

基于GARCH-BP模型的股票价格波动性研究

严彦文, 王彩云

中国石油大学(北京)理学院, 北京

收稿日期: 2024年12月26日; 录用日期: 2025年2月17日; 发布日期: 2025年2月28日

摘要

股票市场快速发展, 股票价格波动性研究备受关注, 准确预测股价走势对投资者决策和市场稳定意义重大。鉴于股票价格波动的不确定性与非线性特征, 单一模型预测效果欠佳。为此, 本文提出将GARCH与BP神经网络相结合的组合预测方法, 以中国农业银行股票日收盘价数据为例, 基于误差修正思想构建组合模型, 运用BP神经网络对GARCH模型的残差数据进行预测校正。研究结果表明组合模型预测效果优于单一模型, 验证了该组合模型在提高股票价格预测准确度方面的有效性。

关键词

预测, GARCH模型, BP神经网络, 波动性

Research on Stock Price Volatility Based on GARCH-BP Model

Yanwen Yan, Caiyun Wang

College of Science, China University of Petroleum, Beijing

Received: Dec. 26th, 2024; accepted: Feb. 17th, 2025; published: Feb. 28th, 2025

Abstract

With the rapid development of the stock market, the study of stock price volatility has attracted much attention, and accurate prediction of stock price movements is of great significance to investors' decision-making and market stability. In view of the uncertainty and nonlinear characteristics of stock price volatility, the prediction effect of a single model is not good. For this reason, this paper proposes a combined prediction method combining GARCH and BP neural network, taking the daily closing price data of Agricultural Bank of China as an example, constructing a combined model based on the idea of error correction, and utilizing BP neural network to correct the residual data of the GARCH model for prediction. The results show that the combination model predicts better

than a single model, which verifies the effectiveness of the combination model in improving the accuracy of stock price prediction.

Keywords

Prediction, GARCH Model, BP Neural Network, Volatility

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

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

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



Open Access

1. 引言

股票市场在我国金融市场中扮演着关键角色, 对推动我国的经济的发展发挥了至关重要的作用, 其价格波动性研究受到了广泛的关注。股票价格不仅受到宏观经济因素、企业基本面信息的影响, 还受到市场情绪、投资者行为等多种复杂因素的影响[1]。因此, 如何准确地描述和预测股票未来价格的波动趋势, 已成为金融领域中的一项重要研究课题。股票的波动性研究在金融学和投资领域中占据重要地位, 股票价格的波动性不仅直接影响投资者的风险承受能力, 还对金融市场的稳定性和有效性产生深远影响。

目前, 关于股票价格的预测方法主要包括以下三种: 传统投资分析、时间序列分析以及人工智能技术。时间序列分析法主要基于统计学和计量经济学等相关理论, 如移动平均法、趋势外推法等, 然而, 这些方法在处理非平稳、非线性的金融数据时存在一定的局限性。近年来, 随着计算机技术的日新月异, 人工智能和机器学习等方法逐渐被应用到金融领域[2], 为股票价格波动性研究提供了新的视角和工具。本文旨在将机器学习方法引入到时间序列分析中, 对股票价格的波动性进行深入分析研究。广义自回归条件异方差(Generalized Autoregressive Conditional Heteroskedasticity, GARCH)模型作为一种基于时间序列的波动性预测模型, 能够捕捉到股票价格波动性的自相关性和聚集性特征, 但由于对参数选择具有敏感性, 导致预测结果与实际情况存在偏差; 而 BP 神经网络(Back Propagation Neural Network)作为一种典型的机器学习模型, 能够从海量数据中提取有价值的信息, 有效应对时间序列中复杂的非线性问题, 但由于其容易产生过拟合风险, 一定程度上也会影响预测准确率。尽管这两种方法各具优势, 但单独使用其中一种方法往往难以全面准确地描述股票价格的波动性, 因此本文将两种模型相结合, 可以更好地适应股票市场的复杂性和非线性性, 提高股票价格的预测精度和稳定性。

2. 模型方法

2.1. GARCH 模型

Bollerslev [3]于 1986 年扩展了 ARCH 模型, 提出 GARCH 模型, 修正了由于残差序列的异方差函数具有长期自相关性而导致模型阶数过高的问题。GARCH 模型的公式如下:

$$\begin{cases} x_t = f(t, x_{t-1}, x_{t-2}, \dots) + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t} e_t \\ h_t = \omega + \sum_{i=1}^p \eta_i h_{t-i} + \sum_{j=1}^q \lambda_j \varepsilon_{t-j}^2 \end{cases} \quad (1)$$

其中 ε_{t-j}^2 为滞后的残差平方, 即 ARCH 项, h_{t-i} 为 GARCH 项, 这个模型简记为 GARCH(p, q)。GARCH

模型对于参数的选取有两点基本的要求, 一是为了确保条件方差始终为正, 而要求参数非负:

$$\omega > 0, \eta_i \geq 0, \lambda_j \geq 0 \tag{2}$$

二是要求条件方差是平稳的, 这导致了参数有界的约束条件:

$$\sum_{i=1}^p \eta_i + \sum_{j=1}^q \lambda_j < 1 \tag{3}$$

2.2. BP 神经网络模型

BP 神经网络由 Rumelhart 和 McClelland 等人于 1986 年提出, 是一种广泛应用于机器学习和人工智能领域的神经网络模型。它的学习规则是采用动态调整参数的梯度下降方法, 即利用反向传播误差算法使网络能够自动更新权重和偏置, 从而使得误差最小化。

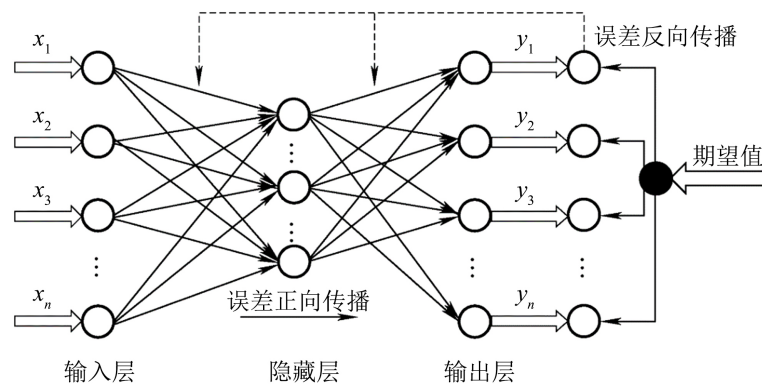


Figure 1. BP neural network structure diagram
图 1. BP 神经网络结构图

BP 神经网络的基本架构如图 1 所示, 它包含了输入层、隐藏层和输出层[4], 网络中的每一层都由若干个神经元构成, 这些神经元之间的连接则通过权重来实现, 从而形成一种多层前馈网络结构。算法的具体步骤如下:

(1) 初始化网络参数。将神经网络的输入数据定义为变量 $X(x_1, x_2, \dots, x_n)$; 设定输入层与隐藏层之间的连接权重为 w_{ij} 、隐藏层与输出层之间的权重为 v_{jk} 、以及隐藏层偏置 b_1 和输出层偏置 b_2 , 将它们分别赋予 $[0, 1]$ 内的数值。

(2) 计算网络隐藏层的输出。根据输入变量序列 X 以及权值 w_{ij} 、偏置 b_1 进行计算, 得到隐藏层的输出 y_j :

$$y_j = f\left(\sum_{k=1}^n w_{kj} x_k + b_1\right) \tag{4}$$

其中, f 表示输入层到隐藏层的激活函数, 取激活函数为 $f(x) = 1/(1 + e^{-x})$ 。

(3) 计算网络输出层的输出。根据得到的隐藏层的输出 y_j 、权值 v_{jk} 和偏置 b_2 , 计算网络结构的预测值 o_k :

$$o_k = f\left(\sum_{j=1}^n v_{jk} y_j + b_2\right) \tag{5}$$

(4) 计算输出层的误差。根据期望达到值和网络预测值, 得出输出层的误差 E :

$$E = \frac{1}{2} \sum_{k=1}^n (d_k - o_k)^2 \quad (6)$$

其中, d_k 表示第 k 个期望值, o_k 为网络的第 k 个输出值。

(5) 将输出层的输出 o_k 代入误差式, 则有:

$$E = \frac{1}{2} \sum_{k=1}^n \left(d_k - g \left(\sum_{j=1}^n v_{jk} y_j + b_2 \right) \right)^2 \quad (7)$$

再将隐藏层的输出 y_j 代入上式, 有:

$$E = \frac{1}{2} \sum_{k=1}^n \left(d_k - g \left(\sum_{j=1}^n v_{jk} f \left(\sum_{i=1}^n w_{ij} x_i + b_1 \right) + b_2 \right) \right)^2 \quad (8)$$

由上式可知, 误差函数 E 是网络各层权值 v_{jk} 、 w_{ij} 与偏置 b_1 、 b_2 的函数, 因此可以使用梯度下降法求解误差函数 E 的极小点:

$$\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}} \quad (9)$$

$$\Delta b_k = -\eta \frac{\partial E}{\partial b_k} \quad (10)$$

权重和偏置的调整应遵循使模型的误差逐步减小的原则, 因此应使调整量与误差的梯度下降成正比, 其中 $\eta \in (0,1)$ 为学习率。

(6) 更新权重和偏置。计算出连接到各层神经元的误差 δ_k , 根据以下公式调整神经网络的连接权值和偏置:

$$w_{jk}(l+1) = w_{jk}(l) + \eta \delta_k o_j \quad (11)$$

$$b_k(l+1) = b_k(l) + \eta \delta_k \quad (12)$$

2.3. 组合预测模型原理

传统的单一模型在拟合股票价格时, 只能捕捉到实际数据的部分信息, 难以全面准确地刻画股票的动态变化, 预测准确度不高[5]。为了实现对股票价格更精确的预测, 本文结合了两种不同预测领域的模型, 即时间序列方法中的 GARCH 模型和机器学习方法中的 BP 神经网络模型, 组合预测的技术路线如图 2 所示, 预测算法的具体步骤如下:

- (1) 获取中国农业银行历史股票价格数据;
- (2) 对数据预处理获得对数收益率序列, 进行平稳性检验和 LM 检验, 进而对 GARCH 模型识别和定阶;
- (3) 建立 GARCH 模型进行预测, 并计算 GARCH 模型的方差;
- (4) 提取拟合好的 GARCH 模型残差序列, 连同股票收益率与 GARCH 模型方差一起作为 BP 神经网络的输入数据, 拟合 BP 神经网络模型;
- (5) 运用 BP 神经网络对 GARCH 模型的残差进行建模预测;
- (6) 将修正好的残差值再带回 GARCH 模型中, 得到组合预测结果。

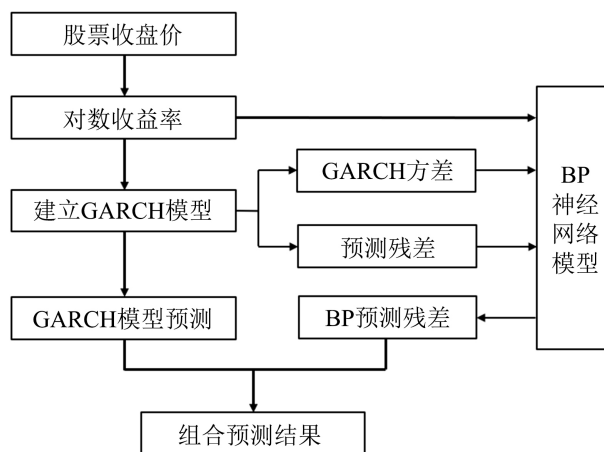


Figure 2. The technical route predicted by the combined model
图 2. 组合模型预测的技术路线

3. GARCH 模型实证分析

3.1. 数据的选取与处理

本文选取中国农业银行从 2021 年 5 月 6 日至 2024 年 1 月 30 日的每日交易收盘价作为研究对象，一共有 669 个股票数据。为了让收益率更符合股票市场中的实际情况，并保持随机变量的二阶矩稳定，对股票日收盘价取对数后再进行一阶差分处理，从而得到对数收益率序列，共有 668 个数据，其计算公式为：

$$R_t = \log(P_t) - \log(P_{t-1}) \tag{13}$$

其中 R_t 为 t 时刻的对数收益率， P_t 、 P_{t-1} 分别表示 t 时刻和 t 前一时刻的股票日收盘价。

3.2. 平稳性检验与 LM 检验

为了确保时间序列分析的准确性、可预测性以及模型建立的有效性，在 GARCH 模型建立前需要进行平稳性检验和异方差性检验。本文采用单位根检验来确定股票收益率序列是否平稳，ADF 检验结果如表 1 所示：

Table 1. Stability test results of return series
表 1. 收益率序列平稳性检验结果

检验项	数值
ADF 值	-13.7397
P 值	<1.1e-25
1%置信水平 ADF 值	-3.4403
5%置信水平 ADF 值	-2.8659
10%置信水平 ADF 值	-2.5691

从单位根检验结果可知，股票收益率的 P 值接近于 0，且 ADF 统计量的值为-13.7397，在 0.1、0.05 和 0.01 等不同显著性水平下，ADF 检验统计量均小于对应的临界值，则认为中国农业银行的股票收益率

序列是平稳的。

由于 GARCH 模型的均值方程仅有常数项, 将中国农业银行的股票收益率数据与其均值作差即可得到残差序列, 对其进行 LM 检验, 结果如表 2 所示:

Table 2. LM test results for residual series
表 2. 残差序列的 LM 检验结果

Null hypothesis: no ARCH effects		
Chi-squared 值	自由度	P 值
59.674	10	4.178e-09

设置 LM 检验的自由度为 10, 从检验结果可以看出, 卡方统计量的值为 59.674, 对应的 P 值接近于 0, 可以拒绝原假设, 说明该残差序列存在 ARCH 效应, 表现出显著的条件异方差性, 因此运用 GARCH 模型进行后续的建模分析与股价预测是合理且可行的。

3.3. GARCH 模型的建立

在实践研究中, GARCH(1,1)模型形式简单且实用性强, 能够满足众多应用场景的需求, 这使得它在金融领域的波动性分析中得到广泛运用, 因此本文建立 GARCH(1,1)模型来预测股票收益率。输出的结果如下表 3 所示:

Table 3. Table of GARCH(1,1) model coefficients
表 3. GARCH(1,1)模型系数表

参数	Estimate	Std. Error	t value	Pr(> t)
μ	0.000179	0.000240	0.74707	0.45502
ω	0.000003	0.000003	1.11675	0.26410
α_1	0.159324	0.049608	3.21166	0.00132
β_1	0.821670	0.045022	18.25060	0.00000

根据输出的 GARCH 模型的参数结果, 可以得到拟合的方程为:

$$r_t = 0.000179 + \varepsilon_t, \varepsilon_t = \sigma_t e_t \tag{14}$$

$$\sigma_t^2 = 3 \times 10^{-6} + 0.159324 \sigma_{t-1}^2 + 0.82167 \varepsilon_{t-1}^2 \tag{15}$$

从表 3 可以看出, GARCH 项的系数 α_1 和 ARCH 项的系数 β_1 对应的 P 值均接近于 0, 说明 GARCH 模型的方差方程中这两部分的拟合结果都是显著的。观察可知, α_1 和 β_1 之和为 0.980994, 小于 1, 说明建立的 GARCH 模型是有效的。

4. GARCH-BP 组合模型实证分析

4.1. 数据来源

组合模型是在拟合 GARCH 模型的基础上建立 BP 神经网络对其残差序列进行修正预测, 所以数据

的选取同 GARCH 模型部分, 依旧以中国农业银行从 2021 年 5 月 6 日至 2024 年 1 月 30 日的每日交易收盘价作为研究对象, 共有 669 个股票数据。将不同时刻的股票价格数据看成时间序列 $\{q(t), t=1, 2, \dots, n\}$, 其预测模型可以描述为:

$$q(t) = \sigma[q(t-1), q(t-2), \dots, q(t-l)] \tag{16}$$

其中, σ 表示非线性函数, 预测窗口的大小 l 表示利用前 l 天的股票数据预测后 1 天的值, 本文预测窗口的大小为 15, 构造样本如下:

$$D = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_i, Y_i), \dots, (X_{653}, Y_{653}), (X_{654}, Y_{654})\} \tag{17}$$

式中, $X_i = \{q_i, q_{i+1}, \dots, q_{i+l-1}\}$, 表示从第 i 天开始的 l 天的历史股票数据集 $Y_i = q_{i+l}$, 表示第 $i+l$ 天的股票数据。训练样本示意图如图 3 所示:

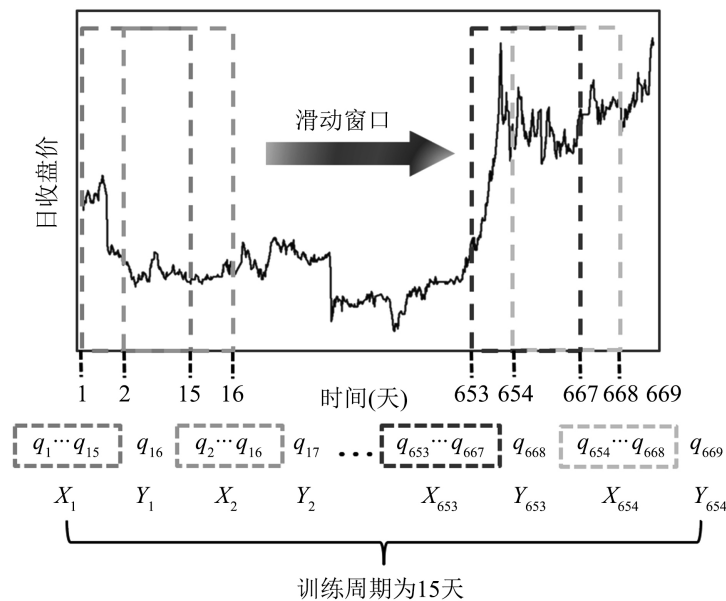


Figure 3. Schematic diagram of training samples
图 3. 训练样本示意图

为了加快训练的速度与预测精度, 在将数据输入 BP 神经网络前一般需要进行归一化处理, 使得不同的数据统一到同一范围内。归一化公式如下:

$$a = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{18}$$

最后, 将预测结果进行反归一化处理, 以获得原始数值。

$$x_i = a(x_{\max} - x_{\min}) + x_{\min} \tag{19}$$

其中 x_i 是初始数据, x_{\min} 、 x_{\max} 分别为序列中的最小值和最大值, a 为经过归一化处理后的数据, 其范围为 $[0,1]$ 。

4.2. 组合模型的建立与预测

基于上述训练样本构建 BP 神经网络模型, 其拟合结果如图 4 所示:

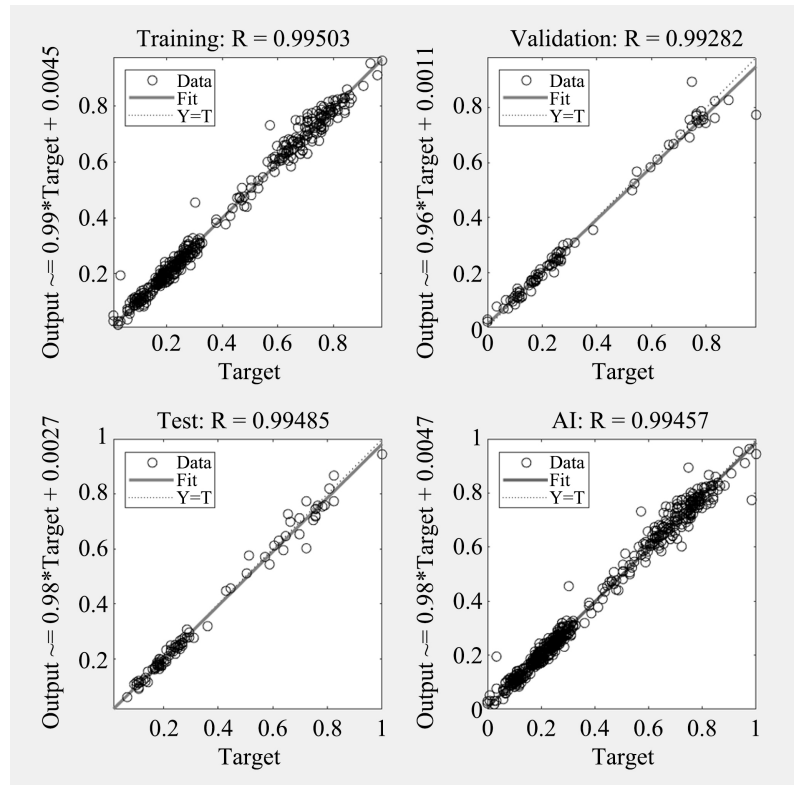


Figure 4. BP neural network model fitting results
图 4. BP 神经网络模型拟合结果

从拟合结果可以看出, 训练好的 BP 神经网络模型在训练集和测试集上的拟合优度均接近于 1, 说明实际数据拟合的很好, 利用 BP 神经网络来预测股票收盘价是可行有效的。

组合模型中, 建立 BP 神经网络模型是为了预测 GARCH 模型的残差序列, 预测的训练周期选为 5 日, 即用连续 5 天的历史数据作为残差的预测依据。每天的历史数据包括 GARCH 模型的残差、GARCH 模型的方差以及对数收益率, 按时间顺序依次将连续 5 天的数据作为 BP 神经网络的输入数据, 其后一天的残差值作为目标输出数据。

根据训练好的 BP 神经网络模型对未来 25 个交易日的残差数据进行预测, 其结果如表 4 所示:

Table 4. Residual value of GARCH model predicted by BP neural network
表 4. BP 神经网络预测的 GARCH 模型残差数值

0.00140290	0.01332771	0.02614103	-0.00270965	-0.00510912
0.00546183	-0.00579877	-0.00561953	-0.00400343	-0.00189212
0.00539392	0.00726650	-0.00028073	0.01152046	-0.00255672
0.00217936	0.01151830	-0.00103313	-0.00032976	-0.00013037
0.01276713	0.00477311	0.00272148	0.00672187	0.00409929

将上述预测残差代入到之前拟合好的 GARCH 模型中可以得到对数收益率序列的预测值, 再将收益率转化为股票收盘价, 即可得到中国农业银行未来 25 个交易日的预测股票价格。

4.3. 预测精度分析

为了分析不同模型关于股票价格波动率的预测精度, 本文选取三种常用的模型性能评估指标对预测值与实际值之间的误差进行比较, 它们分别为 MAE (平均绝对误差)、MSE (均方误差)以及 RMSE (均方根误差), 其运算公式如下:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (20)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (21)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (22)$$

上式中, n 为样本总量, y_i 和 \hat{y}_i 分别表示真实值和预测值。为了直观地判断预测值与真实值之间的差异, 现绘制出单一模型 GARCH 模型、BP 神经网络模型以及两者组合模型的收盘价预测结果对比图, 如图 5~7 所示, 预测精度对比如表 5 所示:

Table 5. Comparison of prediction errors of different models

表 5. 不同模型预测误差对比

方法	MAE	MSE	RMSE
GARCH 模型	0.07218	0.0086404	0.092954
BP 神经网络模型	0.04642	0.0051764	0.071947
组合模型	0.04621	0.0032366	0.056891

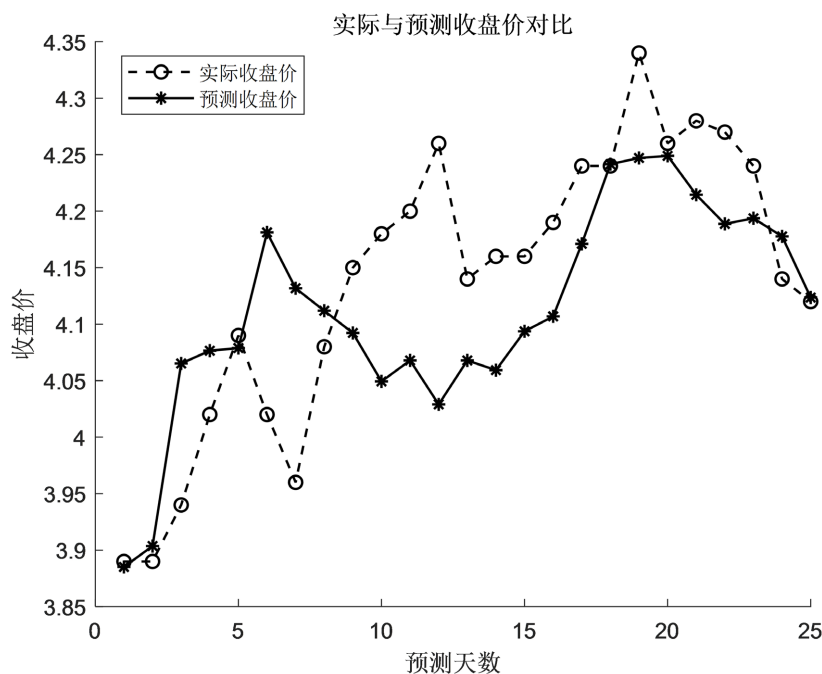


Figure 5. Comparison of actual and predicted closing prices of GARCH model

图 5. GARCH 模型的实际与预测收盘价对比

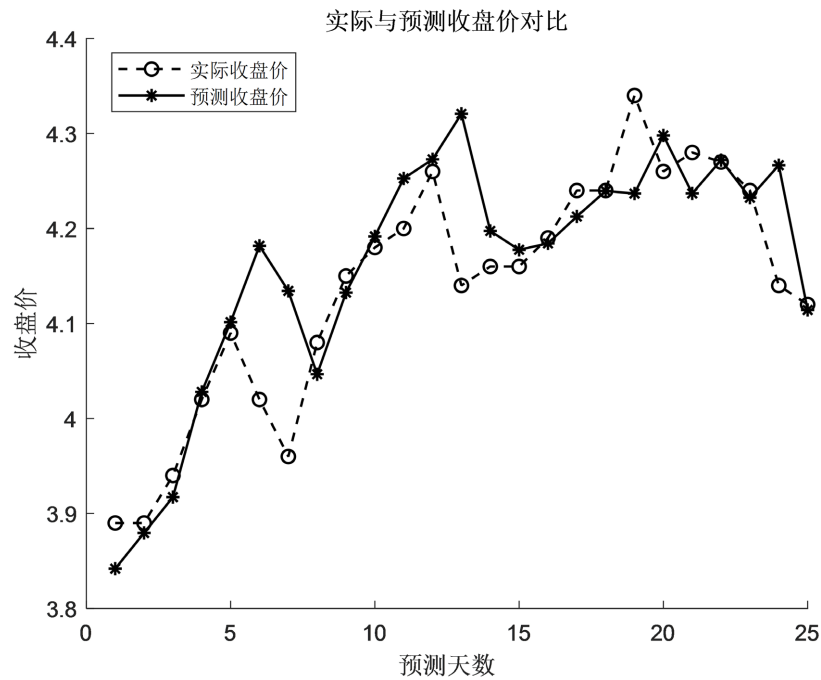


Figure 6. Comparison of actual and predicted closing price of BP neural network model
图 6. BP 神经网络模型的实际与预测收盘价对比

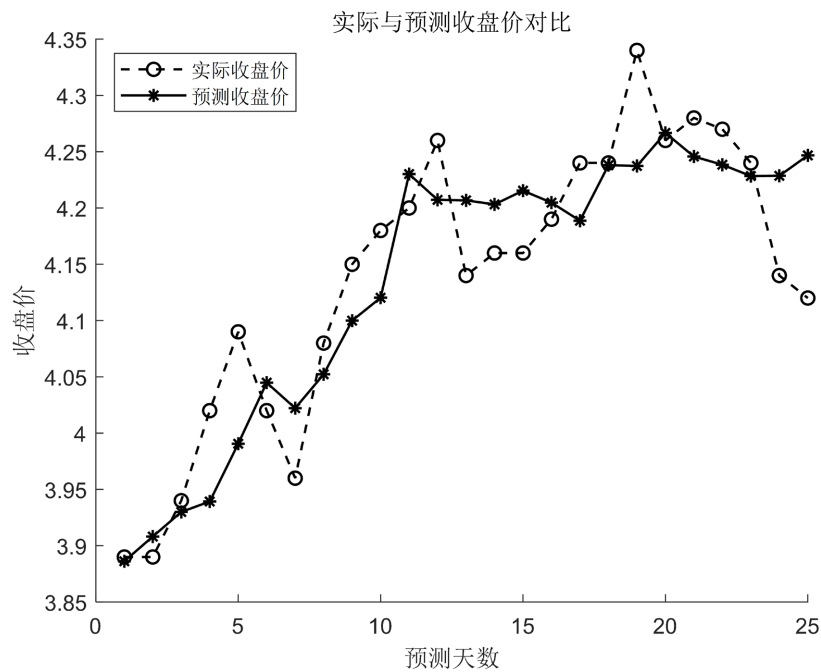


Figure 7. Comparison of actual and predicted closing prices of the combined model
图 7. 组合模型的实际与预测收盘价对比

从以上的预测结果对比图可以观察到, BP 神经网络模型和组合预测模型的实际与预测收盘价之间的偏差较小, 预测的准确度相对较高, 说明这两种模型拟合效果不错, 都很好地描述了中国农业银行未来 25 个交易日股票价格的走势。

组合模型的本质是对 GARCH 模型的残差做出修正处理, 从表 5 的预测误差结果对比可以看出, 引入 BP 神经网络对残差进行预测的组合模型相较于 GARCH 模型的准确度的确有很大的提高, 而且无论选择哪种模型评价指标, 组合预测模型的误差都最小, 预测效果最佳。

5. 结论

针对股票价格变化具有不确定性和非线性的规律, 单一模型很难准确预测未来股票价格的走势, 因此本文将 GARCH 模型和 BP 神经网络模型相结合, 构建了二者的组合预测模型, 并基于实际的股票价格数据通过实证分析, 验证了组合模型的有效性。对于拟合的三种预测模型, 经过三种误差评价指标的比较可知, 组合模型的预测精度最高, 其次是 BP 神经网络模型, 相对较差的是 GARCH(1,1)模型。这是因为 GARCH 模型善于捕捉时间序列数据中的自相关性和波动性, 而 BP 神经网络有很好的非线性学习能力, 通过将这两种模型相结合, 可以充分利用其优势, 提高股票价格的预测精度。

基金项目

中国石油大学(北京)油气资源与探测国家重点实验室“1158”项目(No. PRP/DX-2306)。

参考文献

- [1] 蒋辉, 徐桂烽. 基于灰色系统和 BP 神经网络的股指动态预测[J]. 数学的实践与认识, 2013, 43(22): 28-37.
- [2] 高同, 朱海龙. 基于 GARCH 类模型和 BP 神经网络模型的波动率预测——基于上证综指日度数据[J]. 金融经济, 2020(9): 37-45.
- [3] Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, **31**, 307-327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- [4] 姚培福, 许大丹. BP 神经网络在股票预测中的应用研究[J]. 广东自动化与信息工程, 2006(1): 7-9.
- [5] 曹晓, 孙红兵. 基于灰色 GARCH 模型和 BP 神经网络的股票价格预测[J]. 软件, 2017, 38(11): 126-131.