

基于神经网络和随机波动模型的农产品期货价格预测

钱佳怡

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2025年3月7日; 录用日期: 2025年3月27日; 发布日期: 2025年4月30日

摘 要

本文研究了棉花期货价格的预测问题, 采用了BP神经网络模型和Heston随机波动模型两种模型并进行比较。首先使用BP神经网络模型对样本数据进行训练、验证和测试, 并利用该模型对未来五天的棉花期货价格进行预测。其次, 使用极大似然估计法估计Heston随机波动模型中的参数, 并基于估计后的模型采用滚动时间窗法对样本外五天的棉花期货价格进行预测。最后, 引入五种损失函数对两种方法的预测精度进行比较。实证结果表明, Heston随机波动模型的预测效果优于BP神经网络模型。

关键词

BP神经网络, Heston随机波动模型, 极大似然估计, 滚动时间窗法

Agricultural Futures Price Forecasting Based on Neural Networks and Stochastic Volatility Models

Jiayi Qian

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Mar. 7th, 2025; accepted: Mar. 27th, 2025; published: Apr. 30th, 2025

Abstract

This paper examines the problem of forecasting cotton futures prices, in which we utilize the BP neural network model and the Heston stochastic volatility model, and the prediction accuracy of the two models is compared. Firstly, this paper utilizes the BP neural network model to train, validate and test sample data. And then using this model predict the cotton futures price for five days outside

the sample. After that, we use the maximum likelihood estimation method to estimate the parameters of the Heston stochastic volatility model. Based on the estimated model, the scrolling time window method was used to forecast cotton futures prices five days out of the sample. Finally, introducing five loss functions to compare the prediction accuracy of the two methods, the empirical results show that the Heston stochastic volatility model is better than the BP neural network model in prediction.

Keywords

BP Neural Network, Heston Stochastic Volatility Model, Maximum Likelihood Estimation, Scroll Time Window Method

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

棉花作为核心纺织原料,其期货价格波动直接影响全球产销链条运行。当前电商平台重塑交易模式,加速价格发现进程,但同步加剧市场波动风险。研究期货价格形成机制,对优化产业风险管理、提升市场运行效率具有双重价值,既为生产者提供套期保值依据,亦助力构建数字化时代的价格预警体系。

棉花期货价格的研究旨在寻找棉花价格波动规律,帮助市场参与者更好地把握市场变化趋势,制定更为准确的决策。其主要目的与意义有:预测未来棉花价格走势。通过对棉花期货价格的研究,分析各种市场的供需情况,对未来市场走势的预测更加准确。

2. 文献综述

棉花期货价格是由多种因素影响的,如产量、进口量、市场需求、政策等因素。而江知航(2021) [1]在现有的基础上增加气候因素对价格的影响并应用双向长期记忆网络(BTLSTM)模型与长短期记忆(LSTM)模型对棉花价格的波动进行预测。研究结果表明:BTLSTM的预测效果优于LSTM,拟合测试集的误差最小,且价格预测精度较高。通过SWA算法优化LSTM和BTLSTM网络,可以实现快速收敛和较强的泛化性能。该模型可以更准确地表达棉花市场价格波动规律。此外,张兆同、余潜(2017) [2]也是从影响棉花价格波动的多个因素进行分析。研究者采用灰色关联分析法从众多影响因素中选出了4个影响最大的因素:国际市场因素、替代品因素、居民消费价格指数和棉花进口量。并且通过主要的影响因素去优化了RBF神经网络模型的输入节点。该研究分别从数据中选取训练集和测试集,而经过训练后的网络拟合效果很好。结果表明:预测精度较高,泛化能力强,可以较好地捕捉棉花价格变化的本质规律,为准确预测棉花市场价格提供参考。

然而,现实中,棉花期货价格受多因素影响,全面考虑这些因素预测难度大且数据可能不全或不准。因此,学者多采用基于时间序列的方法来预测棉花期货价格。曹建飞(2021) [3]针对2013年9月16日至2019年5月13日中国棉花价格,提出多尺度组合预测模型。采用EEMD分解价格序列,聚类分析重构分量,解释波动特征。比较多种预测方法后,选择最优模型预测重构序列,集成预测结果。该组合模型较单一及其他组合模型预测精度更高。袁新晨(2021) [4]通过分析纺织原料价格指数构成,采集棉花价格指数历史数据,并用马尔科夫链建模及VB编程设计预测系统,预测2020年棉花价格的月度和日度走势。研究表明,马尔科夫链预测的棉花价格区间符合市场数据变动,短期稳定的棉花价格指数更适合用此方

法预测。高欣宇、余国新(2014) [5]以 2013 年 1 月 2 日至 2014 年 6 月 31 日期间的棉花期货价格作为研究对象,使用 ARIMA 模型和 EGARCH-EWMA 模型进行短期价格预测,并将两者进行对比分析。结果表明,EGARCH-EWMA 模型在准确度和可行性方面优于 ARIMA 模型。该研究采用 EGARCH 模型估计滞后系数,以确定衰减因子,克服了无法科学地判定衰退因子的不足。张立杰、寇纪淞、李敏强、朱新杰(2013) [6]分析了通过 2008 年至 2011 年间的月度棉花价格数据建立了基于自回归移动平均的 ARIMA (1, 1, 1)模型。结果表明,ARIMA (1, 1, 1)模型可以很好地模拟国内棉花价格且平均相对误差百分比低于 4%。基于此,研究者还建立了支持向量机模型来处理 ARIMA 模型中的残差,并将自回归移动平均模型与 SVM 模型进行组合预测,并将单一预测模型与组合预测模型进行比较。结果表明,组合预测模型比单一预测模型具有更好的预测精度和改进作用。

综上所述,现有研究在模型精度、变量覆盖、动态适应性等方面取得进展,但仍需解决数据质量、多因素交互、模型鲁棒性及可解释性等核心问题。而神经网络模型以处理非线性数据及自动降噪能力擅长预测复杂市场,Heston 模型则通过显式波动率建模和明确经济参数提供有效预测。因此,本文使用两种模型对棉花期货价格进行预测并对比两种模型。

3. 神经网络

3.1. 数据选取说明

本文的棉花价格主要以郑州商品交易所的每日棉花收盘价作为研究对象,选取了从 2022 年 1 月 4 日至 2023 年 4 月 17 日共计 310 天的数据作为实证研究的数据基础。预测过程分别以 2022 年 8 月 3 日至 2023 年 3 月 31 日连续 160 天、2022 年 6 月 15 日至 2023 年 3 月 31 日连续 210 天、2022 年 3 月 1 日至 2023 年 3 月 31 日连续 260 天、2022 年 1 月 4 日至 2023 年 3 月 31 日连续 300 天的价格数据作为训练样本,以 2023 年 4 月 3 日至 2023 年 4 月 10 日连续五天的价格数据作为测试样本,2023 年 4 月 11 日至 2023 年 4 月 17 日连续五天的价格数据作为预测样本。因为期货交易具有休息日且在此期间不进行交易,所以训练和预测样本选择五天的数据,即分别以 $H = 165$ 、215、265、305 天的历史数据预测未来五天棉花期货的价格。历史价格数据选取情况如表 1。

Table 1. Sampling interval

表 1. 样本选取区间

训练样本	测试样本	预测样本
$H = 165$ 天		
2022 年 8 月 3 日至 2023 年 3 月 31 日	2023 年 4 月 3 日至 2023 年 4 月 10 日	
$H = 215$ 天		
2022 年 6 月 15 日至 2023 年 3 月 31 日	2023 年 4 月 3 日至 2023 年 4 月 10 日	2023 年 4 月 11 日
$H = 265$ 天		
2022 年 3 月 1 日至 2023 年 3 月 31 日	2023 年 4 月 3 日至 2023 年 4 月 10 日	至 2023 年 4 月 17 日
$H = 305$ 天		
2022 年 1 月 4 日至 2023 年 3 月 31 日	2023 年 4 月 3 日至 2023 年 4 月 10 日	

3.2. 数据进行预处理

因为棉花价格的数据噪声较大,所以在构建模型前对样本数据进行归一化处理,其次由于数据是非

线性的，所以本文的函数选择的是 S 型曲线函数且需要将数据输入值控制在[0, 1]内。此外，为提高 BP 神经网络的学习率，本文采用极差法对数据进行标准化处理。

3.3. 预测模型的选择

从数据来源看，BP 神经网络输入分为静态数据和时间序列动态数据。BP 网络能学习历史数据规律，进行静态及时间序列预测。但静态指标预测棉花价格因素复杂、收集难度大，而时间序列预测模型则综合考虑多种因素通过均衡价格体现。因此，本文采用基于历史数据的模型预测棉花价格。

3.4. 设计 BP 神经网络模型

从数据来源看，BP 神经网络输入分为静态数据和时间序列动态数据。BP 网络能学习历史数据规律，进行静态及时间序列预测。但静态指标预测棉花价格因素复杂、收集难度大，而时间序列预测模型则综合考虑多种因素通过均衡价格体现。因此，本文采用基于历史数据的模型预测棉花价格。

(1) 网络层数的选定

据理论研究，三层 BP 神经网络能拟合任意函数。因此，本文选用含一个隐含层的三层 BP 神经网络进行预测分析，以满足效果并降低成本和复杂度。

(2) 输入输出层神经节点

根据棉花价格的数据特征，输入层的神经节点个数设定为 5，同时，输出层的神经节点个数设定为 1，也就是说，用连续五天的棉花价格数据来预测第六天的价格。

(3) 隐藏层神经节点

因无确定隐含层神经元个数的成熟理论，本模型依据以往经验估算。通过改变隐含层神经节点数实验，选择输出误差最小的节点数为最优。本文采用的经验公式是： $l = \sqrt{n + m} + p$ ，其中 l 为隐含层神经节点的个数， n 、 m 分别为输入层和输出层的神经节点个数， p 为大于 1 且小于 10 的常数。将数据进行标准化处理后，把不同区间的测试集和训练集代入模型，最终得到每个节点的均方根误差和，最后将隐含层神经节点设定为 8。各个隐藏节点的均方根误差如表 2。

Table 2. Root mean squared error (RMSE) of each hidden node across different intervals
表 2. 不同区间各个隐藏节点的均方根误差

	$H = 165$	$H = 215$	$H = 265$	$H = 305$	合计
3	0.0040673	0.00097869	0.00062239	0.0006023	0.00627
4	0.0032983	0.0008674	0.00065206	0.0005861	0.0054
5	0.0029321	0.0010907	0.0005864	0.0006332	0.00524
6	0.0031548	0.0007521	0.00064447	0.0005143	0.00507
7	0.0033369	0.0010205	0.00063817	0.0005165	0.00551
8	0.0030465	0.00064195	0.00063423	0.0006041	0.00493
9	0.0039482	0.00076512	0.00068292	0.0005396	0.00594
10	0.003698	0.0012076	0.00064793	0.0005689	0.00612
11	0.0041189	0.0006959	0.0004779	0.0006744	0.00597
12	0.0031878	0.00089116	0.00056534	0.0006053	0.00525

(4) 算法及参数设定

构建模型时，用 trainlm 函数训练神经网络，因其能快速收敛至全局最优。设学习效率为 0.01，目标

误差为 0.000001, 最大迭代 1000 次。为提升拟合度, 设两停止条件: 当目标误差或最大迭代次数时停训, 确保神经网络高效达到最佳拟合。

3.5. 结果展示

将 2022 年 8 月 3 日至 2023 年 4 月 10 日即 $H = 165$ 天的样本数据值代入模型中进行训练, 训练结果如图 1。

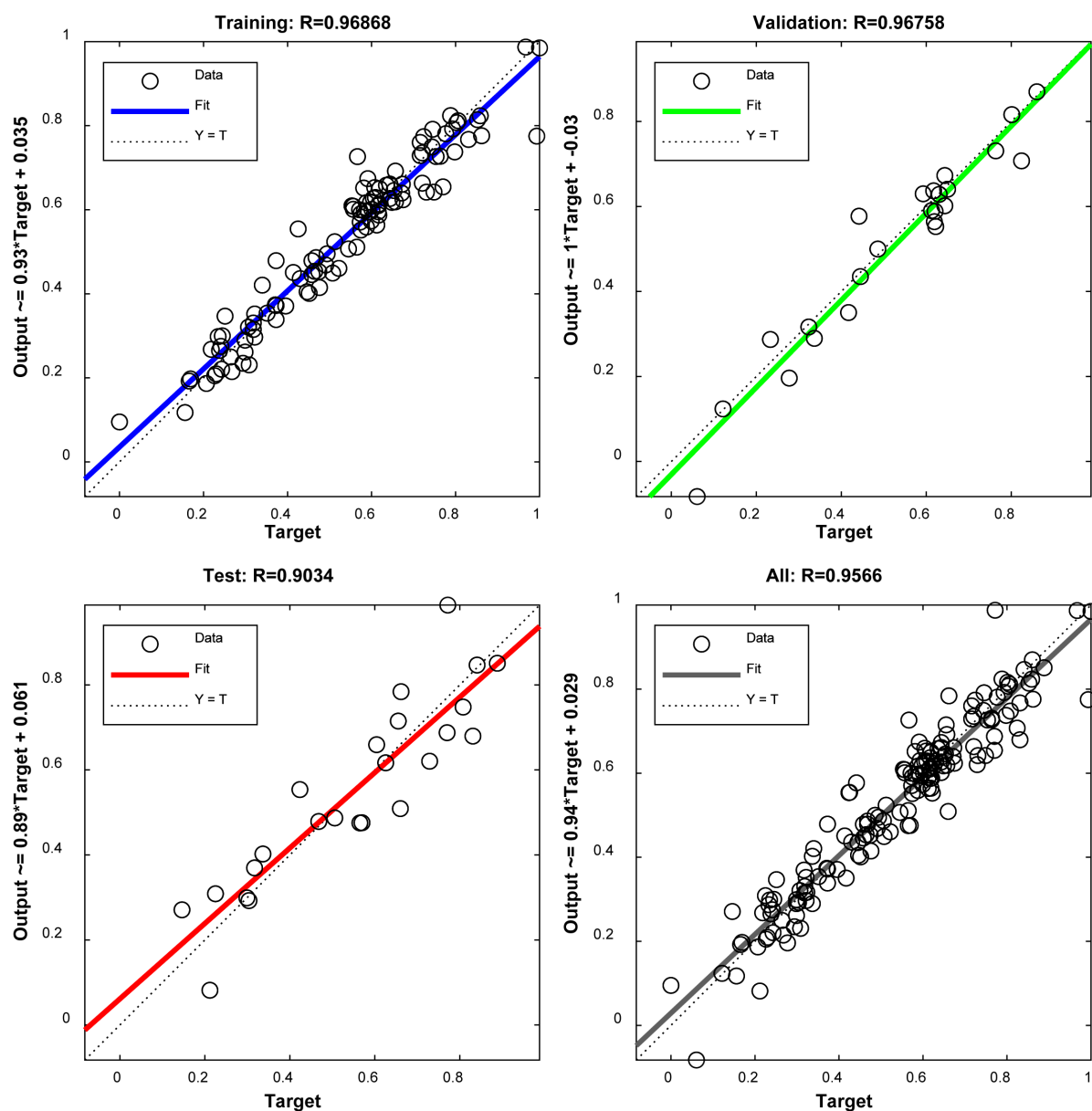


Figure 1. Training results for the configuration with $H = 165$

图 1. $H = 165$ 的训练结果

对于训练结果, 大部分数据点都在拟合线附近分布, 说明模型对于大部分数据点的拟合效果比较好。从曲线拟合的角度看, 模型的回归系数为 0.96868, 表明模型的训练效果较好。在训练过程中, 模型对目

标输出曲线和输出曲线之间的偏差进行了较好的调整, 因此总体来看, 模型拟合效果比较准确。然而, 需要注意的是, 尽管模型训练效果较好, 但未来预测中仍有可能出现个别值与真实数据相差较大的情况。

在对 BP 神经网络模型进行检验时, 获得了该模型测试集的真实数据与预测数据的比较图。如图 2。

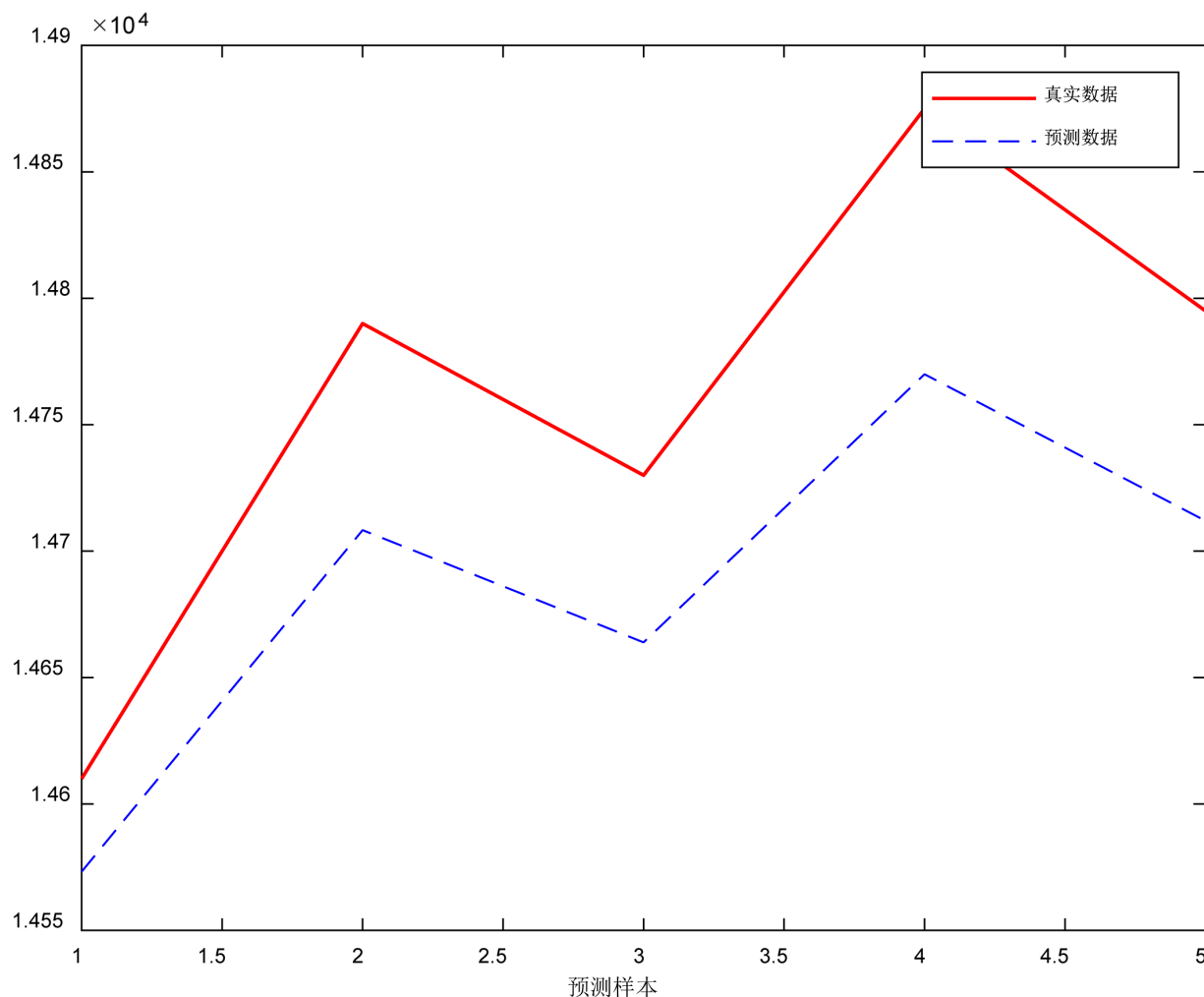


Figure 2. Comparison plot of actual vs. predicted data for the test set

图 2. 测试集的真实数据与预测数据的对比图

将 2023 年 4 月 3 日至 2023 年 4 月 10 日的棉花价格代入模型中进行测试得到图 3, 可以看出, 预测数据与真实数据之间的距离较为接近并且两者之间的波动一致, 其中最为接近的值是第一组。通过计算可知, 测试集的预测数据与真实数据的均方根误差(RMSE)为 128.0165。

选取 2023 年 4 月 11 日至 17 日棉花期货价格, 用训练好的 BP 模型预测这五天价格。预测数据稍有滞后, 但偏差小, RMSE 为 77.8799, MAE 为 74.5348。偏差可能源于预测数据的累积误差和期货价格的大范围波动。

本节再将 2022 年 8 月 3 日至 2023 年 3 月 31 日连续 160 天、2022 年 6 月 15 日至 2023 年 3 月 31 日连续 210 天、2022 年 3 月 1 日至 2023 年 3 月 31 日连续 260 天、2022 年 1 月 4 日至 2023 年 3 月 31 日连续 300 天的价格数据作为训练样本进行以上操作, 得到表 3, 表 4, 并对两个表格进行分析。

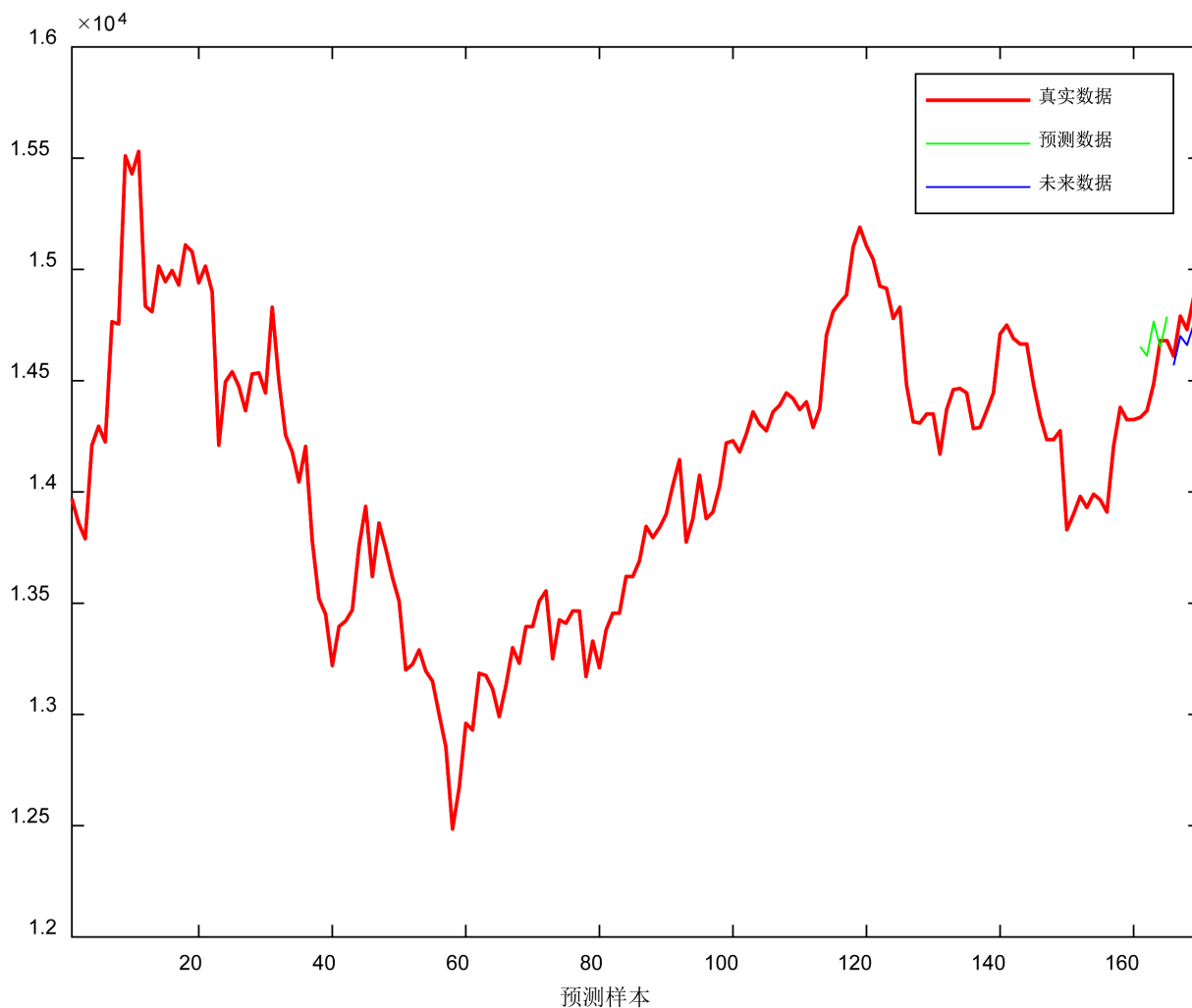


Figure 3. Neural network model prediction of out-of-sample price movement trends

图 3. 神经网络模型预测样本外价格变化趋势

Table 3. Comparison of prediction performance between actual and forecasted data across different time intervals

表 3. 不同区间对测试集预测数据与真实数据的效果

区间	RMSE	R ²	MAE	MBE
$H = 165$	128.0165	-1.1146	117.4096	0.4405
$H = 215$	123.9175	-0.98136	90.9104	-69.9559
$H = 265$	129.9129	-1.1777	102.826	-87.7126
$H = 305$	105.8081	-0.44456	91.691	-23.7338

从表 3 可以发现, 当 $H = 165$ 、215、305 时, 随着训练样本区间的增长, 测试集中的预测数据与真实数据之间的均方根误差(RMSE)逐渐减小, 这表明训练效果良好, 训练集的拟合效果随着样本增加拟合效果逐渐变好。

表 4 显示, $H = 165$ 、215、305 时, 训练样本区间增长使预测更准确。但 $H = 265$ 时, 样本外预测误差最大。说明增加训练样本不一定能提高预测效果, 需适当选择和调整模型。

Table 4. The out-of-sample prediction performance across different intervals
表 4. 不同区间的样本外预测效果

区间	RMSE	MAE	MSE	RMSPE	MAPE
$H = 165$	77.8799	74.5348	6065.2757	0.0059	0.00004
$H = 215$	76.2843	72.9021	5819.2888	0.0058	0.00004
$H = 265$	93.7803	90.1642	8794.7411	0.0071	0.0001
$H = 305$	47.2724	40.5376	2234.6777	0.0037	0.00004

3.6. 小结

通过上述的损失函数可知神经网络模型的预测效果不太理想。虽然当 $H = 165$ 、 215 、 305 时，样本的区间增大使得样本外的预测误差在逐渐减小，但是当 $H = 265$ 时，测试集的预测数据与真实数据之间的误差和样本外的预测误差均达到最大值。因此，在实际应用中，需要根据具体情况选择合适的样本数量，以实现有效的风险管理和投资决策。

4. Heston 随机波动模型

4.1. Heston 随机波动模型简介

Heston 随机波动模型是 Steven L. Heston 于 1993 年提出的金融风险管理工具。它采用两个随机过程(股票价格和波动率)来描述金融市场特性,适用于期权定价和风险管理。模型认为股票价格和波动率均随机且相互关联,能更好反映资产价格的真实特征,包括偏度、尖峰厚尾等,有助于管理金融风险。

Heston 模型是一个比较简单并且非常有效的随机波动模型,其模型表达式如下:

$$\begin{cases} \frac{dS(t)}{S(t)} = \mu dt + \sigma(t) dW_s(t) \\ d\sigma^2(t) = (\alpha + \beta\sigma^2(t))dt + \sigma_0\sigma(t) dW_\sigma(t) \end{cases} \quad (1)$$

表达式中, $W_s(t)$, $W_\sigma(t)$ 是相关系数 ρ 的两个布朗运动。模型共有五个参数,分别为 μ , α , β , σ_0 , ρ , 有时候也会将 μ 取为无风险利率 r 。

对该模型进行变形,可以得到以下的表达式:

$$\begin{cases} dS(t)/S(t) = \mu dt + \sigma(t) dW_1(t) \\ d\sigma^2(t) = (\alpha + \beta\sigma^2(t))dt + \sigma_1\sigma(t) dW_1(t) + \sigma_2\sigma(t) dW_2(t) \end{cases} \quad (2)$$

式中, $W_1(t)$, $W_2(t)$ 是两个相对独立的布朗运动。经过变形后,模型的参数变为 μ , α , β , σ_1 , σ_2 , 先前的 $\rho = \frac{\sigma_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}$ 。变形后的 Heston 模型便于数据模拟和参数估计。本文将用极大似然估计法来估计其参数,并需解析后验概率密度函数和极大似然估计的表达式。

4.2. 棉花价格的 Heston 预测模型构建与实证分析

Heston 模型所采用的数据是来自郑州商品交易所的棉花期货价格数据,采用了从 2022 年 1 月 4 日到 2023 年 4 月 10 日的日收盘价。用简单收益率通过日收盘价计算,而已实现波动率则通过五分钟的棉花价格数据来计算,从而更好地通过极大似然方法对模型参数进行估计及预测。

4.2.1. 模型处理

通过之前描述的 Heston 模型, 要将其进行欧拉格式的离散化处理, 并写出其表达式:

$$\begin{cases} S(t_{n+1}) - S(t_n) = \mu S(t_n) \Delta t_n + \sigma(t_n) S(t_n) \Delta W_1(t_n) \\ \sigma^2(t_{n+1}) - \sigma^2(t_n) = (\alpha + \beta \sigma^2(t_n)) \Delta t_n + \sigma_1 \sigma(t_n) \Delta W_1(t_n) + \sigma_2 \sigma(t_n) \Delta W_2(t_n) \end{cases} \quad (3)$$

其中, $n = 0, \dots, M-1, M$ 为样本量, $\Delta W_1(t_n) = \Delta W_1(t_{n+1}) - \Delta W_1(t_n), i = 1, 2$ 。因为通常情况下一年有 252 个交易日, 所以本节选取 $\Delta t_n = \Delta t_{n+1} - t_n = \frac{1}{252}$ 。

在本文中用 $(y_1, y_2, \dots, y_n)$ 表示一天内棉花期货价格的简单收益率序列, $\sigma^2(t_n)$ 表示一天内棉花价格的已实现波动率, 而第 t 天已实现波动率是通过第 t 天的高频收益率来计算的, 主要是将第 t 天的高频收益相加得到的总和, 它的计算公式表达式为:

$$RV_t = \sigma^2(t_n) = \sum R_t^2 = \sum \left(\ln \frac{y_t}{y_{t+1}} \right)^2 \quad (4)$$

4.2.2. 参数极大似然估计

研究人员张磊[7]对 Heston 模型进行研究, 从而推导出了模型参数的后验密度函数和极大似然估计的解析表达式。这个过程详细说明了参数估计的理论推导过程。

通过张磊[7]的求估模型参数的详细推导, 本文各参数的极大似然估计为下文所示。

4.2.3. 样本外的预测及预测结果分析

首先用训练样本估计 Heston 模型参数, 再用测试样本验证预测效果。采用滚动时间窗法预测样本外值, 并引入损失函数评估模型预测效果。

一、样本外预测

在本文的样本外预测过程中采用了滚动时间窗方法。将数据分为两个不同的部分, 一部分用于估计模型参数的训练样本, 另一部分则用于预测的测试样本。首先使用训练样本来估计模型中的参数, 在此基础上利用估计所得的模型参数对未来进行预测。具体过程如下:

1、数据的两个部分分别为训练样本和测试样本。为了进行估计, 本节选取了四个不同长度的训练样本, 分别为 165 天、215 天、265 天和 305 天, 而 2023 年 4 月 11 日至 2023 年 4 月 17 日的 5 个交易日的收盘价作为本文的测试样本($t = H+1, H+2, \dots, H+M$, 其中 $M = 5$, H 为估计样本区间的的天数)。

2、利用包含 165 个数据的训练样本(即 $t = 1, 2, \dots, H = 165$)对模型的参数进行估计。基于的给定待估参数值, 利用模型计算出第 166 天的价格预测值。为了提高预测的准确性, 我们进行了 165 次的估计和预测, 得到了 165 个预测价格。我们将这 165 个预测价格的平均值作为次日的预测价格, 记为 P_{H+1} , 也就是说, 我们在基于前 165 个样本数据计算出的第 166 天预测价格。

3、对于长度为 165 天的训练样本, 采取向后移动一天的方法, 仍然使用从 $t = 2$ 到 $t = 166$ 的数据对模型进行参数估计。在得到新的参数估计值后, 利用模型计算出第 167 天的价格预测值。接着重复进行 165 次估计和预测的过程, 每次估计和预测使用的是从 $t = 3$ 到 $t = 167$, $t = 4$ 到 $t = 168$ 等等, 长度为 165 的数据样本。最后, 利用这些预测值的平均数, 计算出第 167 天、第 168 天等等直到第 $H+M$ 天的预测价格。这样就可以利用 Heston 随机波动模型对未来棉花价格进行预测, 并得到最后一天的预测结果 P_{H+M} 。

4、在保持 2023 年 4 月 11 日至 2023 年 4 月 17 日的 5 个交易日收盘价作为测试样本并不变的情况下, 重复进行第二步和第三步的操作, 分别使用训练样本区间为 $H = 165$ 天、215 天、265 天和 305 天的样本数据, 以得到针对四个不同估计区间的共计 20 个棉花期货价格预测值。用 P_m 来代表测样本区间的

真实价格，其中 $m = H + 1, H + 2, \dots, H + M$ 。

经过以上步骤的计算得到以下结果。

二、预测结果分析

用不同区间的训练样本对模型参数进行估计，表 5 是不同区间的参数估计值。

Table 5. The parameter estimates for different intervals

表 5. 不同区间的参数估计值

参数	$H = 165$	$H = 215$	$H = 265$	$H = 305$
$\hat{\mu}$	-0.407078	-0.4331973	-0.265873	-0.2583617
$\hat{\alpha}$	0.006934277	0.1615073	0.00742805	0.008280862
$\hat{\beta}$	1.107415e-06	1.141309e-06	1.160037e-06	1.297781e-06
$\hat{\sigma}_1$	0.0004315474	0.0002074812	0.0002395282	0.0001731509
$\hat{\sigma}_2$	0.1227431	0.007493404	0.1825468	0.2032789

通过上述的方法得到区间 $H = 305$ 时的真实数据与预测数据的对比图，如图 4。

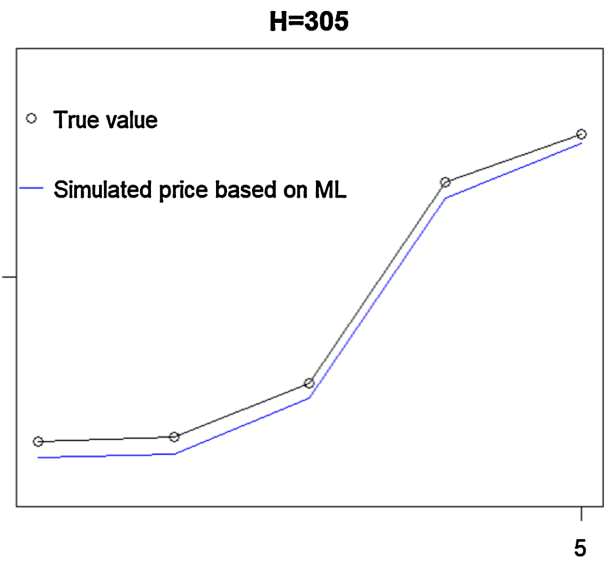


Figure 4. A comparison chart of real data versus predicted data when $H = 305$
图 4. $H = 305$ 时的真实数据与预测数据的对比图

图 4 给出了在极大似然估计下的 Heston 模型的真实数据与预测数据的对比图，从图中可以发现，真实数据与预测数据的波动趋势一致，并且通过计算可知均方根误差为 13.85854，说明基于极大似然估计的 Heston 模型能够有效地对棉花期货价格进行预测。

本研究引入了五个不同的损失函数，用于评估该估计方法对预测结果的影响。基于五个损失函数得到 Heston 模型中预测数据与真实数据之间的误差，并将各个损失函数值展示在表 6 中。

由表 6 中的五种损失函数可以发现，随着样本区间的增大，五种损失函数的值均在减小。由此可以说明，采用极大似然估计方法下的 Heston 随机波动模型可以有效地对棉花期货价格进行预测。同时，这也说明使用极大似然估计方法对 Heston 随机波动模型中的参数进行估计具有较好的精度和可靠性。这为

运用 Heston 模型进行期货价格预测提供了可行性和依据。

Table 6. The out-of-sample prediction performance across different estimation intervals
表 6. 不同估计区间的样本外预测效果

区间	MSE	MAE	RMSPE	RMSE	MAPE
$H = 165$	697.0373	25.9944	0.001812167	26.40146	3.283949e-06
$H = 215$	748.7916	27.18391	0.002043173	27.36406	4.174555e-06
$H = 265$	355.0031	17.03247	0.001266533	18.84153	1.604107e-06
$H = 305$	192.0591	13.5843	0.0009608442	13.85854	9.232215e-07

5. 结果对比

5.1. 比较两种模型

将两种模型的损失函数展示在表 7 中。

Table 7. Comparison of loss functions
表 7. 损失函数对比

区间	损失函数	BP	ML
$H = 165$	MSE	6065.28	697.0373
	MAE	74.5348	25.9944
	RMSPE	0.0059	0.001812167
	RMSE	77.8799	26.40146
	MAPE	0.00004	3.28E-06
$H = 215$	MSE	5819.29	748.7916
	MAE	72.9021	27.18391
	RMSPE	0.0058	0.002043173
	RMSE	76.2843	27.36406
	MAPE	0.00004	4.17E-06
$H = 265$	MSE	8794.74	355.0031
	MAE	90.1642	17.03247
	RMSPE	0.0071	0.001266533
	RMSE	93.7803	18.84153
	MAPE	0.0001	1.60E-06
$H = 305$	MSE	2234.68	192.0591
	MAE	40.5376	13.5843
	RMSPE	0.0037	0.000960844
	RMSE	47.2724	13.85854
	MAPE	0.00004	9.23E-07

从表 7 可以发现, 当 $H=165$ 、215、265、305 时, 在 MSE、MAE、RMSPE 等五个指标上 Heston 模型的预测效果更好。此外, Heston 随机波动模型的误差随着样本区间的增大都在不同程度地减小, 然而 BP 神经网络模型的误差在 $H=265$ 时反而增大, 由此可以看出, 与 BP 神经网络相比, Heston 模型在预测棉花期货价格上更具有优势。

5.2. 两种模型的差异机制

针对棉花期货数据噪声高、波动复杂的特点, 神经网络模型凭借非线性处理能力和自适应降噪优势更适配数据驱动场景; 然而其'黑箱'特性与过拟合风险导致预测效果较 Heston 模型存在劣势。相比之下, Heston 模型通过显式波动率建模提供清晰的参数经济含义, 在复杂数据环境下表现出更强的预测稳定性。

6. 结论、建议和展望

6.1. 结论

本文主要应用了两种模型对棉花期货价格进行了预测, 并引入五个不同的损失函数对预测结果进行了评价。结果显示, BP 神经网络模型和 Heston 随机波动模型都可以用于棉花期货价格的预测, 但是相比较而言, Heston 随机波动模型的预测效果更好。由此可见, 对于金融市场中价格预测问题, 尤其是在涉及到连续性、波动性等复杂数据特征的情况下, Heston 随机波动模型是一种有效的建模方法。

6.2. 建议与展望

虽然本文通过两种模型实现了棉花期货价格的预测, 但是未来还有很大的进步空间并且还有很多更深层次的问题值得去深思。比如, 在本文中只是单纯的以随机波动的角度来预测棉花的期货价格, 但是本文并没有更深入地去分析造成价格波动的原因, 因此, 在未来的工作中可以将导致期货价格波动的各种因素加入到模型的预测之中; 其次本文在两种模型中只考察了 4 种估计区间对棉花期货价格预测的效果, 所以未来可以考虑区间更长的预测效果。

最后, 通过对 BP 神经网络模型和 Heston 随机波动模型进行价格预测效果比较发现基于极大似然估计的 Heston 模型在价格预测方面有一定的优势。因此, 生产者、投资者和纺织业经营者等可以尝试将基于极大似然估计法的 Heston 模型应用到其他期货品种的价格预测中, 以推动该模型在期货品种价格预测方面的发展, 这样合理地应用预测模型能够比较好的发挥预测优势。

参考文献

- [1] 江知航, 王艳霞, 颜家均, 等. 基于 BILSTM 的棉花价格预测建模与分析[J]. 中国农机化学报, 2021, 42(8): 151-160.
- [2] 张兆同, 余潜. 灰色关联分析与 RBF 神经网络在我国棉花价格预测中的应用研究[J]. 价格月刊, 2017(9): 31-36.
- [3] 曹建飞. 基于多尺度组合模型的我国棉花价格预测[J]. 统计与决策, 2021, 37(2): 176-180.
- [4] 袁新晨. 基于马尔科夫链的棉花价格波动预测与企业成本管理分析[D]: [硕士学位论文]. 苏州: 苏州大学, 2020.
- [5] 高欣宇, 余国新. 对我国棉花期货价格预测的方法研究——基于 EGARCH-EWMA 模型与 ARIMA 模型比较[J]. 价格理论与实践, 2014(12): 85-87.
- [6] 张立杰, 寇纪淞, 李敏强, 等. 基于自回归移动平均及支持向量机的中国棉花价格预测[J]. 统计与决策, 2013(6): 30-33.
- [7] 张磊. 随机波动率模型参数估计: 贝叶斯和极大似然方法[D]: [硕士学位论文]. 北京: 清华大学, 2013.