

基于ARIMA模型的螺纹钢价格预测分析

安江鸿¹, 李宜阳¹, 成常杰²

¹上海工程技术大学数理与统计学院, 上海

²欧冶云商股份有限公司, 上海

收稿日期: 2023年3月31日; 录用日期: 2023年5月24日; 发布日期: 2023年5月31日

摘要

螺纹钢在房屋、桥梁、道路等土建工程建设上应用广泛。螺纹钢的价格波动会给交易双方带来较大的风险。螺纹钢期货在一定程度上可以规避价格波动带来的风险。螺纹钢期货价格的预测也成为当下的研究热点。本文在对最近六年的螺纹钢期货价格进行分析的基础上对螺纹钢期货价格进行预测。本文采用ARIMA模型, 将螺纹钢期货价格的历史数据进行可视化, 通过时间序列历史值及历史值预测的误差确定的模型对螺纹钢期货价格进行预测。本文所建模型在所选取的时间段内, 对螺纹钢期货价格的预测值与真实值相比, 平均误差为0.03%, 能够很好地刻画期货价格的走势与波动。

关键词

螺纹钢期货, ARIMA模型, 价格预测

Price Prediction Analysis of Rebar Based on ARIMA Model

Jianghong An¹, Yiyang Li¹, Changjie Cheng²

¹School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai

²Ouyeel Co., Ltd., Shanghai

Received: Mar. 31st, 2023; accepted: May 24th, 2023; published: May 31st, 2023

Abstract

Steel rebar is widely used in the construction industry. With the change of market demand, the opacity of the trading market brings great risks to the price of steel rebar. Steel rebar futures can avoid the above risks to a certain extent. The prediction of steel bar futures price has become a hot research topic. Based on the analysis of the steel bar futures price in the last six years, this paper predicts the steel bar futures price in the future. In this paper, ARIMA model is used to visualize

the historical data of steel bar futures price, and the price is predicted by the model of determining the historical value of time series and the error of historical value prediction. In the selected time period, the model established in this paper compares the predicted value of steel bar futures price with the real price, and the average error is 0.03%, which can describe the trend and fluctuation of futures price well.

Keywords

Rebar Futures, ARIMA Model, Price Forecasts

Copyright © 2023 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

螺纹钢是中国钢材产品中产量最大的品种之一，在基础设施建设领域有大量的应用，对国计民生影响很大。随着中国经济的飞速发展，对螺纹钢的需求不断扩大。但是，螺纹钢市场信息的不通畅、交易价格的不透明等因素带来的螺纹钢价格变动风险不断增强。为了规避上述的风险，中国于 2009 年 3 月推出螺纹钢期货。

螺纹钢期货价格的变化在很大程度上反映了未来一段时间螺纹钢现货市场的供需情况，对螺纹钢现货的价格具有一定的引领作用。在螺纹钢现货价格出现较大的波动前，在螺纹钢期货的价格上可以提早体现。所以螺纹钢期货在一定程度上可为螺纹钢现货市场规避由价格剧烈波动带来的风险，同时对于提高市场的透明度、建立公平公正的现货市场有着重要意义。

本文在 2017 年 1 月 2 日~2022 年 12 月 30 日这段时间螺纹钢期货价格的基础上用 ARIMA 模型预测螺纹钢期货的价格。ARIMA 模型是一种针对时间序列数据进行预测的模型。因其理论基础比较坚实、短期预测效果好等优点在农产品、能源、股票和期货等价格的预测方面应用广泛。

关于螺纹钢期货和螺纹钢现货价格的预测国内已经有学者用不同于 ARIMA 模型的数学模型进行了研究[1] [2] [3] [4]，有学者用 ARIMA 模型预测钢材价格[5]和铁矿石价格[6]。得到了比较好的预测结果。

本文在所选样本时间段内，对螺纹钢期货价格的预测值与真实的价格相比，样本内平均误差为 0.03%，能够很好地刻画期货价格的走势与波动。本文所建 ARIMA 模型可以为有关部门、钢铁企业、螺纹钢买卖双方、期货投资者更好把握市场，进行稳健决策提供一定的依据。

2. ARIMA 模型简介

ARIMA 模型，称为差分整合移动平均自回归模型，又称整合移动平均自回归模型，是时间序列预测分析方法之一。

其模型 ARIMA(p, d, q) 形式为

$$\varphi(L)(1-L)^d X_t = \theta(L)\varepsilon_t$$

其中 L 为滞后算子， d 为差分阶数， $\{\varepsilon_t\}$ 是零均值白噪声序列， $\varphi(L)$ 和 $\theta(L)$ 分别为 p 阶与 q 阶平稳的滞后算子，对应的公式为

$$\varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p;$$

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q.$$

式中 p, d, q 均为非负整数。

3. 数据选择与处理

3.1. 数据选择与可视化

本文是从 Akshare 接口直接下载近六年(2017年1月2日~2022年12月30日)螺纹钢期货价格的数据。共采集到样本数据 1439 个, 将这些跟时间相关的数据构成的初始时间序列记为 X_t 。由于数据来源比较权威且 Akshare 已经实现从采集, 清洗加工等一系列流程, 所以本文没有对下载的原始数据进行清洗。

用 python 软件对 X_t 可视化得到如图 1 所示的螺纹钢期货价格时序图。应用 ARIMA 模型, 需要保证时间序列是平稳的。下面对初始时间序列 X_t 进行平稳性检验。

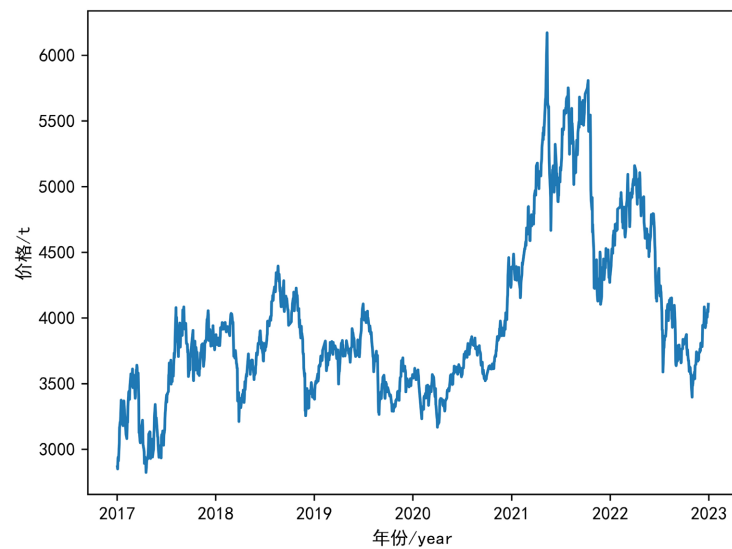


Figure 1. Timing chart of rebar futures prices
图 1. 螺纹钢期货价格时序图

3.2. 模型平稳性检验

利用单位根检验 - ADF 方法检验初始时间序列 X_t 是否含有单位根。通过 python 软件编程运行得到结果如表 1 所示, 初始时间序列 X_t 检验统计量大于各个临界值, 且 $p \approx 0.111718 > 0.05$, 接受原假设。即认为初始时间序列 X_t 是非平稳时间序列, 不能应用 ARIMA 模型。

Table 1. Test results of ADF method for X_t

表 1. 时间序列 X_t 的 ADF 方法检验结果

	t -统计量	p -值
增广迪基 - 富勒检验统计量	-2.515626	0.111718
	1% level	-3.434846
测试临界值	5% level	-2.863525
	10% level	-2.567827

4. 模型建立

4.1. 初始时间序列，一阶差分后，二阶差分后的时序图

由于初始时间序列是非平稳的，需要通过 d 阶差分来实现时间序列平稳。如图 2 所示，从上到下分别为原始数据，一阶差分后数据，二阶差分后数据的时序图。记一阶差分后时间序列为 Z_t ，二阶差分后时间序列为 Y_t 。由图 2 可知经一、二阶差分后的时间序列 Z_t ， Y_t 中的数据基本围绕基准线零线进行上下波动，并不存在显著的趋势或周期，可初步假设时间序列 Z_t ， Y_t 已经达到平稳。

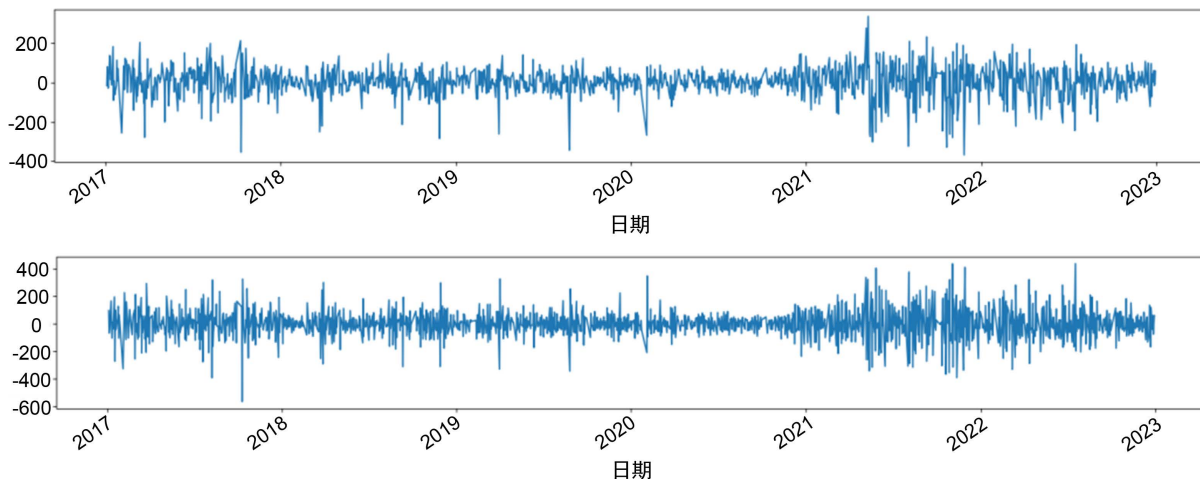


Figure 2. Time sequence diagrams after first and second difference

图 2. 一阶差分，二阶差分后的时序图

4.2. 初始时间序列 d 阶差分的平稳性检验

对于一阶差分后的时间序列 Z_t 是否真正平稳，需要进行单位根检验。采用 ADF 方法检验时间序列 Z_t 是否还含有单位根。单位根检验结果如表 2 所示，可以看出，经过一阶差分后的时间序列 Z_t 检验统计量小于各个临界值，且 $p \approx 0.00 < 0.05$ 。可判断该时间序列为平稳序列。

Table 2. Test results of ADF method for Z_t

表 2. 时间序列 Z_t 的 ADF 方法检验结果

	t -统计量	p -值
增广迪基 - 富勒检验统计量	-39.874853	0.000000
1% level	-3.434846	
测试临界值	5% level	-2.863525
	10% level	-2.567827

同时，对二阶差分 Y_t 的平稳性也做出了验证，如表 3 所示，二阶差分后的数据也是平稳的。

4.3. 白噪声序列检验

因为差分后的数据构成的时间序列有可能是白噪声(纯随机)序列，而对一个白噪声序列没有必要建模。因此对差分后的序列需要进行相关性检验来判断是否是白噪声序列。

Table 3. Test results of ADF method for Y_t
表 3. 时间序列 Y_t 的 ADF 方法检验结果

		t -统计量	p -值
增广迪基 - 富勒检验统计量		-15.035372	9.766×10^{-28}
	1% level	-3.434911	
测试临界值	5% level	-3.434911	
	10% level	-2.567842	

利用混成检验法(LB)对时间序列 Z_t 和 Y_t 进行白噪声检验, 由表 4 可知在前 20 阶延迟下 Q 检验统计量 Z_t 序列的 p 值大于 0.05, 即认为时间序列 Z_t 属于白噪声序列。而 Y_t 序列的 p 值小于 0.05 (见表 5), 与对应时间序列为白噪声序列的假设不成立, 即认为时间序列 Y_t 属于非白噪声序列。因此选择使用二阶差分后时间序列 Y_t 。

Table 4. First twenty order test Q -statistic of Z_t
表 4. 时间序列 Z_t 的前 20 阶 Q 检验统计量

	负载均衡统计值	负载均衡 p -值
10	13.892107	0.177970
20	23.876902	0.247881

Table 5. First twenty order test Q -statistic of Y_t
表 5. 时间序列 Y_t 的前 20 阶 Q 检验统计量

	负载均衡统计值	负载均衡 p -值
10	414.963164	6.137814×10^{-83}
20	425.441160	1.061280×10^{-77}

4.4. 模型建立

经过平稳性检验以及白噪声检验后, 认为原数据二阶差分后的数据符合预测的条件, 基于此在通过穷举的方法对二阶差分后的数据 Y_t 进行遍历, 通过不断缩小范围, 找出 AIC + BIC + HQ (赤池信息量准则 + 贝叶斯信息准则 + HQ 信息准则)最小值所对应的 p 和 q , 可以确定模型参数 $p=0, q=1$ 。由此建立模型 ARIMA(0, 2, 1), 即模型参数自回归项数为 0, 滑动平均项数为 1, 使用的数据为二阶差分后数据。

5. 模型检验

为了验证模型是否适用, 需要对模型进行检验, 首先检验残差序列是否为正态分布, 结果如表 6 所示。因为 $p < 0.05$, 拒绝原假设, 接受备择假设, 认为残差序列具有正态性。

Table 6. Residual normal test diagram
表 6. 残差正态检验图

正态检验结果	统计量 = 85.010532	p -值 = 3.468944×10^{-19}
--------	-----------------	-------------------------------------

如图 3, 图 4 所示, QQ 图散点基本在线上, 直方图也呈现正态性, 因此模型建立成功。

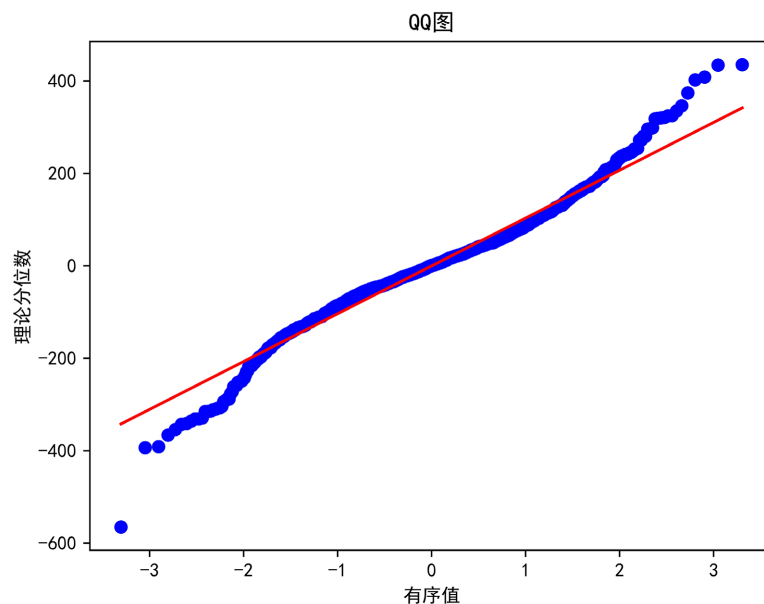


Figure 3. QQ-diagram

图 3. QQ-图

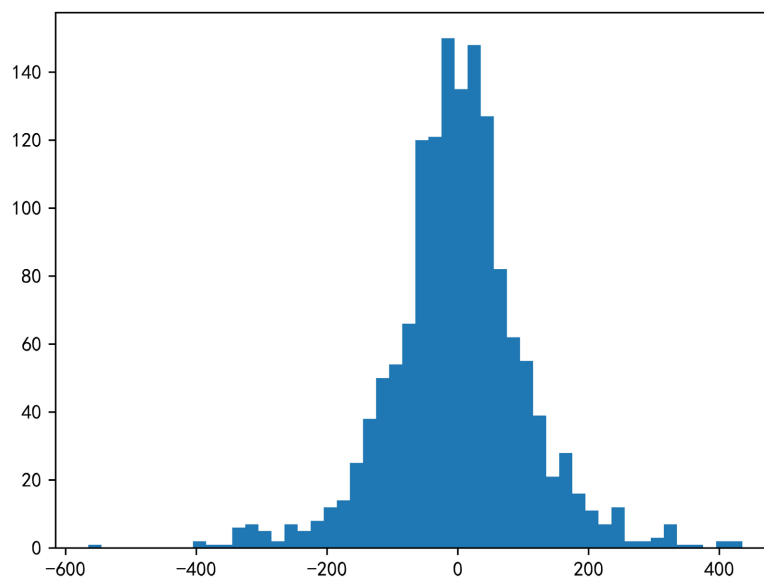


Figure 4. Histogram

图 4. 直方图

6. 数据预测

6.1. 样本内预测

通过上面的分析, 可以确定 $ARIMA(0, 2, 1)$ 模型是螺纹钢期货价格时序图谱线(见图 1)最佳的拟合预测模型, 下面给出样本内预测结果。

根据表 7 可得预测值与真实值平均误差为 0.03%。

Table 7. Prediction result of ARIMA (0, 2, 1) in sample**表 7.** ARIMA (0, 2, 1) 模型样本内预测结果

日期	收盘价	预测值	准确性指标
2022-12-26	4006.00	4105.87	0.00%
2022-12-27	4066.00	4163.33	0.07%
2022-12-28	4050.00	4050.59	0.01%
2022-12-29	4048.00	4048.12	0.00%
2022-12-30	4105.00	4102.47	0.06%

6.2. 样本外内预测

由表 8 可得样本外预测结果，平均误差为 1.17%。

Table 8. Prediction result of ARIMA (0, 2, 1) out of sample**表 8.** ARIMA (0, 2, 1) 模型样本外预测结果

日期	收盘价	预测值	准确性指标
2023-1-3	4100.00	4107.79	0.19%
2023-1-4	4061.00	4108.49	1.17%
2023-1-5	4015.00	4109.19	2.35%
2023-1-6	4031.00	4109.89	1.96%
2023-1-9	4105.00	4111.98	0.17%

7. 结论

本文以 2017 年 1 月 2 日~2022 年 12 月 30 日螺纹钢价格期货价格的数据为样本，选取在时间序列预测上作用明显的 ARIMA 模型对螺纹钢期货价格进行预测，建立了 ARIMA 钢材价格预测模型。通过上面的一系列预测过程，ARIMA 模型成功的预测出了螺纹钢期货的价格，样本内平均误差为 0.03%。结果与真实值之间误差较小，预测结果比较好。但样本外预测结果与真实值之间误差相对样本内的预测结果比较大，平均误差为 1.17%。短期预测比长期预测效果好。这是因为 ARIMA 模型的预测仅从时间序列自身出发，没有考虑其它因素的影响。但螺纹钢期货的影响因素不仅仅只有时间，还有各种各样的因素，比如政策的影响等。在时间并不是数据变化的唯一影响因素时，ARIMA 模型的发挥会受到影响，预测结果的误差会比较大。故推荐将 ARIMA 模型的预测作为一种参考。

基金项目

上海工程技术大学大学生创新项目：螺纹钢期货价格影响因素分析(项目编号：CS2221003)。

参考文献

- [1] 陈海鹏, 卢旭旺, 申铨京, 等. 基于多元线性回归的螺纹钢价格分析及预测模型[J]. 计算机科学, 2017, 44(A2): 61-64, 97.
- [2] 朱明桐, 陈洋洋. 基于 LSTM 模型螺纹钢期货价格预测研究[J]. 新营销, 2019(4): 90-91.

- [3] 李文君. 基于 GAM 模型的中国螺纹钢期货价格预测研究[J]. 中国证券期货, 2021(3): 16-13.
- [4] 王越, 孙若莹. 基于高斯混合隐马尔可夫模型的螺纹钢价格波动预测[J]. 冶金经济与管理, 2022(2): 36-40.
- [5] 王雪飞, 刘志伟. 基于 ARIMA 模型的中国钢材市场价格预测[J]. 经济研究, 2011(1): 20-21, 23.
- [6] 杨锦伟, 李宜懋. 基于 ARIMA 模型的中国铁矿石价格指数预测[J]. 数学的实践与认识, 2020, 50(11): 289-298.