

基于 ARFIMA 模型的钢材价格预测研究

陈雪¹, 申建红^{1,2*}, 徐文慧¹, 朱琛¹

(1. 青岛理工大学 管理工程学院, 山东 青岛 266520; 2. 山东省高校智慧城市建设管理研究中心, 山东 青岛 266520)

摘要: 钢材价格的准确预测有利于施工企业拟定合理的材料采购策略。针对当前钢材价格的预测研究中均未考虑其价格变动的长记忆性, 导致建模过程中有效信息丢失, 预测误差增大。建立了考虑长记忆性的 ARFIMA 钢材价格预测模型, 以青岛市 2014 年 1 月到 2019 年 6 月螺纹钢的价格为研究对象进行了钢材价格预测, 并利用 ARFIMA 模型和 ARIMA 模型的预测值与真实值进行对比分析, 实验结果显示: ARFIMA 模型较 ARIMA 模型的钢材价格预测精准度提高了 1.7%, 且预测效果更稳定。

关键词: 钢材价格预测; ARFIMA 模型; 时间序列; 长记忆性

中图分类号: TU511.3+3

文献标识码: A

Research on Steel Price Forecasting Based on ARFIMA Model

CHEN Xue¹, SHEN Jianhong^{1,2*}, XU Wenhui¹, ZHU Chen¹

(1. School of Management Engineering, Qingdao University of Technology, Qingdao, Shandong 266520, China;

2. University Research Center for Smart City Construction and Management of Shandong Province, Qingdao, Shandong 266520, China)

Abstract: The accurate prediction of steel prices is helpful for construction companies to formulate reasonable material procurement strategies. For the current steel price prediction research, the long memory of its price changes is not considered, resulting in the loss of effective information in the modeling process and the increase of prediction errors. In this paper, the ARFIMA steel price prediction model considering long memory was established. The steel price prediction was carried out based on the price of rebar in Qingdao from January 2014 to June 2019. The predicted values of the ARFIMA model and ARIMA model were used for comparative analysis of the true value. The experimental results show that the accuracy of the steel price prediction of the ARFIMA model is 1.7% higher than that of the ARIMA model, and the prediction effect is more stable.

Key words: steel price forecast; ARFIMA model; time series; long memory

近年来,随着我国城镇化步伐的加快,建筑产业不断发展,工程承包企业之间的竞争日趋激烈。钢材作为工程建设的主要原材料,其价格直接影响着工程承包企业的成本和利润。科学合理地运用价格预测技术,准确把握钢材价格的变化趋势,是工程承包企业制定合理采购策略和库存优化的依据。

目前,国内外针对钢材价格的预测研究相对薄弱。陈海鹏等人综合考虑螺纹钢原材料结算价

与货币汇率等因素对螺纹钢价格变动的影响,提出了一种基于岭估计法的线性回归预测模型,通过与未来螺纹钢的交易价格进行对比验证,得到该模型预测精度较高^[1]。陈希等人针对钢材价格变动性较大的现象,利用 Levenberg-Marquardt 方法改进的 BP 神经网络建立了钢材价格预测模型,同时结合 1990—2008 年的数据对模型进行训练预测,结果显示所建立的钢材价格预测模型有较好的预测精准度^[2]。胡六星为提高工程造价效率以

收稿日期:2020-04-10

基金项目:国家自然科学基金资助项目(71471094)

作者简介:陈雪(1994-),女,陕西宝鸡人,硕士研究生,主要从事工程项目管理方面的研究。

* 通讯作者:申建红(1970-),男,山东青岛人,博士,教授,主要从事工程项目管理方面的研究。

钢材价格为研究对象,在提炼出其趋势项的基础上利用残差序列建立了自回归移动平均模型(Auto-Regressive Moving Average Model, ARMA),并应用该模型对钢材价格进行分析预测,同时结合真实价格进行精度验证,得到其预测结果误差低于5%^[3]。王雪飞等人通过建立差分自回归移动平均模型(Autoregressive Integrated Moving Average model, ARIMA),对多种钢材价格进行预测分析,实验结果表明该模型的短期预测效果较好^[4-5]。王松等人应用广义自回归条件异方差(GARCH)模型与ARMA模型相结合的组合模型对近十六年 Myspic 钢材价格的变动率进行实证分析,发现钢材价格的时间序列中含有着明显的长短记忆特征^[6]。Kaijian He 等人对初始时间序列进行空间重构,并采用曲线去噪方法分离和剔除噪声干扰,实现了对原始时间序列中的动态混沌数据特征进行建模。其所提出的模型在统计上比传统的标杆模型具有更强的鲁棒性^[7]。Tsung-Yin Ou 等人提出了一种极限学习机器(ELM)法与灰色关联分析(GRA)法相结合的材料价格组合预测模型,其预测能力优于 ARIMA 和 GARCH 模型,可以更准确、快速地预测原材料价格^[8]。

目前关于钢材价格预测的研究只考虑了钢材价格变动的短记忆性,并未考虑其价格变动的长记忆性对预测结果的影响。为此,本文从螺纹钢价格时间序列出发,引入钢材价格变动的长记忆性指标,从而建立了基于分数差分自回归移动平均(Autoregressive fractionally integrated moving average, ARFIMA)模型的钢材价格预测模型,通过对钢材价格时间序列变动特性的分析并借助 MATLAB 软件编程进行了钢材价格的预测^[9]。并利用 ARFIMA 模型和 ARIMA 模型的预测值与真实值进行对比分析,实验结果显示,ARFIMA 模型的钢材价格预测精度更高。

1 基本模型

1.1 ARMA 及 ARIMA 模型

ARMA 模型可认为由自回归模型(AR)和移动平均模型(MA)组成,通常可记为 ARMA(p, q),其形式为:

$$X_t = \varphi_0 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (1)$$

式中: $\{\varepsilon_t\}$ 是白噪声序列, p 为自回归模型阶数, q 为移动平均模型阶数, $\varphi_1, \varphi_2, \cdots, \varphi_p$ 和 $\theta_1, \theta_2, \cdots, \theta_q$ 为参

数,且 $\varphi_p \neq 0, \theta_q \neq 0$ 。

若使用后移算子 B 表示,ARMA 模型可简写为:

$$\Phi(B)X_t = \Theta(B)\varepsilon_t \quad (2)$$

$$BX_t = X_{t-1} \quad (3)$$

ARMA 模型适合于分析平稳时间序列,即其均值不受时间影响且方差收敛的狭义平稳过程。目前,由于多数时间序列呈现出非平稳性,因此 ARMA 模型不再适用。而 ARIMA 模型则适用于分析非平稳时间序列,它是在 ARMA 模型的基础上加入差分处理,通过剔除时间序列中非平稳趋势,得到平稳时间序列,从而展开分析研究。其模型 ARIMA(p, d, q) 形式为:

$$\varphi(L)(1-L)^d X_t = \theta(L)\varepsilon_t \quad (4)$$

式中: L 为滞后算子, d 为差分参数, $\{\varepsilon_t\}$ 是白噪声序列, $\varphi(L)$ 和 $\theta(L)$ 分别为 p 阶与 q 阶平稳的滞后算子,可表示为:

$$\varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p \quad (5)$$

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q \quad (6)$$

式中: p, d, q 均为非负整数。

1.2 ARFIMA 模型

在 ARIMA 模型中,当 $0 < d < 1$ 时, X_t 具有无限方差。此时,使用一般的识别方法进行时间序列分析,可得到运算时必须进行一阶差分。但由于一阶差分处理会丢失时间序列中的低频信息,从而造成时间序列的过度差分,使得 ARIMA 模型只适用于具有短记忆性的时间序列分析。ARFIMA(p, d, q) 模型形式同 ARIMA 模型,不同的是差分参数 d 采用了介于 -0.5 和 0.5 之间的实数,它是用来分析时间序列中的长记忆性成分,而参数 p 和 q 是用来分析时间序列中的短记忆性成分。其中分数差分算子为:

$$(1-L)^d = 1 - dL + \frac{d(d-1)L^2}{2!} - \frac{d(d-1)(d-2)L^3}{3!} + \cdots \quad (7)$$

对于任何 $d > -1$ 的实数,上式都可以用一个超几何函数表示为:

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} L^k \quad (8)$$

式中: $\Gamma(x)$ 为 Gamma 函数。差分参数 d 可以通过赫斯指数(H) 来确定^[10], $d = H - 0.5$ 。

对于一个由 N 个数值组成的时间序列集合 X_t ,将其划分成长度为 n 的 m 个子序列,并将这 m 个子序列的均值都定义为:

$$\bar{x}_l = \frac{1}{n} \sum_{k=1}^n x_{k,l}, 1 \leq k \leq n, 1 \leq l \leq m \quad (9)$$

式中: $x_{k,l}$ 为第 l 个时间序列中第 k 个观测值。

计算每个子序列的累计均值离差为:

$$X_{k,l} = \sum_{i=1}^k (x_i - \bar{x}_l), 1 \leq k \leq n \quad (10)$$

根据上式,计算出每个子序列的极差为:

$$R_l = \max_{1 \leq k \leq n} X_{k,l} - \min_{1 \leq k \leq n} X_{k,l} \quad (11)$$

则其重标极差均值可表示为:

$$(R/S)_n = \frac{1}{m} \sum_{l=1}^m \left(\frac{R_l}{S_l} \right) \quad (12)$$

式中: S_l 表示第 l 个子序列的标准差为:

$$S_l = \sqrt{\frac{1}{m} \sum_{k=1}^m (x_{k,l} - \bar{x}_l)^2} \quad (13)$$

如果时间序列满足:

$$(R/S)_n = \alpha n^H \quad (14)$$

则该时间序列符合赫斯特率,式中, α 为系数, H 为赫斯指数。其中 H 可以由公式(14) 两侧取对数,并结合最小二乘法拟合得到。当 $0 \leq H < 0.5$ 时,则表示时间序列具有反持久性,其表现出强烈的异变特性;当 $H = 0.5$ 时,则表示时间序列为随机序列;当 $0.5 < H < 1$ 时,则表示时间序列具有长记忆特性,且数值越大长记忆性越强。

因此,ARFIMA 模型能够避免过度差分,最大限度地保留时间序列在低频处的信息,更适用于具有长记忆性的时间序列分析研究。

2 基于 ARFIMA 模型的钢材价格预测

本文在借助 SPSS 软件缺失值分析功能对钢材价格原始数据进行补齐的基础上,通过平稳性检验、长记忆性检验、数据拟合等过程利用 ARFIMA 模型对青岛市直径为 18 mm 的 HRB400 螺纹钢价格进行了分析预测,同时结合 ARIMA 模型对该钢材价格的预测结果进行对比分析。

2.1 原始数据处理

本文数据选用青岛市 2014 年 1 月 1 日到 2019 年 6 月 30 日直径 18 mm 的 HRB400 钢材价格作为研究对象。共采集到样本数据 1 875 个,由于各种因素导致收集的原始数据存在部分缺失,共缺失 131 个。

为保证预测结果的准确性,本文采用 SPSS 软件缺失值分析功能对缺失数据进行了分析,并在此基础上将数据补齐。其中单变量统计结果见表

1,通过 SPSS 数据分析中的多变异分析(MVA)得到钢材价格变动时间序列服从正态分布,且多变量检验(Little's MCAR)的显著性水平小于 5%,由此可判断原始数据为随机缺失型数据。本文采用期望-最大似然估计法(EM)对存在缺失的原始数据进行填补。

表 1 单变量统计

Tab. 1 Univariate statistics

个案例	平均值	标准 偏差	缺失		极值数	
			计数	百分比	低	高
V ₁	1 875	3 335.92	873.114	131	6.53	0

2.2 ARFIMA 模型

利用经数据处理后的钢材日价格共得到 287 个周价格样本作为初始时间序列,记为 X_t 。图 1 为利用 SPSS 软件对该时间段内青岛市直径为 18 mm 的 HRB400 螺纹钢价格数据绘制的时序图,其价格在早期呈现出下降趋势,后期呈现出上升趋势,初步识别为非平稳时间序列。并利用单位根检验-ADF 方法检验时间序列 X_t 是否含有单位根。

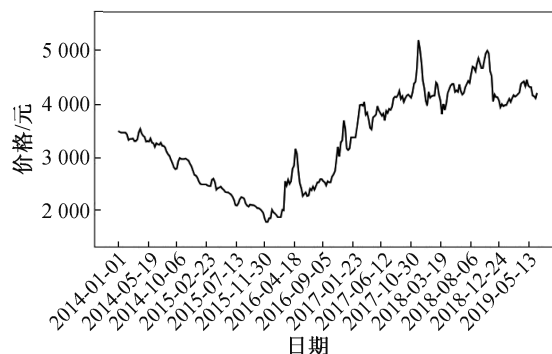


图 1 螺纹钢价格时序图

Fig. 1 Rebar price timing chart

单位根检验结果如图 2 所示,初始时间序列 X_t 检验统计量大于各个临界值,此时原假设成立,即认为初始时间序列 X_t 是非平稳时间序列。

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.279 835	0.766 4
Test critical values:		
1% level	-2.573 277	
5% level	-1.941 965	
10% level	-1.615 940	

图 2 时间序列 X_t 的 ADF 检验结果

Fig. 2 ADF test result of time series X_t

2.3 长记忆性检验

图 3 为时间序列 X_t 的自相关函数图,图中表明钢材价格的变动没有明显的季节性变化,且自

相关系数逐渐减小呈缓慢下降的趋势,在滞后 80 阶后才开始小于 0,表现出长期相关特性,可初步判断时间序列 X_t 含有一定程度的长记忆性。根据公式(8)—公式(13) 计算得到 H 为 0.835,赫斯指数大于 0.5 进一步表明时间序列 X_t 确实具有较强的长记忆性。

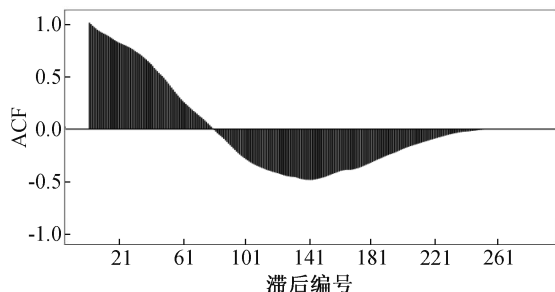


图3 时间序列 X_t 自相关函数图

Fig. 3 Time series X_t autocorrelation function graph

2.4 模型建立

将样本数据分为两部分,取 2014 年 1 月 1 日到 2019 年 5 月 12 日的周价格数据,即前 280 个样本数据作为建模序列,记为 Y_t ,用来构建 ARFIMA 模型和确定模型参数。取 2019 年 5 月 13 日至 2019 年 6 月 30 日的周价格数据,即后 7 个样本数据用来验证所建立的 ARFIMA 模型的可行性与精确度。

应用 R/S 分析法求得时间序列 Y_t 的赫斯指数为 0.835,则所建模型中的分数差分参数 d 为 0.335。对时间序列 Y_t 进行分数差分剔除其中存在的长记忆性,从而获得只含有短记忆的时间序列,记为 Z_t 。图 4 为时间序列 Z_t 的时序图,由图可知经分数差分后的时间序列 Z_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,此时可初步判断时间序列 Z_t 已达到平稳。

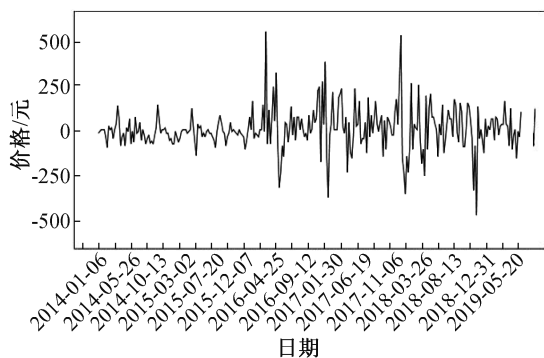


图4 时间序列 Z_t 时序图

Fig. 4 Time series Z_t timing diagram

采用 ADF 方法检验时间序列 Z_t 是否还含有单位根,结果如图 5 所示,可以看出经过分数差分处理后的时间序列 Z_t 检验统计量小于各个临界值,此时可判断该时间序列为平稳时间序列。通过利用混成检验(LB)法对该时间序列纯随机性进行检验,可知在前 6 阶和前 12 阶延迟下 Q 检验统计量的 p 值均小于 0.05,于是时间序列 Z_t 为纯随机序列的假设不成立,即认为时间序列 Z_t 属于非白噪声序列。

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-14.569 05	0.000 0
Test critical values:		
1% level	-2.573 307	
5% level	-1.941 969	
10% level	-1.615 937	

图5 时间序列 Z_t 的 ADF 检验结果

Fig. 5 ADF test result of time series Z_t

通过分数差分处理后的时间序列 Z_t 的自相关系数在滞后 2 阶时其值小于 2 倍的标准差,同时偏自相关系数在滞后 1 阶时其值小于 2 倍的标准差,于是可以确定模型参数 $p=2, q=1$ 。由此可建立模型为 ARFIMA(2,0.335,1),并对时间序列 Z_t 进行数据拟合,同时采用 LB 检验其残差序列,可知在前 6 阶和前 12 阶延迟下 Q 检验统计量的 p 值均大于 0.05,此时认为该残差序列是白噪声序列,即可表明 ARFIMA(2,0.335,1)是显著有效的。为了与 ARIMA 模型的预测结果进行对比分析,本文同时利用 ARIMA 模型对时间序列 Y_t 进行数据处理,模型参数对比如表 2 所示。

表2 ARFIMA 和 ARIMA 模型参数估计值

Tab. 2 ARFIMA and ARIMA model parameter estimates

参数	d	φ_1	φ_2	θ_1
ARFIMA 模型	0.335	0.862	-0.158	0.261
ARIMA 模型	1	0.834	-0.151	0.233

2.5 模型预测

为了验证 ARFIMA 模型的适用性,根据以上模型参数估计,确定以 ARFIMA(2,0.335,1)模型对直径为 18 mm 的 HRB400 钢材价格进行拟合预测。图 6 为 ARFIMA(2,0.335,1)模型的拟合预测结果与真实值的比较,结果显示 ARFIMA(2,0.335,1)具有很好的拟合效果。

本文通过分别利用 ARIMA 模型和 ARFIMA 模型对 2019 年 5 月 13 日至 2019 年 6 月 30 日,7 周的钢材周价格进行预测分析,同时结合 7 周的真实值进行累计差值和相对误差的精度验证,如表 3

表3 ARFIMA模型和ARIMA模型预测结果

Tab.3 ARFIMA model and ARIMA model prediction results

时间	真实值	ARFIMA 预测值	ARIMA 预测值	ARFIMA 差值	ARIMA 差值	ARFIMA 相对误差/%	ARIMA 相对误差/%
2019.5.13—5.19	4 330	4 341	4 408	11	78	0.25	1.80
2019.5.20—5.26	4 300	4 318	4 423	18	123	0.42	2.86
2019.5.27—6.02	4 300	4 322	4 429	22	129	0.51	3.00
2019.6.03—6.09	4 140	4 202	4 314	62	174	1.50	4.20
2019.6.10—6.16	4 130	4 131	4 178	1	48	0.02	1.16
2019.6.17—6.23	4 090	4 119	4 233	29	143	0.71	3.50
2019.6.24—6.30	4 190	4 158	4 108	-32	-82	-0.76	-1.96

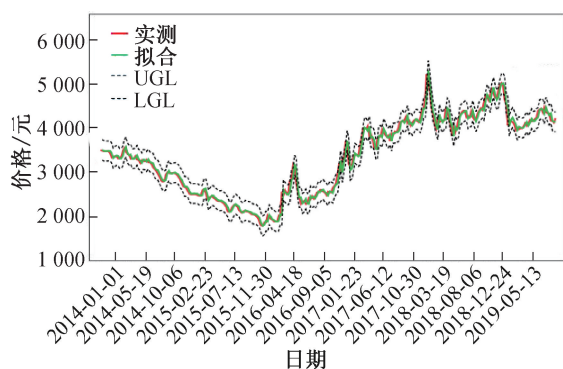


图6 ARFIMA(2,0.335,1)模型预测拟合图

Fig.6 ARFIMA (2, 0.335, 1) model prediction fit graph

所示。从表中可以看出,ARIMA模型预测值较真实值的累计差值为777元,平均相对误差为2.08%;ARFIMA模型预测值较真实值的累计差值为175元,平均相对误差为0.38%。综上所述,在钢材价格预测中ARFIMA模型较ARIMA模型预测结果精度更高。

3 结论

本文以2014年1月~2019年6月青岛市直径18 mm的HRB400钢价格数据为样本,建立了ARFIMA(2,0.335,1)钢材价格预测模型,并与ARIMA模型进行对比分析,得出以下几点结论:

1) ARFIMA模型中的分数差分较ARIMA模型的整数差分能更好地保证钢材价格原始时间序列信息的完整性,避免有效信息的缺失。

2) 螺纹钢价格时间序列具有较强的长记忆性,ARFIMA模型在保持ARIMA短期预测精度的基础上,充分考虑了钢材价格变动的长记忆性。

3) 实验结果显示,ARFIMA模型预测值较真实值平均相对误差仅为0.38%,相较于ARIMA模型的钢材价格预测精准度提高了1.7%,且模型拟合

度更高、预测效果更稳定。对于房地产企业及时掌握市场动向,降低库存资金占用率,制定短期钢材采购计划有一定的参考意义。

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(11A): 61-64+97.
- [2] 陈希,王景强,王玉峰.基于神经网络钢材价格的分析与预测[J]. 自动化与仪表, 2010(12): 7-10.
- [3] 胡六星.基于时间序列的建筑工程造价预测研究[J]. 太原理工大学学报, 2012, 43(6): 706-709+714.
- [4] 王雪飞,刘志伟.基于ARIMA模型的中国钢材市场价格预测[J]. 中国城市经济, 2011(1): 20-21+23.
- [5] BOX GE E P, JEKINS G M, REINSEL G C, et al. Time Series Analysis, Forecasting and Control [M]. 5th ed. Hoboken, New Jersey: John Wiley & Sons, 2016.
- [6] 王松,王超发,孙金岭.钢材价格指数波动率的实证分析[J]. 统计与决策, 2017(19): 173-175.
- [7] HE K J, LU X J, ZOU Y C, et al. Forecasting Metal Prices with a Curvelet Based Multiscale Methodology[J]. Resources Policy. 2015(45): 144-150.
- [8] OU Tsungyin, CHENG Chenyang. Dynamic Cost Forecasting Model Based on Extreme Learning Machine—A Case Study in Steel Plant[J]. Computers & Industrial Engineering, 2016(101): 544-553.
- [9] GRANGER C W J, JOYEUX R. An Introduction to Long Memory Time Series Models and Fractional Differencing [J]. Journal of Time Series Analysis, 1980, 1(6): 15-39.
- [10] ELLIS C. Estimation of the ARFIMA(p,d,q) Fractional Differencing Parameter(d) Using the Classical Rescaled Adjusted Range Technique[J]. International Review of financial Analysis, 1999, 8(1): 53-65.

(责任编辑 周雪梅)