拟交易框架，我们能够科学地观察基于波动率的量化交易策略的盈利能力。

3. 研究方法

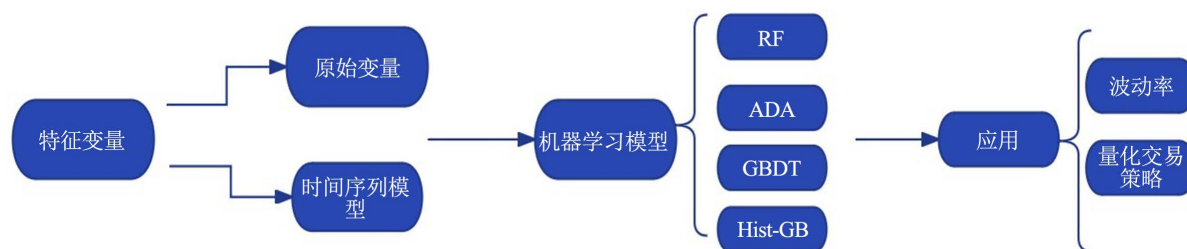


Figure 1. Experimental framework
图 1. 实验流程框架

本研究中的整体框架如图 1 所示，我们的研究目标是使用时间序列模型和机器学习相结合的方法来预测香港股市中的 7 只股票的未來波动率，最后使用未來波动率构造合适的量化交易策略。首先，我们会使用原始股票数据集构造有利于对波动率预测的特征变量。在建模阶段，我们首先会使用 GARCH、GJR-GARCH 和 EGARCH 模型对历史波动率进行建模，在本研究中我们将使用滑动窗口的方法对时间序列模型进行动态建模。然后使用构建的特征和时间序列模型建模预测的波动率相结合，组成全新的特征变量，再使用 4 种机器学习模型对未來的波动率进行预测。最终，我们根据不同双融合模型预测的未來

波动率进行模拟交易实验，比较不同双融合模型下的波动率量化交易策略的获利能力。

3.1. 数据准备和预处理

香港拥有完善的市场政策和悠久的金融产品交易历史。在数据选取方面，我们从香港交易所选取了 7 只股票作为样本，它们分别是汇丰控股(0005)、香港交易所(0388)、电能实业(0006)、港铁公司(0066)、长江基建集团(1038)、腾讯控股(0700)和比亚迪股份(1211)。我们在雅虎金融网站下载了这些股票的日间交易数据集，时间跨度为 2010 年 1 月 1 日至 2023 年 12 月 31 日 [16]。这些股票数据包括五个变量：开盘价(Open)、最高价(High)、最低价(Low)、收盘价(Close)和成交量(Volume)。除了这些原始数据，我们还会构造波动率相关的特征。首先我们根据收盘价计算出股票的每日收益率，记为 R_t ，而波动率的定义则为过去一段时间范围内的股票每日收益率的标准差，在此项研究中，我们定义波动率为过去 30 天的标准差，即 $n = 30$ 。

$$R_t = \frac{Close_t - Close_{t-1}}{Close_{t-1}}$$

$$Volatility_t = \sqrt{\frac{1}{n-1} \sum_{i=0}^{n-1} (R_{t-i} - \mu)^2}$$

除了当天的波动率 $Volatility_t$ ，过去五天的波动率特征 $Volatility_{t-1}$ 、 $Volatility_{t-2}$ 、 $Volatility_{t-3}$ 、 $Volatility_{t-4}$ 和 $Volatility_{t-5}$ 也被用作机器学习模型的特征变量。这些自变量使算法更容易识别股票波动率变化的趋势。同时，我们使用 $Volatility_{t+1}$ 作为机器学习模型的因变量，即未来一天的波动作为将要预测变量。

另外，我们会使用传统的波动率预测时间序列模型，例如 GARCH、GJR-GARCH 和 EGARCH 模型对历史波动率进行建模，我们将使用这三种模型预测的波动率作为特征输入进机器学习模型中。GARCH 模型是一种经典的时间序列模型，经常用于预测估计金融资产收益率的波动性。GARCH 模型能够捕捉到波动率的聚集性和厚尾性。标准的 GARCH(p,q)模型的公式如下所示。其中预测变量 σ_t^2 是时间 t 的条件方差， ϵ_t 是收益率的扰动项。剩下的 ω ， α_i 和 β_j 是模型参数。

$$\epsilon_t \sim N(0,1)$$

$$\epsilon_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

而 GJR-GARCH 模型是 GARCH 模型的更新版，它通过引入一个指示函数来捕捉金融资产收益率的不对称波动性，其公式表示如下。其中 N_{t-i} 为指示函数，当 ϵ_{t-i} 小于 0 时， N_{t-i} 取 1；否则取 0 [17]。

$$\sigma_t^2 = \omega + \sum_{i=1}^p (\alpha_i + \gamma_i N_{t-i}) \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

EGARCH 模型则是另一种 GARCH 模型的更新版，EGARCH 模型的特点是对数波动率的建模，这样就不需要对模型参数的范围进行限制。EGARCH 模型的表示如下所示 [18]。

$$\ln(\sigma_t^2) = \omega + \sum_{i=1}^p \alpha_i \left(\left| \frac{\epsilon_{t-i}}{\sigma_{t-i}} \right| - \sqrt{\frac{2}{\pi}} \right) + \sum_{j=1}^q \beta_j \ln(\sigma_{t-j}^2)$$

这三种模型的主要区别在于它们对波动率的建模方式不同。在此研究中我们参考了 [19] 中的建模方

式，并对其加以改进，使用滑动窗口法实时更新时间序列模型，使其预测的 σ 值更具备参考意义。在与机器学习模型结合的过程中，在其它特征变量固定的情况下，我们将使用一个时间序列模型搭配一个机器学习的双融合模型形式进行最终建模。

3.2. 机器学习模型

在对波动率预测的回归模型中，我们一共选择了 4 种传统的机器学习模型：RF、ADA、GBDT 和 Hist-GB。在每个模型的参数选择上，我们在这些机器学习模型的预训练过程中使用网格搜索法来挑选模型的最佳参数，以便每个模型都具备最好的泛化能力和适应能力。我们使用时间序列模型和机器学习模型相结合的方式预测未来的波动率。其中时间序列模型为 GARCH、GJR-GARCH 和 EGARCH，与 4 种机器学习模型的结合过程中将产生 12 种双融合模型组合，例如 GARCH-RF、GARCH-ADA、GJR-GARCH-ADA、GJR-GARCH-GBDT、EGARCH-RF 和 EGARCH-Hist-GB 等模型。在计算实际值与模型预测值之间的误差时，我们使用三个常见的回归指标：R Squared、均方根误差(RMSE)和平均绝对误差(MAE)。其中 R Squared 越接近于 1，则表示模型的解释性越好，模型的预测可靠性越高。当 RMSE 和 MAE 的值越小，越接近于 0，则表示机器学习模型的拟合效果越好。这三者的公式如下所示。

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

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

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

3.3. 波动率量化交易策略

股票量化交易是一种利用数学统计模型和计算机程序，根据市场的历史数据和交易规律，自动进行股票买卖的交易方式。其根本目的是通过科学的方法，降低投资者自身的情绪干扰，提高效率和收益。其中量化交易的类别有很多种，例如趋势跟踪策略，套利策略和均值回归策略。而此文涉及的波动率量化交易策略参考了均值回归的核心思想。均值回归策略是根据股票的价格与其历史均值的偏离程度，判断其是否存在超买或超卖的现象，并在当前交易价格和历史均值差价较大时，买入或卖出，等待价格回归其股票本身应有的价值时，从而获利[20]。在本中，我们使用历史波动率取代历史价格，同时我们将定义一条上轨道线和下轨道线。其中上轨道线为过去 70 天里波动率的最大值，下轨道线为 70 日的波动率均线。公式定义如下所示。

$$UpperTrack_t = \text{Max}(Volatility_t, Volatility_{t-1}, \dots, Volatility_{t-69})$$

$$LowerTrack_t = \frac{Volatility_t + Volatility_{t-1} + \dots + Volatility_{t-69}}{70}$$

当机器学习模型预测的未来一天的波动率 $Volatility_{t+1}$ 大于当天的上轨道 $UpperTrack_t$ 时，我们认为此时单股的交易市场的波动率较大，即市场投资者的情绪处于比较恐慌或者狂热状态，策略会对股票进行买入。而当 $Volatility_{t+1}$ 小于 $LowerTrack_t$ 时，我们认为此时个股在交易市场中的波动率回归到比较平稳的阶段，即投资者恐慌或者热度情绪散去，策略会卖出持有的股票。此文使用一些表现指标来评价波动率量化交易策略的收益和风险。个股在测试集上的回测收益可以简单表示为模拟交易结束后的总资产减去

最初的投资金额，为了更方便科学地观察收益情况，我们还可以使用年化收益率来评估投资策略效果。它表示投资一年时的预期收益率，计算单位为%。在评估风险方面，我们使用最大回撤，计算单位也为%。其表示投资策略在一段时间内发生的最大亏损幅度。最大回撤的数值越小越好，越大说明风险越大。策略有效性也越差。

4. 实验结果和讨论

在本章节中，我们会详细地报告研究中的波动率量化交易策略在不同双融合模型和不同股票下的表现。其中 R Squared、RMSE 和 MAE 会被用于评估双融合模型的效果；策略收益，年化收益率和最大回撤会被用于评估量化交易策略的表现。我们会在汇丰控股(0005)、香港交易所(0388)、电能实业(0006)、港铁公司(0066)、长江基建集团(1038)、腾讯控股(0700)和比亚迪股份(1211)这 7 只香港股票进行模拟回测测试。数据的时间跨度为 2010 年 1 月 1 日至 2023 年 12 月 31 日；其中 2010 年~2019 年的日间交易数据被用于训练双融合模型，剩下的 2020 年~2023 年的日间交易数据被用于测试双融合模型表现效果和量化交易回测。

4.1. 回归结果

在本节中，我们首先提供了 3 种时间序列模型和 4 种机器学习模型组成的 12 种双融合模型的预测效果，我们会使用均值 ± 标准差(Mean ± SD)来评估回归预测的总体效果的有效性。其中图 2 显示的是 GJR-GARCH-ADA 双融合模型在 0005 股票上的波动率预测表现。从图中可以看出机器学习模型在拟合波动率的任務上表现较为优异。

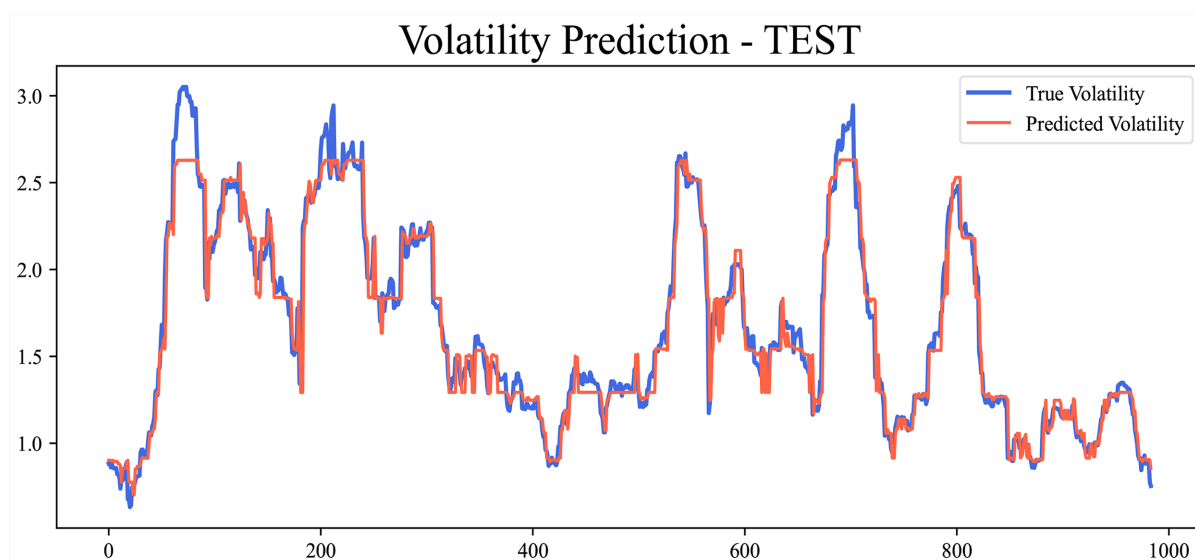


Figure 2. Volatility prediction of stock 0005

图 2. 股票 0005 的波动率预测

在表 1 和图 3 中，我们详细地展示了每种时间序列模型和机器学习模型组合后的双融合模型在 7 只香港股票波动率上的平均表现。从箱线图中可以清晰地看出 GARCH、GJR-GARCH 和 EGARCH 分类下的机器学习模型预测效果下不会有太大的分别，所有对应的机器学习模型基本上有着相同的分布。现在针对机器学习模型进行讨论，我们可以看出相比于其它模型，GBDT 模型拟合效果比较一般，同时误差也是最大的。而 ADA 模型的表现稍微比 GBDT 模型好，误差的上下限范围比 GBDT 小。而 RF 模型和

Hist-GB 模型均有比较好的拟合效果。

Table 1. Average regression results for time series models and machine learning models
表 1. 时间序列模型和机器学习模型回归结果的平均表现

时间序列	机器学习	R Squared	RMSE	MAE
GARCH	RF	0.915 ± 0.066	0.189 ± 0.097	0.087 ± 0.035
	ADA	0.899 ± 0.068	0.209 ± 0.096	0.117 ± 0.042
	GBDT	0.893 ± 0.069	0.219 ± 0.108	0.107 ± 0.046
	Hist-GB	0.923 ± 0.067	0.179 ± 0.097	0.087 ± 0.036
GJR-GARCH	RF	0.916 ± 0.065	0.187 ± 0.096	0.086 ± 0.034
	ADA	0.901 ± 0.068	0.206 ± 0.096	0.116 ± 0.041
	GBDT	0.886 ± 0.067	0.225 ± 0.106	0.110 ± 0.049
	Hist-GB	0.924 ± 0.068	0.178 ± 0.098	0.086 ± 0.035
EGARCH	RF	0.916 ± 0.065	0.187 ± 0.096	0.085 ± 0.035
	ADA	0.898 ± 0.068	0.210 ± 0.097	0.117 ± 0.041
	GBDT	0.892 ± 0.066	0.216 ± 0.100	0.105 ± 0.040
	Hist-GB	0.924 ± 0.067	0.177 ± 0.099	0.084 ± 0.036

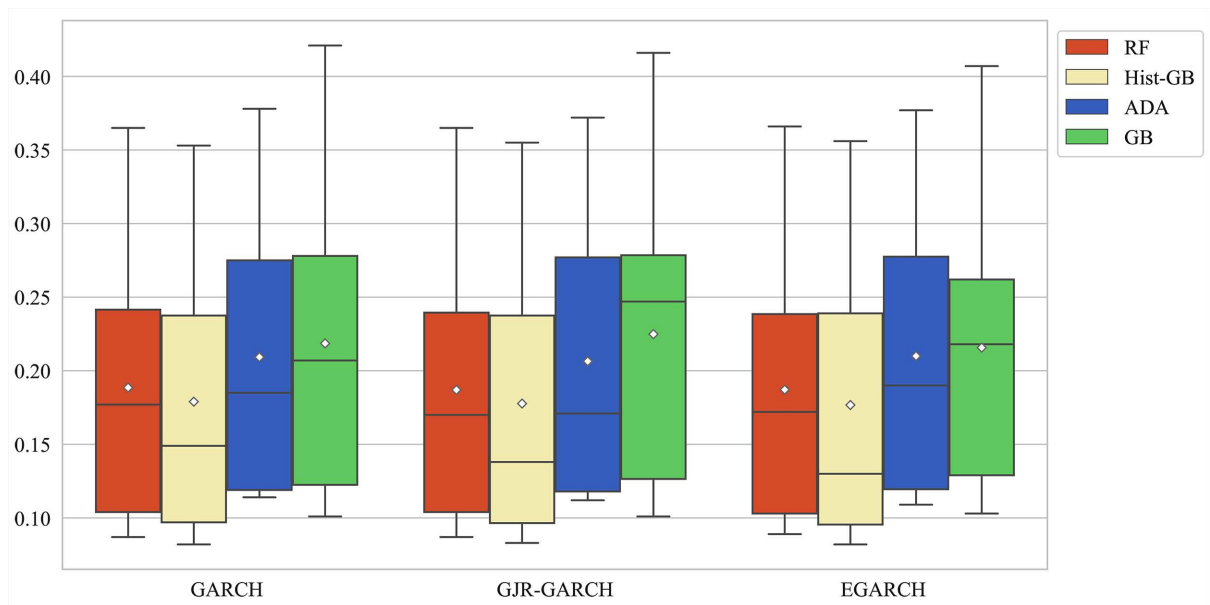


Figure 3. Comparisons of RMSE
图 3. RMSE 的对比

4.2. 模拟交易结果

类似于章节 4.1，在本节中，我们提供了基于时间序列模型和机器学习模型分类下的量化交易策略获利表现。在模拟交易开始前，系统默认初始资金为 100000 元，最终我们根据获利情况和风险控制来评估不同双融合模型下的投资表现。其中图 4 表示的是 GJR-GARCH-ADA 双融合模型在 0005 股票上的量化交易回测可视化，从图中我们可以看出 GJR-GARCH-ADA 双融合模型能够根据预测波动率准确地选择在

股票的低价值期买入。同时能够在股票恢复至正常价值或者高价值时期发出卖出信号，从而获取利润。此模型在回测结束后资金为 140046.25 元，具备了较好的盈利能力。

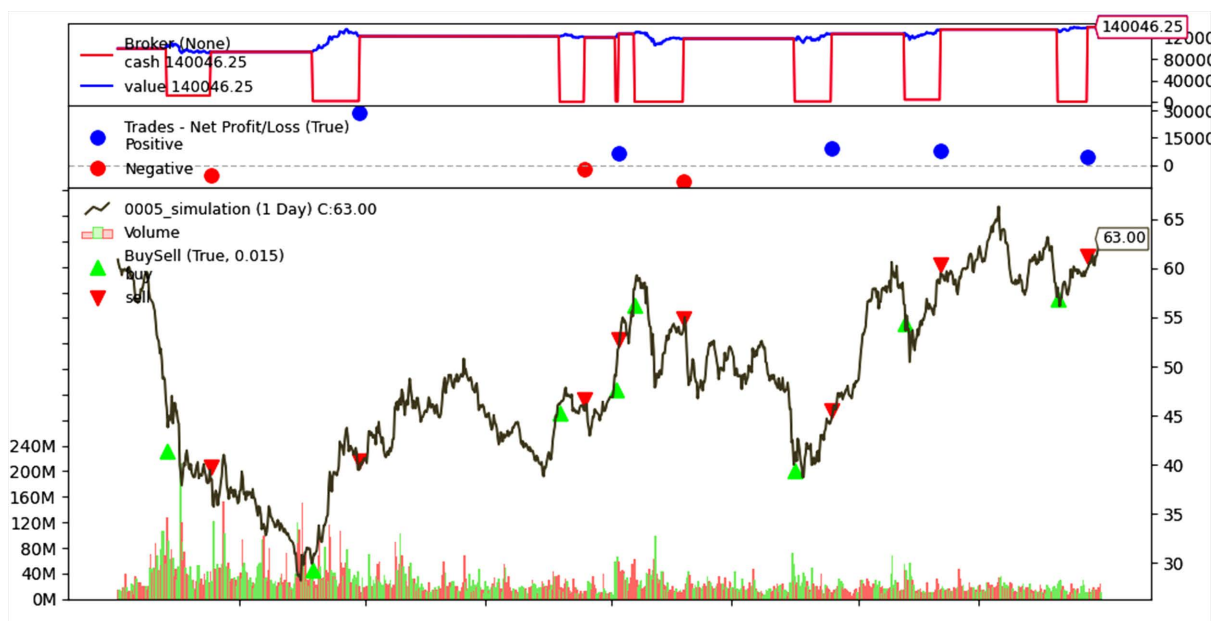


Figure 4. Trading simulation result of stock 0005

图 4. 股票 0005 的模拟交易结果

为了更加直观地观察所有双融合模型在每只股票上的量化交易表现，我们把 7 只股票的收益表现和风险表现做了汇总，并在表 2 中显示。为了更符合现实中的实际情况，方便投资者和研究人员了解策略的利润上限和下限，我们使用了 95% 的置信区间取代了均值±标准差来表示获利表现和风险控制表现。从中我们可以看出，所有股票的 95% 置信区间范围内的利润和年化收益率均为正数。但是区间范围内的数值均有较大的差异，例如汇丰控股(0005)、香港交易所(0388)、电能实业(0006)和港铁公司(0066)的收益率不高，但是长江基建集团(1038)、腾讯控股(0700)和比亚迪股份(1211)均有较好的表现，年化收益率达到非常高的水平，可以给投资者带来丰厚的利润。造成这样的现象的原因是每只股票背后的公司都是独立的；行业也有较大区别。公司的决策层会根据市场、政治、经济等原因做出不一样的发展决定，从而影响着公司的未来和对应的股票价格。我们构造的双模型波动率量化交易策略在面对不同的股票时，依然有着比较正向的反馈，也证明了我们策略构造的合理性和参考性。

Table 2. Average investment results for each stock

表 2. 每只股票投资结果的平均表现

股票	回测收益	年化收益率	最大回撤
汇丰控股(0005)	[14131.55, 26674.50]	[3.458, 6.165]	[25.013, 28.075]
香港交易所(0388)	[14090.24, 37377.89]	[3.318, 8.366]	[29.906, 38.814]
电能实业(0006)	[6223.03, 15676.78]	[1.541, 3.773]	[18.402, 24.028]
港铁公司(0066)	[2948.62, 14567.25]	[0.709, 3.511]	[14.090, 18.971]
长江基建集团(1038)	[17723.82, 34242.73]	[4.274, 7.725]	[13.123, 24.656]
腾讯控股(0700)	[64337.28, 79780.30]	[13.526, 16.192]	[17.477, 19.202]
比亚迪股份(1211)	[18341.70, 50677.73]	[4.127, 10.947]	[22.989, 38.609]

Table 3. Average investment results for time series models

表 3. 时间序列模型投资结果的平均表现

时间序列	回测收益	年化收益率	最大回撤
GARCH	[17971.66, 37028.33]	[4.110, 8.041]	[20.435, 27.548]
GJR-GARCH	[20229.00, 37767.96]	[4.660, 8.246]	[19.465, 26.971]
EGARCH	[18166.34, 38891.05]	[4.163, 8.335]	[20.777, 27.669]

Table 4. Average investment results for machine learning models

表 4. 机器学习模型投资结果的平均表现

机器学习	回测收益	年化收益率	最大回撤
RF	[16388.35, 31247.66]	[3.899, 7.033]	[15.196, 24.227]
ADA	[10561.16, 35903.20]	[2.487, 7.636]	[21.550, 31.257]
GBDT	[21551.60, 47197.88]	[4.868, 10.022]	[20.501, 28.306]
Hist-GB	[21408.19, 42481.08]	[4.904, 9.226]	[21.998, 27.452]

在模型方面，我们首先展示了基于时间序列模型分类下的量化交易策略获利表现，如表 3 所示。可以看出 3 种不同的时间序列模型有着相似的表现，其中基于 GJR-GARCH 和 EGARCH 下的模型表现更好，收益区间比较优秀。GARCH 模型只是稍微逊色于这两者。造成 3 种模型表现有差异的原因是不同模型对波动率预测的高点和低点均有不同的结果，因此造成了不同时间序列模型在量化交易回测时期的买入点和卖出点出现差异，从而出现了不同的获利情况。在风险控制方面，我们可以看到 3 种模型的最大回撤均处于 20%~27% 之间。差距不是非常地明显。接下来我们按照机器学习模型分类来进行计算量化交易策略获利表现，如表 4 所示。我们依然可以发现虽然 GBDT 在预测能力上表现不如其它 3 种机器学习模型，但是在量化交易的买点和卖点上选择得当，具有最好的收益能力。于此同时，GBDT 回撤能力的控制也比 ADA 和 Hist-GB 模型好。在预测中表现优秀的 RF 模型虽然收益能力不如 GBDT，但是在风险控制上有着最有效的表现，其最大回撤的 95% 置信区间上限比其它 3 种模型均低几个百分点。

Table 5. Average investment results for time series models and machine learning models

表 5. 时间序列模型和机器学习模型投资结果的平均表现

时间序列	机器学习	回测收益	年化收益率	最大回撤
GARCH	RF	[7057.85, 38935.16]	[1.953, 8.621]	[10.141, 31.460]
	ADA	[-5353.19, 47817.09]	[-0.772, 10.038]	[15.911, 36.467]
	GBDT	[7355.13, 58283.19]	[1.861, 12.459]	[17.354, 32.158]
	Hist-GB	[8507.41, 57397.32]	[2.190, 12.255]	[18.381, 30.061]
GJR-GARCH	RF	[7880.61, 42081.99]	[2.059, 9.328]	[11.069, 26.299]
	ADA	[-634.72, 54076.05]	[0.328, 11.324]	[14.813, 37.630]
	GBDT	[9345.29, 56937.05]	[2.444, 12.128]	[12.556, 32.904]
	Hist-GB	[13300.03, 49001.53]	[3.341, 10.674]	[19.315, 31.160]
EGARCH	RF	[8952.33, 38000.09]	[2.359, 8.477]	[9.322, 29.977]
	ADA	[-5382.14, 48870.00]	[-0.830, 10.281]	[17.414, 36.187]
	GBDT	[4972.81, 69354.98]	[1.600, 14.178]	[18.898, 32.552]
	Hist-GB	[7038.36, 56423.13]	[1.921, 12.008]	[19.004, 30.431]

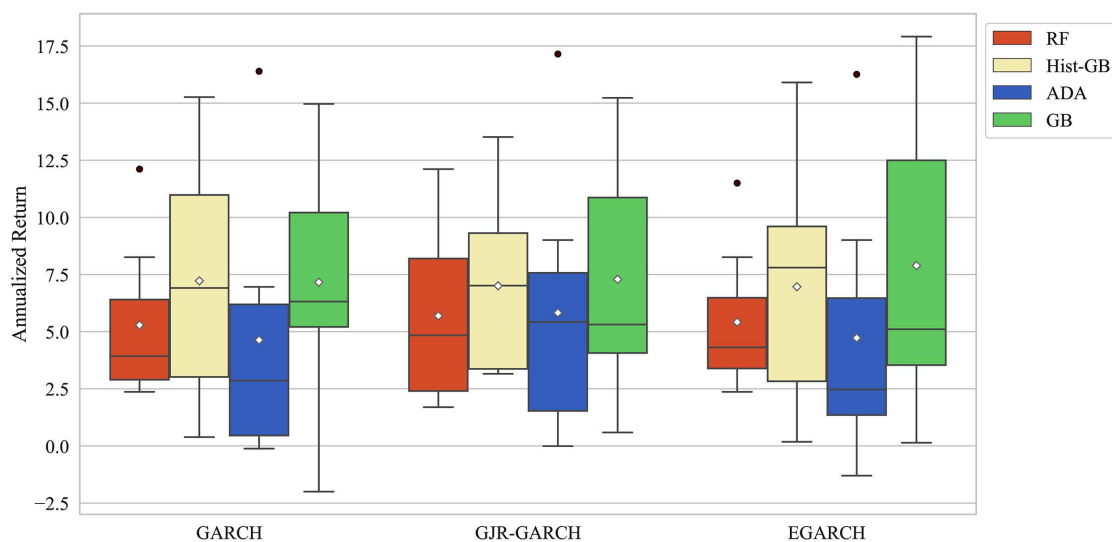


Figure 5. Comparisons of annualized return

图 5. 年化收益率的对比

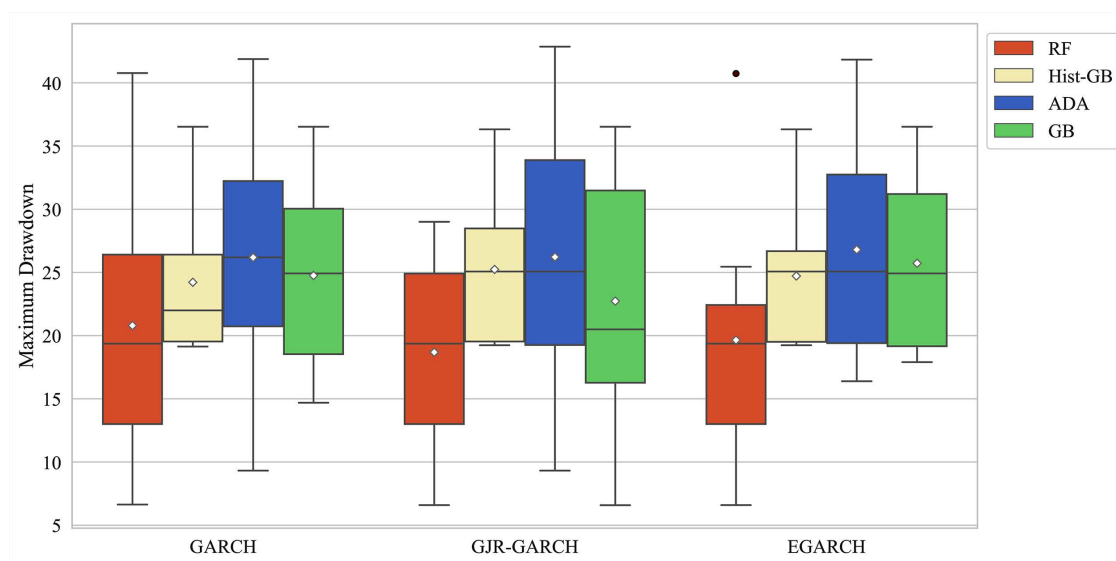


Figure 6. Comparisons of maximum drawdown

图 6. 最大回撤的对比

我们对模型再进行细分, 可以把 3 种时间序列模型和 4 种机器学习模型组成的 12 种双融合模型的量化交易结果做成表 5。其中更加直观的年化收益率和最大回撤可视化结果如图 5 和图 6 所示。与表 3 的结果类似, 我们从表 5、图 5 和图 6 中看出 GJR-GARCH 和 EGARCH 下的 4 种机器学习模型表现总体上优于 GARCH 模型下的机器学习模型。在相同的时间序列模型下, 机器学习模型的获利情况也各不相同。比较意外的是 ADA 模型的表现。在表 4 中, ADA 的表现和其它模型不会有非常大的差距, 但是在细分后, 其平均收益的下置信区间都出现了负数, 即表示投资者极有可能面临亏损。而且其 95% 置信区间上限的数值也不是非常高。因此可以推断出 ADA 模型在 4 种机器学习模型中综合表现最一般。与表 4 结果相似的是, 不同时间序列模型下的 GBDT 依然具有最好的收益表现, 最大回撤也处于比较正常的水平。Hist-GB 则紧跟 GBDT 后面, 甚至在 GARCH 模型下表现优于 GBDT。而 RF 模型依然处于比较稳定的状

态, 虽然收益率表现不如 GBDT 和 Hist-GB, 但是在最大回撤控制上具有最好的表现, 即在股票买入和卖出的过程中资产价值不会出现非常剧烈的波动。对于投资者而言, 决定使用哪一种波动率量化交易策略取决于自身的预期收益和风险承受能力。对于比较激进, 敢于冒险, 并且能接受一定风险和亏损的投资者, GJR-GARCH-GBDT、EGARCH-GBDT、GJR-GARCH-Hist-GB 和 EGARCH-Hist-GB 等双融合模型比较适合在波动率量化交易策略中使用。而对于比较保守, 风险承受能力不强的投资者, GJR-GARCH-RF 或者 EGARCH-RF 等双融合模型比较适合在波动率量化交易策略中使用。最后, 本研究提供的波动率量化交易策略的思路和应用给投资者带来了更多的收益, 为投资者和研究人员将波动率, 时间序列模型, 机器学习模型, 与量化交易的结合提供了有意义的实战参考。

5. 结论与展望

本研究以香港股市中的 7 只股票为研究对象, 运用了 GARCH、GJR-GARCH 和 EGARCH 这 3 种统计时间序列模型和 RF、ADA、GBDT 和 Hist-GB 这 4 种机器学习模型, 对上述股票的未来一天波动率进行了预测, 并基于预测结果构建了量化交易策略, 评估了投资表现。实验结果表明这些新结合的模型中 GBDT 和 Hist-GB 得到了最优的获利结果。综上所述, 通过使用真实的股票交易数据证实了时间序列模型和机器学习模型组合的双融合模型在波动率量化交易策略中具有应用价值, 本研究为大众提供了全新的投资交易思路和方法, 也为构造量化交易策略提供了可靠的参考。在未来的研究中, 我们可以进一步地扩展研究样本, 不仅仅在香港市场的股票进行实验, 还可以在其它国家或者地区的市场进行不同主题股票进行实验。以及探索更多的模型、训练方法、其它交易模式来提高波动率量化交易策略的效果和稳健性。

作者贡献

颜轲越: 收集香港市场股票的交易数据, 建模流程框架的构思, 建模实验检验, 撰写原稿。

王宁: 对香港交易的股票提供选股和投资建议, 对原稿进行润色。

李莹: 监督项目进度, 对建模流程细节提供建议指导, 对原稿进行润色。

参考文献

- [1] 颜轲越, 王祎萌, 李莹. 基于机器学习的共享单车需求预测[J]. 计算机科学与应用, 2022, 12(3): 697-706. <https://doi.org/10.12677/CSA.2022.123071>
- [2] 李莹, 颜轲越. 基于改进 LRFMC 模型的航空公司客户分类[J]. 计算机科学与应用, 2022, 12(5): 1341-1349. <https://doi.org/10.12677/CSA.2022.125133>
- [3] Wang, Y. and Yan, K. (2022) Prediction of Significant Bitcoin Price Changes Based on Deep Learning. *5th International Conference on Data Science and Information Technology*, Shanghai, 22-24 July 2022, 1-5. <https://doi.org/10.1109/DSIT55514.2022.9943971>
- [4] Wang, Y. and Yan, K. (2023) Machine Learning-Based Quantitative Trading Strategies across Different Time Intervals in the American Market. *Quantitative Finance and Economics*, 7, 569-594. <https://doi.org/10.3934/QFE.2023028>
- [5] Yan, K., Wang, Y. and Li, Y. (2023) Enhanced Bollinger Band Stock Quantitative Trading Strategy Based on Random Forest. *Artificial Intelligence Evolution*, 4, 22-33. <https://doi.org/10.37256/aie.4120231991>
- [6] Li, Y. and Yan, K. (2023) Prediction of Barrier Option Price Based on Antithetic Monte Carlo and Machine Learning Methods. *Cloud Computing and Data Science*, 4, 77-86. <https://doi.org/10.37256/ccds.4120232110>
- [7] Kumbure, M.M., Lohrmann, C., Luukka, P. and Porras, J. (2022) Machine Learning Techniques and Data for Stock Market Forecasting: A Literature Review. *Expert Systems with Applications*, 197, Article ID: 116659. <https://doi.org/10.1016/j.eswa.2022.116659>
- [8] 颜轲越, 李莹, 王祎萌. 基于信号分解和机器学习模型的股票价格预测[J]. 计算机科学与应用, 2022, 12(4): 1080-1088. <https://doi.org/10.12677/CSA.2022.124111>
- [9] Wang, Y. and Yan, K. (2023) Application of Traditional Machine Learning Models for Quantitative Trading of Bitcoin.

-
- Artificial Intelligence Evolution*, **4**, 34-48. <https://doi.org/10.37256/aie.4120232226>
- [10] Basak, S., Kar, S., Saha, S., Khaidem, L. and Dey, S.R. (2019) Predicting the Direction of Stock Market Prices Using Tree-Based Classifiers. *The North American Journal of Economics and Finance*, **47**, 552-567. <https://doi.org/10.1016/j.najef.2018.06.013>
- [11] Yan, K. and Wang, Y. (2023) Prediction of Bitcoin Prices' Trends with Ensemble Learning Models. *5th International Conference on Computer Information Science and Artificial Intelligence*, Chongqing, 16-18 September 2022, 900-905. <https://doi.org/10.1117/12.2667793>
- [12] Alsulmi, M. and Al-Shahrani, N. (2022) Machine Learning-Based Decision-Making for Stock Trading: Case Study for Automated Trading in Saudi Stock Exchange. *Scientific Programming*, **2022**, Article ID: 6542862. <https://doi.org/10.1155/2022/6542862>
- [13] Vijh, M., Chandola, D., Tikkiwal, V.A. and Kumar, A. (2020) Stock Closing Price Prediction Using Machine Learning Techniques. *Procedia Computer Science*, **167**, 599-606. <https://doi.org/10.1016/j.procs.2020.03.326>
- [14] Khan, W., Ghazanfar, M.A., Azam, M.A., Karami, A., Alyoubi, K.H. and Alfakeeh, A.S. (2022) Stock Market Prediction Using Machine Learning Classifiers and Social Media, News. *Journal of Ambient Intelligence and Humanized Computing*, **13**, 3433-3456.
- [15] Sun, H. and Yu, B. (2020) Forecasting Financial Returns Volatility: A GARCH-SVR Model. *Computational Economics*, **55**, 451-471. <https://doi.org/10.1007/s10614-019-09896-w>
- [16] (2024) Yahoo Finance. <https://finance.yahoo.com/>
- [17] Monfared, S.A. and Enke, D. (2014) Volatility Forecasting Using a Hybrid GJR-GARCH Neural Network Model. *Procedia Computer Science*, **36**, 246-253. <https://doi.org/10.1016/j.procs.2014.09.087>
- [18] Epaphra, M. (2016) Modeling Exchange Rate Volatility: Application of the GARCH and EGARCH Models. *Journal of Mathematical Finance*, **7**, 121-143. <https://doi.org/10.4236/jmf.2017.71007>
- [19] Karasan, A. (2021) Machine Learning for Financial Risk Management with Python. O'Reilly Media, Inc., Sebastopol.
- [20] Lo, H.C. and Chan, C.Y. (2023) Mean Reverting in Stock Ratings Distribution. *Review of Quantitative Finance and Accounting*, **60**, 1065-1097. <https://doi.org/10.1007/s11156-022-01121-4>