

具有趋势项的残差自回归移动平均模型的修正预测

谢飞*, 武新乾

河南科技大学数学与统计学院, 河南 洛阳

收稿日期: 2024年2月27日; 录用日期: 2024年3月21日; 发布日期: 2024年3月27日

摘要

为了研究新数据产生背景下时序模型的预测问题, 本文针对含多项式趋势项的残差ARMA模型探讨了无需重新拟合的不变模型修正预测法。采用K折交叉验证, 并以平均RMSE作为评价指标, 确定最佳的多项式拟合次数。基于最小二乘法和线性时间序列建模方法进行了数值模拟和实证分析。结果显示, 与需要重新拟合的改变模型修正预测法相比, 无需重新拟合的不变模型修正预测法具有一定的优越性, 计算成本小, 且优于传统的未修正预测法, 可以看作是一种简单易行的修正预测方法。

关键词

K折交叉验证, ARMA, 修正预测

A Correcting Prediction Method of Residual ARMA Models with Polynomial Trend

Fei Xie*, Xinqian Wu

School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang Henan

Received: Feb. 27th, 2024; accepted: Mar. 21st, 2024; published: Mar. 27th, 2024

Abstract

In order to investigate the predictive issues of time series models under the context of new data generation, an invariant model correcting prediction method of residual ARMA models with polynomial trend components is discussed in this paper. K-fold cross-validation is used to determine the optimal degree of polynomial fitting in the sense of the average root mean squared error (RMSE).

*通讯作者。

Numerical simulation and empirical analysis are conducted based on the least squares method and linear time series modeling approach. The results show that the invariant model correcting prediction method, which does not need to re-estimate models and has lower computational cost, exhibits certain advantages over the changing model correcting prediction method. Furthermore, the invariant model correcting prediction method outperforms the traditional uncorrected prediction method. Thus, the invariant model correcting prediction method can be viewed as a simple and feasible correcting prediction approach.

Keywords

K-Fold Cross-Validation, ARMA, Correcting Prediction

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

时间序列分析是数据分析的一个重要分支。通过分析时间序列,可以深入了解各种现象的发展趋势,并将其应用于如资源规划、环境控制等领域。多数现有研究专注于模型选优以提高预测精度。例如,李文武等人(2022) [1]为提高综合能源系统负荷的预测精度,提出了一种结合聚合混合模态分解和时序卷积神经网络(TCN)的预测框架。杨芸珍等人(2020) [2]为了提高 Klobuchar 模型的电离层延迟修正精度,提出了基于 ARIMA 误差修正预测的精细化方法。但面临持续增长的数据量,如何有效地利用新数据进行修正预测以提高准确性,这是一个值得关注的问题。

利用新数据进行修正预测主要有两种策略。一种策略是将新数据纳入到原始数据集中,然后基于这个更新后的数据集重新拟合模型,再利用重新拟合后的模型进行预测,这种方法称为重新拟合模型下的修正预测(简称“改变模型修正预测”)方法。另一种策略是在新数据产生后,并不需要重新拟合模型,而是利用平稳时间序列的特性对已有的预测结果进行修正[3],这种方法称为拟合模型不变下的修正预测(简称“不变模型修正预测”)方法。例如,郭祥琳等(2018) [4]通过介绍时间序列修正预测方法,验证其在建筑物沉降监测中的适用性。刘军等(2011) [5]考虑到电离层变化的周期性,为了进一步减小极值点处预测误差,提出了改进的修正预测法,用上一周期的预测误差对修正预测法结果进行改正。

本文主要在 ARMA 模型的修正预测基础上,研究具有多项式趋势的残差 ARMA 模型的修正预测方法。

2. 模型构建方法

考虑含有多项式趋势项的残差自回归移动平均模型

$$\begin{cases} x_t = a_0 + a_1 t + \dots + a_n t^n + u_t \\ u_t = \phi_1 u_{t-1} + \dots + \phi_p u_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \\ E(\varepsilon_t) = 0, \text{var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(u_s \varepsilon_t) = 0, \forall s < t \end{cases} \quad (1)$$

其中 x_t 代表时间序列在 t 时刻的值, a_0, a_1, \dots, a_n 代表多项式回归模型的系数, n 代表多项式的最高次数,

u_t 代表在时间 t 处的误差项, $\varphi_1, \varphi_2, \dots, \varphi_p$ 代表自回归部分的模型参数, $\theta_1, \theta_2, \dots, \theta_q$ 代表移动平均部分的模型参数, ε_t 是残差 ARMA 模型中的误差项。

2.1. 多项式拟合次数的确定

通过 K 折交叉验证确定多项式拟合的最佳次数 m , 通常选取 K 等于 10 [6]。依次对一次多项式、二次多项式……执行一次 K 折交叉验证。在交叉验证的每一轮中, 数据首先被分割为 K 份, 每次选取一份作为测试集, 其余部分作为训练集[7]。接着, 根据给定的多项式次数对训练集进行模型拟合, 并记录该次数下模型在测试集上的均方根误差(RMSE)。这样, 每一个多项式次数将得到 K 个 RMSE 值, 将平均 RMSE 值作为该多项式次数下的平均误差。最后, 找出具有最小平均误差的多项式次数, 将其确定为最佳次数。

基于得到的最佳次数 m , 进行回归分析, 得到最小二乘估计的回归系数, 获得回归方程, 然后, 对回归方程进行 F 检验和 t 检验, 经过检验之后, 对残差做进一步的分析。

2.2. 残差 ARMA 模型参数的确定

基于拟合的趋势模型, 得到残差序列。首先对其进行平稳性检验, 若不平稳, 考虑使用差分、对数转换等方法进行平稳化处理; 对于平稳的残差序列, 基于自相关函数(ACF)和偏自相关函数(PACF)图形以及 AIC 和 BIC 准则, 对模型进行识别和阶数的确定。最后, 利用最小二乘法或极大似然估计法来确定模型的参数。

2.3. u_t 的修正预测

对于一个平稳可逆的 ARMA(p, q) 模型, 它的传递形式为[3]

$$u_t = \sum_{j=0}^{\infty} G_j \varepsilon_{t-j} \quad (2)$$

式中 $\{G_0, G_1, G_2, \dots\}$ 为 Green 函数。

通过待定系数法, 可以得到 ARMA(p, q) 模型场合下 Green 函数的递推公式为

$$\begin{cases} G_0 = 1 \\ G_k = \sum_{j=1}^k \phi'_j G_{k-j} - \theta'_k, k \geq 1 \end{cases} \quad (3)$$

式中 ϕ'_j, θ'_k 为

$$\phi'_j = \begin{cases} \phi_j, 1 \leq j \leq p \\ 0, j > p \end{cases}, \theta'_k = \begin{cases} \theta_k, 1 \leq k \leq q \\ 0, k > q \end{cases} \quad (4)$$

已知在旧数据 $u_t, u_{t-1}, u_{t-2}, \dots$ 的基础上, u_{t+l} 的预测值为

$$\hat{u}_t(l) = G_l \varepsilon_t + G_{l+1} \varepsilon_{t-1} + \dots \quad (5)$$

如果获得 1 个新观察值 u_{t+1} , 则 u_{t+l} 的修正预测值为

$$\hat{u}_{t+1}(l-1) = G_{l-1} \varepsilon_{t+1} + \hat{u}_t(l) \quad (6)$$

式中 $\varepsilon_{t+1} = u_{t+1} - \hat{u}_t(1)$ 是 u_{t+1} 的一步预测误差。

如果获得 s 个新观察值 u_{t+1}, \dots, u_{t+s} ($1 \leq s \leq l$), 则 u_{t+l} 的修正预测值为

$$\hat{u}_{t+s}(l-s) = G_{l-s} \varepsilon_{t+s} + \dots + G_{l-1} \varepsilon_{t+1} + \hat{u}_t(l) \quad (7)$$

式中, $\varepsilon_{t+i} = u_{t+i} - \hat{u}_{t+i-1}$ (1) 是 u_{t+i} ($i=1,2,\dots,s$) 的一步预测误差。

3. Monte Carlo 模拟算例

不妨设模型(1)中 $a_0 = 3, a_1 = 2, a_2 = 1, \phi_1 = 0.7, \theta_1 = -0.15, \varepsilon_t \sim N(0,1)$ 。此时, 模型为

$$\begin{cases} x_t = 3 + 2t + t^2 + u_t \\ u_t = 0.7u_{t-1} + \varepsilon_t + 0.15\varepsilon_{t-1} \end{cases} \quad (8)$$

在 Monte Carlo 模拟中, 进行 1000 次模拟实验, 每次实验生成 105 个数据, 其中前 100 个用于构建模型, 剩余的 5 个用于验证。基于模型(8)拟合的平均 MAE (平均绝对误差)为 0.767, 平均 MSE (均方误差)为 0.924; 预测的平均 MAE 为 1.213, 平均 MSE 为 2.461。表 1 和表 2 分别给出了不变模型修正预测结果和改变模型修正预测结果。

Table 1. Mean MAE and mean MSE for modified predictions of the invariant model

表 1. 不变模型修正预测的平均 MAE 和平均 MSE

	1 期修正	2 期修正	3 期修正	4 期修正
平均 MAE	1.176	1.141	1.107	1.073
平均 MSE	2.387	2.312	2.246	2.179

Table 2. Mean MAE and mean MSE of modified predictions by changing models

表 2. 改变模型修正预测的平均 MAE 和平均 MSE

	1 期修正	2 期修正	3 期修正	4 期修正
平均 MAE	1.206	1.167	1.081	0.956
平均 MSE	2.455	2.306	2.006	1.528

由表 1 和表 2 可以看出, 两种修正预测法的平均 MAE 和平均 MSE 均小于未修正预测的平均 MAE 和平均 MSE, 并且随着新数据的产生和修正期数的增加, 平均 MAE 和平均 MSE 逐渐减小, 显示了修正预测的优势。进一步地, 在初期修正时, 不变模型修正预测法的平均 MAE 和平均 MSE 较低, 而在后期修正时, 改变模型修正预测法的平均 MAE 和平均 MSE 较低, 说明了这两种方法各有优劣。然而, 考虑到改变模型修正预测法需要不断地重新拟合模型, 将导致计算成本的大量增加, 在这个意义下, 不变模型修正预测法不失为一种简单易行的方法。

4. 实证分析

选取 2023 年 1 月到 6 月的 118 期上证指数收盘价做为原始数据, 记为 $\{x_t\}(t=1,2,\dots,118)$, 如图 1 所示。为确定多项式 $a_0 + a_1t + \dots + a_nt^n$ 的最佳次数及其系数, 采用了 10 折交叉验证方法。通过计算不同次数多项式的平均 RMSE, 以平均 RMSE 最小的次数确定多项式的最佳次数, 结果如表 3 所示。当多项式为 2 阶时, 其平均 RMSE 达到最小值 43.381, 再使用最小二乘法对二次多项式进行拟合, 得到的趋势项模型为

$$x_t = 3160.48432 + 4.73715x - 0.03868x^2 + u_t \quad (9)$$

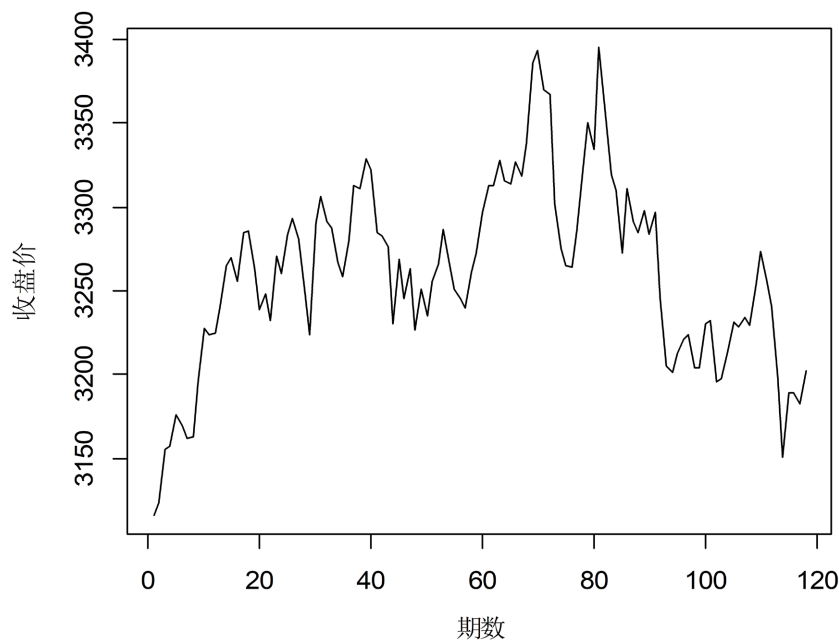


Figure 1. Shanghai Composite Index closing prices from January to June 2023 $\{x_t\}$

图 1. 2023 年 1 月至 6 月每期上证指数收盘价 $\{x_t\}$

Table 3. Average RMSE for polynomial fitting

表 3. 多项式拟合的平均 RMSE

次数	1 次	2 次	3 次	4 次
平均 RMSE	59.575	43.381	50.233	72.433

对模型(9)进行 F 检验, 统计量值为 68.68, 对应的 P 值为 0.00, 说明该模型显著成立。表 4 的多项式参数估计和检验结果显示, 常数项、一次项和二次项的 P 值均为 0.00, 均通过了 t 检验, 即参数估计值显著。

Table 4. Parameter estimates and statistics

表 4. 参数估计及统计量

变量	系数	t 统计量	P 值
截距	3160.48432	300.49	0.00
一次项	4.73715	11.61	0.00
二次项	-0.03868	-11.64	0.00

通过去掉多项式趋势项, 得到残差序列 $\{u_t\}$, 如图 2 所示。容易看出, 残差序列 $\{u_t\}$ 展现出平稳特性。经 ADF 单位根检验, 其 P 值为 0.00, 表明残差序列是平稳序列。经 Ljung-Box 检验得到 P 值为 0.00, 确定该序列是非白噪声序列。

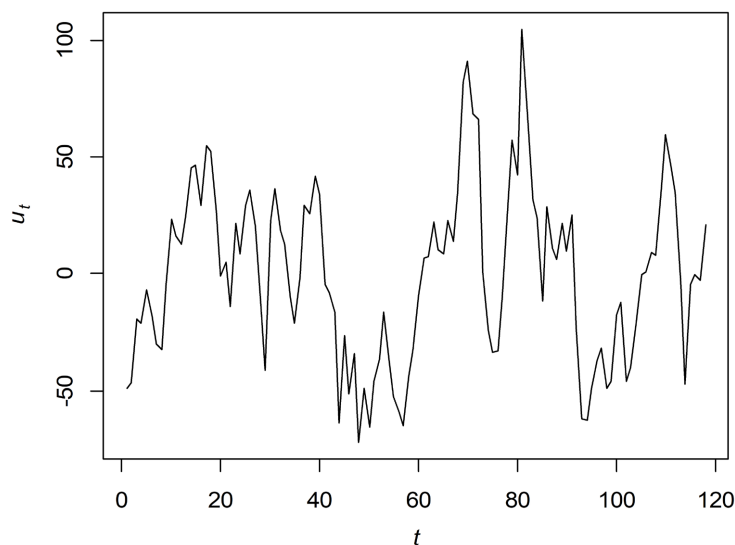


Figure 2. $\{u_t\}$ Timing diagram of residual sequence

图 2. 残差序列 $\{u_t\}$ 时序图

图 3 给出了残差序列 $\{u_t\}$ 的自相关图和偏自相关图。由图 3 可知, 偏自相关图在一阶后截尾, 选择 $p = 1$ 。自相关图显示了拖尾特性, 为确定适当的 q 值, 参考 AIC 和 BIC 的评价准则, 根据表 5 的数据, 当 $p = 1$ 且 $q = 0$ 时, AIC 和 BIC 均达到最小值, 因此, 对残差序列建立 AR(1)模型。采用最小二乘法对数据进行拟合, 最终得到的模型为

$$\begin{cases} x_t = 3160.48432 + 4.73715x - 0.03868x^2 + u_t \\ u_t = 0.795u_{t-1} + \varepsilon_t \end{cases} \quad (10)$$

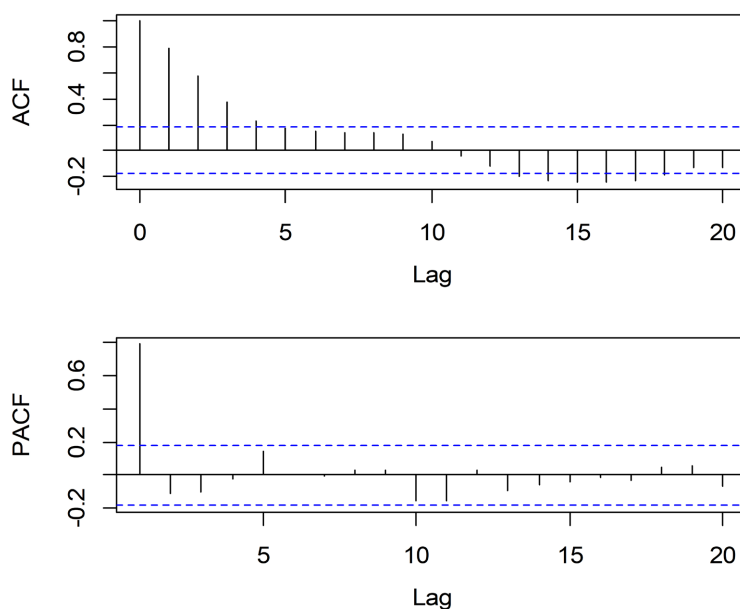


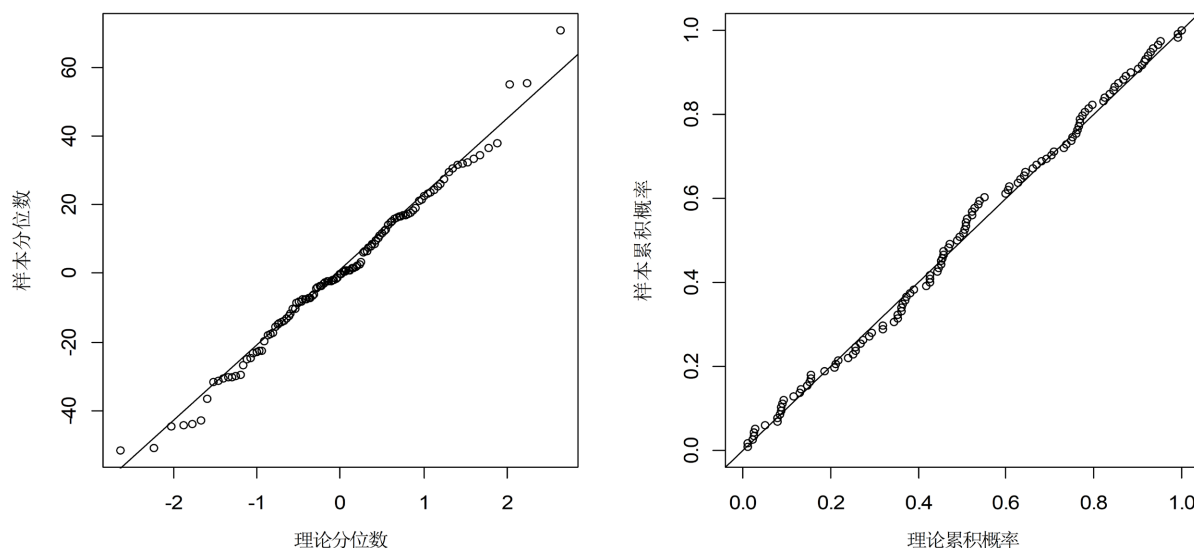
Figure 3. Autocorrelation and partial autocorrelation of residual sequence $\{u_t\}$

图 3. 残差序列 $\{u_t\}$ 的自相关图和偏自相关图

Table 5. AIC and BIC values for residual sequence $\{u_t\}$ modeling**表 5.** 残差序列 $\{u_t\}$ 建模的 AIC 和 BIC 的值

q	AIC	BIC
0	1073.951	1079.492
1	1074.746	1083.058
2	1074.763	1085.846
3	1075.470	1089.323

对序列 $\{\varepsilon_t\}$ 进行白噪声检验, P 值为 0.356, 接受原假设, 即该序列为白噪声。图 4 为序列 $\{\varepsilon_t\}$ 的 QQ 图和 PP 图, 这两个图中, 数据点紧密地沿着参考线排列, 表明 $\{\varepsilon_t\}$ 近似服从正态分布。

**Figure 4.** QQ chart (left) and PP chart (right) of residual sequence $\{\varepsilon_t\}$ **图 4.** 残差序列 $\{\varepsilon_t\}$ 的 QQ 图(左图)和 PP 图(右图)

基于模型(10), 预测了 2023 年 7 月上证指数的前 5 期收盘价, 如表 6 所示。5 期预测的相对误差均未超过 2%。特别是在第 123 期, 预测的相对误差甚至未达 1%。

Table 6. Unrevised predictions based on model (10)**表 6.** 基于模型(10)的未做修正的预测结果

期数	原始值	预测值	相对误差
119	3243.98	3193.295	1.562%
120	3245.35	3185.341	1.849%
121	3222.95	3178.017	1.394%
122	3205.57	3171.178	1.073%
123	3196.61	3164.707	0.998%

经将 7 月份前 4 期收盘价逐步纳入为新数据进行修正预测后, 不变模型修正预测结果见表 7, 改变模型修正预测结果见表 8。由表 7 和表 8 可知, 两种修正预测的误差均得到了降低。特别地, 在不变模型的 2 期修正预测中, 第 121 期和 123 期, 以及 3 期修正预测的第 122 期, 预测的相对误差均不足 0.1%。同样, 在改变模型的 2 期修正预测中, 第 121 期和 122 期, 以及 3 期修正预测的第 122 期, 预测的相对误差也均不超过 0.1%。

Table 7. Revised predictions for the unchanged model

表 7. 不变模型的修正预测结果

期数	1 期修正值	相对误差	2 期修正值	相对误差	3 期修正值	相对误差	4 期修正值	相对误差
120	3225.634	0.607%						
121	3210.048	0.400%	3225.722	0.086%				
122	3196.641	0.279%	3209.100	0.110%	3206.897	0.041%		
123	3184.949	0.365%	3194.854	0.055%	3193.103	0.110%	3192.048	0.143%

Table 8. Revised predictions for changing the model

表 8. 改变模型的修正预测结果

期数	1 期修正值	相对误差	2 期修正值	相对误差	3 期修正值	相对误差	4 期修正值	相对误差
120	3226.056	0.595%						
121	3210.722	0.379%	3225.322	0.074%				
122	3197.448	0.253%	3205.675	0.003%	3203.434	0.067%		
123	3185.808	0.338%	3189.336	0.228%	3187.561	0.283%	3189.572	0.220%

进一步地, 模型(10)拟合的 MAE 为 17.392, MSE 为 503.222; 预测的 MAE 为 384, MSE 为 2077.921。表 9 给出了不变模型修正预测的 MAE 和 MSE, 表 10 给出了改变模型修正预测的 MAE 和 MSE。由表 9 和表 10 可知, 两种修正预测的 MAE 和 MSE 比不修正预测的 MAE 和 MSE 小得多, 表明了修正预测的优势。此外, 两种修正预测方法相比较, 在一个新数据引入并进行 1 期修正下, 改变模型修正预测的 MAE 和 MSE 较低, 但随着后续新数据的逐步引入, 改变模型修正预测的 MAE 和 MSE 却较高, 显示了改变模型修正预测法的不足。

Table 9. MAE and MSE predicted by modified invariant models

表 9. 不变模型修正预测的 MAE 和 MSE

	1 期修正	2 期修正	3 期修正	4 期修正
MAE	13.302	2.686	2.417	4.562
MSE	192.722	7.743	7.031	20.815

Table 10. MAE and MSE for modified predictions by changing models
表 10. 改变模型修正预测的 MAE 和 MSE

	1 期修正	2 期修正	3 期修正	4 期修正
MAE	12.611	3.250	5.592	7.038
MSE	176.101	19.517	43.222	49.537

5. 结束语

本文主要探究了含有多项式趋势的残差自回归移动平均模型的修正预测方法, 模拟结果显示, 与需要重新拟合的改变模型修正预测法相比, 不变模型修正预测法计算成本小, 优于未修正预测法, 可以看出是一种简单易行的修正预测方法。实证分析结果也表明了不变模型修正预测法的优越性和可行性。

基金项目

国家自然科学基金资助项目(11601126); 河南省科技重点攻关项(182102210286)。

参考文献

- [1] 李文武, 张鹏宇, 石强, 等. 基于聚合混合模态分解和时序卷积神经网络的综合能源系统负荷修正预测[J]. 电网技术, 2022, 46(9): 3345-3357.
- [2] 杨芸珍, 刘立龙, 黄良珂, 等. 基于 ARIMA 误差修正预测的 Klobuchar 模型精化[J]. 桂林理工大学学报, 2020, 40(3): 551-556.
- [3] 王燕. 应用时间序列分析[M]. 北京: 中国人民大学出版社, 2005.
- [4] 郭祥琳, 成枢, 程方. 时间序列分析的修正预测在建筑物沉降监测中的应用[J]. 北京测绘, 2018, 32(5): 546-549.
- [5] 刘军, 柴洪洲, 常宜峰, 唐江波. 改进的修正预测法预报电离层[J]. 测绘科学技术学报, 2011, 28(1): 19-22.
- [6] 梁子超, 李智炜, 赖铿, 等. 10 折交叉验证用于预测模型泛化能力评价及其 R 软件实现[J]. 中国医院统计, 2020, 27(4): 289-292.
- [7] 路佳佳. 基于交叉验证的集成学习误差分析[J]. 计算机系统应用, 2023, 32(1): 302-309.