

尼日利亚猴痘疫情不确定逻辑增长模型分析

陈丽丽,丁春晓*

(聊城大学数学科学学院,山东聊城252000)

摘要:猴痘疫情受到各种各样不确定因素的影响,为了处理猴痘传播过程中的不确定性,本文尝试利用不确定逻辑增长模型对尼日利亚严峻的猴痘疫情进行分析。首先,对尼日利亚疾控中心公布的猴痘累计确诊病例数据进行整理,并画出数据的散点图;其次,建立不确定逻辑增长模型,并对模型进行参数估计、残差分析和假设检验;最后,给出尼日利亚猴痘累计确诊病例数的预测值,并得出95%的置信区间。此外,通过交叉验证的方法以及对观测值、预测值和置信区间的直观对比验证模型的可行性。本研究为尼日利亚猴痘疫情防控和政策制定提供一定的理论支撑。

关键词:不确定性理论;不确定回归分析;不确定逻辑增长模型;尼日利亚猴痘疫情

中图分类号:O29 **文献标志码:**A

引用格式:陈丽丽,丁春晓.尼日利亚猴痘疫情不确定逻辑增长模型分析[J].山东大学学报(理学版),2025,60(5):116-121.

Analysis of Mpox outbreak in Nigeria based on uncertain Logistic growth model

CHEN Lili, DING Chunxiao*

(School of Mathematical Sciences, Liaocheng University, Liaocheng 252000, Shandong, China)

Abstract: The Mpox epidemic is affected by various uncertain factors. In order to deal with the uncertainty in the process of Mpox transmission, the uncertain Logistic growth model is used to analyze and forecast the the severa Mpox epidemic in Nigeria, where the Mpox epidemic is severe. Firstly, the cumulative confirmed cases of Mpox published by the Nigeria Center for Disease Control and Prevention are sorted out, and the scatter plot of the data is illustrated. Next, an uncertain Logistic growth model is established. Also, parameter estimation, residual analysis and hypothesis test are carried out for the above model. Finally, the forecast values of the cumulative number of confirmed Mpox cases in Nigeria are given, and 95% confidence intervals are obtained. Besides, the feasibility of the model is verified by cross-validation method and the direct comparison among observed values, forecast values and confidence intervals. This study provides some theoretical support for the prevention and control of Mpox epidemic and policy formulation in Nigeria.

Key words: uncertainty theory; uncertain regression analysis; uncertain Logistic growth model; Mpox outbreak in Nigeria

0 引言

猴痘是猴痘病毒感染所致的一种人畜共患病^[1],其主要症状包括发热、皮疹、淋巴肿大等。猴痘病毒于1970年在一名刚果儿童身上被首次发现,此后,非洲等地不断有猴痘病例出现,大多集中在非洲中部和西部的热带雨林地区,传播链比较短,具有地方流行性特征。而2022年爆发的猴痘疫情,规模空前盛大,涉及了很多从未有过猴痘病例的国家,猴痘病毒已不容忽视。

2022年中旬,猴痘疫情的突然暴发,无疑给各国带来了一次的挑战。猴痘疫情尤其对全球旅游、进出口

收稿日期:2023-09-25;网络出版时间:2024-10-08 12:55:21

基金项目:国家自然科学基金资助项目(12026226,12461100);山东省自然科学基金资助项目(ZR2021MA089,ZR2022MF293)

第一作者:陈丽丽(2001—),女,硕士研究生,研究方向为不确定理论和生物数学. E-mail:1784957151@qq.com

*通信作者:丁春晓(1987—),女,副教授,硕士生导师,博士,研究方向为不确定理论和生物数学. E-mail:dingchunxiao1987@163.com

贸易以及文化交流等方面影响较大。了解猴痘的流行病学特征,分析总结疫情的发展趋势,对于充分了解流行病的传播,做好应急预案具有重要意义^[2]。

专家学者们采用逻辑增长模型进行疫情预测分析。逻辑增长模型最初由比利时数学家 Verhulst 于 1838 年推导出来,后被广泛应用于动物饲养、植物栽培等资源生态环境方面的模拟研究^[3]。逻辑增长模型的特点表现为初期增长缓慢,随后逐渐加速,最终趋于平缓并接近饱和的“S”型增长趋势,因此逻辑增长曲线又被称为“S”型曲线。猴痘疫情传播过程中受到各种各样的不确定因素的影响,因此专家学者将这些不确定因素视为随机变量^[4],采用随机逻辑增长模型进行分析预测^[5]。随机逻辑增长模型作为一种概率论模型,使用前提是假设残差频率稳定,而随着研究的深入发现,残差频率总是不稳定的,本文使用双样本柯尔莫可洛夫-斯米洛夫检验(two-sample Kolmogorov-Smirnov test)、威尔科克森秩和检验(Wilconxon rank sum test)、安萨里-布拉德利检验(Ansari-Bradley test)、双样本 T 检验(two-sample T-test)、双样本 F 检验(two-sample F-test)5 种方法,验证残差不是来自同一分布,即频率不稳定。此时继续运用随机逻辑增长模型分析尼日利亚猴痘疫情是不合理的,因此,本文尝试使用不确定理论对尼日利亚猴痘疫情的累计确诊病例人数进行建模分析。

不确定理论由清华大学刘宝碇教授于 2007 年建立^[6],2009 年进一步完善使其成为公理化体系^[7]。不确定理论满足正则性、对偶性、次可加性及乘积公理^[8]。不确定统计学是不确定理论的重要分支之一,由刘宝碇教授于 2010 年提出,是一套利用不确定理论来收集、分析和解释数据的数学技术^[9]。不确定回归分析是一种利用不确定理论来探索解释变量与响应变量之间关系的统计技术^[10]。它的切入点是把扰动项看作成一个不确定变量,为了估计不确定扰动项,提出了矩估计^[11]、极大似然估计^[12-13]和最小二乘估计^[14]等多种参数估计方法。此外,利用不确定假设检验来评估不确定回归模型的合理性^[15],并对回归系数提出不确定显著性检验^[10]。迄今为止,不确定回归分析已成功应用于经济学^[16-17]、疫情传播^[18]和粮食产量^[19]等领域。本文针对尼日利亚猴痘累计确诊病例建立不确定逻辑增长模型,通过最小二乘法求解模型中未知参数,利用不确定假设检验验证模型的拟合效果,最后对疫情进行预测并得到相应的置信区间。

1 尼日利亚猴痘疫情

尼日利亚疾控中心对当地 2022 年 1 月至 2023 年 1 月猴痘累计确诊病例数据进行了统计(<https://ncdc.gov.ng/news/item/348/>),本文将其数据作图,见图 1。通过对尼日利亚猴痘累计确诊病例的观察与分析,发现猴痘疫情具有逻辑增长的趋势,因此采用逻辑增长模型对数据进行建模分析。设观测数据为

$$(t, y_t), t = 1, 2, \dots, 37, \tag{1}$$

其中, y_t 是图 1 中所示的 t 时的尼日利亚猴痘累计确诊病例数,如 $y_5 = 31$ 指 2022 年 6 月 5 日的累计确诊病例人数为 31, $y_{37} = 786$ 指 2023 年 1 月 22 日的累计确诊病例人数为 786。

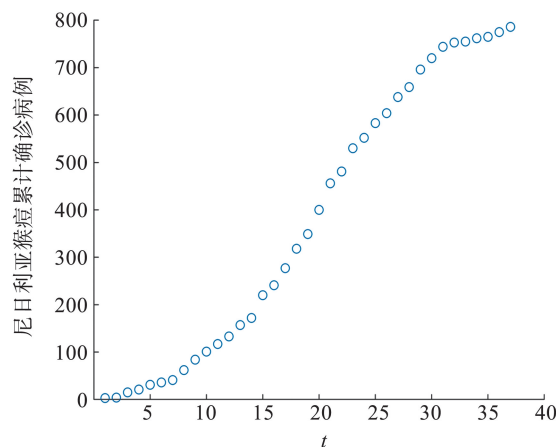


图 1 2022 年 1 月至 2023 年 1 月尼日利亚猴痘累计确诊病例数据

Fig.1 Cumulative confirmed cases of Mpox in Nigeria from January 2022 to January 2023

2 随机逻辑增长模型

采用随机逻辑增长模型对尼日利亚猴痘累计确诊病例进行建模分析,

$$y = \beta_0 / (1 + \beta_1 \exp(-\beta_2 t)) + \varepsilon, \quad \beta_0, \beta_1, \beta_2 > 0, \quad (2)$$

其中, β_0 表示模型的上限, 即疫情在不受外部干预(如疫苗等)且环境条件保持不变的情况下, 累计确诊病例可能达到的最大值, β_1 影响模型的生长速度, 其大小决定了达到模型上限 β_0 的速度快慢, β_2 影响曲线的斜率, β_2 越大, 增长曲线越陡峭, 疾病传播速度越快, ε 是随机扰动项(随机变量)。

利用观测值与预测值作差得到如图 2 所示的 37 个残差 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{37}$ 。为了验证残差是否频率稳定, 采用了 5 种方法进行检验。其中, 双样本 F 检验(two-sample F-test)以及双样本 T 检验(two-sample T-test)属于参数检验方法, 均满足检验的前提条件。在利用 MATLAB 软件对残差进行检验时, 返回值 p 是观察到的数据(或两个数据集之间的差异)与参考分布(或彼此间)一样或更极端的情况出现的概率。如果 p 值小于显著性水平(0.05), 则拒绝原假设, 认为观察到的数据集与参考分布不符(或两个数据集不是来自同一分布)。5 种检验方法的返回值 p 均小于 0.05, 具体细节如下:

(1) 双样本柯尔莫可洛夫-斯米洛夫检验(two-sample Kolmogorov-Smirnov test)

取 37 个残差中的两部分

$$(\varepsilon_1, \varepsilon_3, \dots, \varepsilon_7) \text{ 和 } (\varepsilon_{14}, \varepsilon_{15}, \dots, \varepsilon_{37}),$$

借助 MATLAB 中“kstest2”函数, 通过 two-sample Kolmogorov-Smirnov test 得到 p 值为 0.003 9。这意味着, 上述残差组具有不同的分布函数, 因此不是来自同一个总体。

(2) 威尔科克森秩和检验(Wilconxon rank sum test)

取 37 个残差中的两部分

$$(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_8) \text{ 和 } (\varepsilon_9, \varepsilon_{10}, \dots, \varepsilon_{37}),$$

借助 MATLAB 中“ranksum”函数, 通过 Wilconxon rank sum test 得到 p 值为 0.003 8。这意味着, 上述残差组来自于具有不同中位数的分布, 因此并不是来自于同一个总体。

(3) 安萨里-布拉德利检验(Ansari-Bradley test)

取 37 个残差中的两部分

$$(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{15}) \text{ 和 } (\varepsilon_{33}, \varepsilon_{34}, \dots, \varepsilon_{37}),$$

借助 MATLAB 中“ansaribradley”函数, 通过 Ansari-Bradley test 得到 p 值为 0.005 2。这意味着, 上述残差组具有不同的分散性, 因此并不来自同一个总体。

(4) 双样本 T 检验(two-sample T-test)

取 37 个残差中的两部分

$$(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_8) \text{ 和 } (\varepsilon_9, \varepsilon_{10}, \dots, \varepsilon_{37}),$$

借助 MATLAB 中“ttest2”函数, 通过 two sample T-test 得到 p 值为 0.003 0。这意味着, 上述残差组在满足正态分布和方差齐性假设的前提下, 具有不同的均值, 因此并不来自同一个总体。

(5) 双样本 F 检验(two-sample F-test)

取 37 个残差中的两部分

$$(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_8) \text{ 和 } (\varepsilon_{20}, \varepsilon_{21}, \dots, \varepsilon_{32}),$$

借助 MATLAB 中“vartest2”函数, 通过 two-sample F-test 得到 p 值为 0.034 7。这意味着, 上述残差组在满足正态分布假设的前提下, 具有不同的方差, 因此并不来自同一个总体。

由于 37 个残差不是来自同一总体, 即残差不是来自同一分布, 频率不稳定, 所以将扰动项视为随机变量是不合理的。本文尝试采用不确定变量来刻画该扰动项, 采用不确定逻辑增长模型对尼日利亚猴痘累计确诊病例进行建模。

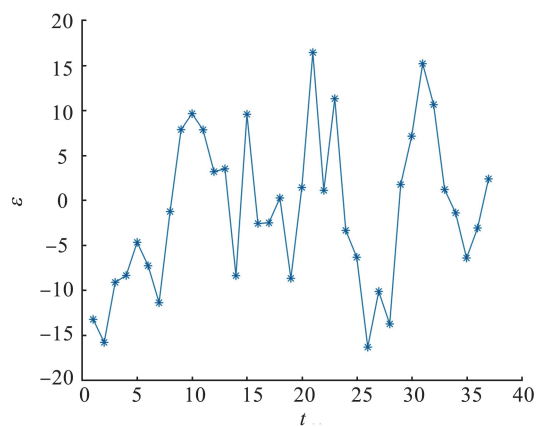


图 2 逻辑增长模型残差图

Fig.2 Residual plot of Logistic growth model

3 不确定逻辑增长模型

为了拟合观测数据,采用不确定逻辑增长模型

$$y = \beta_0 / (1 + \beta_1 \exp(-\beta_2 t)) + \varepsilon, \quad \beta_0, \beta_1, \beta_2 > 0, \tag{3}$$

其中, ε 是不确定扰动项(不确定变量)。

基于观测数据(1),利用最小二乘估计求解不确定逻辑增长模型(3)中未知参数($\beta_0, \beta_1, \beta_2$),即

$$\min_{\beta_0, \beta_1, \beta_2 > 0} \sum_{t=1}^{37} (y_t - \beta_0 / (1 + \beta_1 \exp(-\beta_2 t)))^2. \tag{4}$$

利用 MATLAB 中“lsqnonlin”函数求解上述最小化问题,得到拟合参数值为

$$\beta_0 = 809.098\ 4, \quad \beta_1 = 59.955\ 5, \quad \beta_2 = 0.203\ 2.$$

因此,拟合逻辑增长模型为

$$y = 809.098\ 4 / (1 + 59.955\ 5 \exp(-0.203\ 2t)),$$

如图 3 蓝色实线所示。

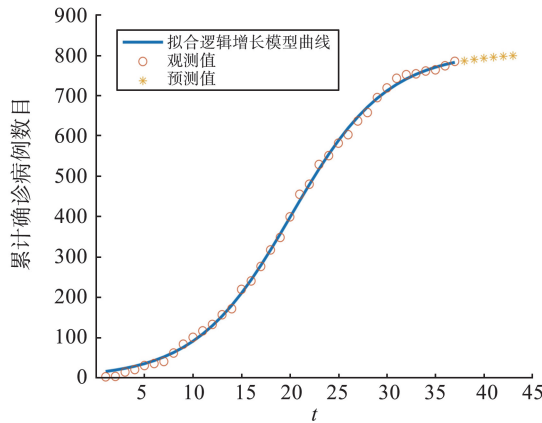


图 3 尼日利亚累计确诊病例数的观测值、拟合逻辑增长模型及预测值的对比图
Fig.3 Comparison of observed, fitted logistic growth model and forecasts of the cumulative number of confirmed cases in Nigeria

将观测值与预测值作差,即

$$\varepsilon_t = y_t - 809.098\ 4 / (1 + 59.955\ 5 \exp(-0.203\ 2t)),$$

得到了 37 个残差 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{37}$ 。假设残差服从正态不确定分布 $\mathcal{N}(e, \sigma)$, 不确定扰动项 ε 的期望值为

$$\hat{e} = \frac{1}{37} \sum_{t=1}^{37} \varepsilon_t = -1.146\ 2, \tag{6}$$

方差是

$$\hat{\sigma}^2 = \frac{1}{37} \sum_{t=1}^{37} (\varepsilon_t - \hat{e})^2 = 8.522\ 0^2. \tag{7}$$

因此,尼日利亚猴痘累计确诊病例数的不确定逻辑增长模型为

$$y = 809.098\ 4 / (1 + 59.955\ 5 \exp(-0.203\ 2t)) + \mathcal{N}(-1.146\ 2, 8.522\ 0). \tag{8}$$

使用不确定假设检验判断不确定逻辑增长模型(8)是否很好地拟合图 1 所示的尼日利亚猴痘累计确诊病例数,即检验 37 个残差 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{37}$ 是否服从正态不确定分布:

$$\mathcal{N}(-1.146\ 2, 8.522\ 0).$$

为此,考虑以下两个假设

$$H_1: e = -1.146\ 2 \text{ 且 } \sigma = 8.522\ 0 \text{ 与 } H_2: e \neq -1.146\ 2 \text{ 或 } \sigma \neq 8.522\ 0.$$

取显著性水平 $\alpha = 0.05$, 根据不确定假设检验,可以得出该检验为

$$W = \{(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{37}) : \text{至少存在 2 个下标 } t \text{ 满足 } 1 \leq t \leq 37 \text{ 使得 } \varepsilon_t < -18.359\ 2 \text{ 或 } \varepsilon_t > 16.066\ 7\}.$$

观察图 4, 只有 $\varepsilon_{21} \notin [-18.359\ 2, 16.066\ 7]$ 时, $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{37}) \notin W$ 。因此接受 H_0 , 即不确定逻辑增长模型可

以很好地拟合图 1 的尼日利亚累计确诊病例数目。

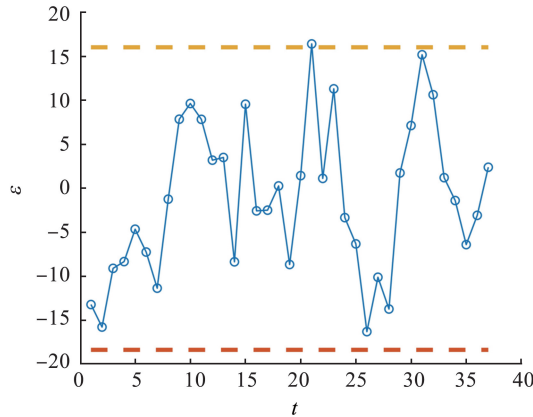


图 4 不确定逻辑增长模型的残差图

Fig.4 Residual plot of the uncertain Logical growth model

利用不确定逻辑增长模型(8)预测 2023 年 1 月 29 日尼日利亚猴痘累计确诊病例数。取 $t=38$, 不确定变量 \hat{y}_{38} 为

$$\hat{y}_{38} = 809.0984 / (1 + 59.9555 \exp(-0.2032 \times 38)) + \mathcal{N}(-1.1462, 8.5220), \tag{9}$$

即 $\hat{y}_{38} \sim \mathcal{N}(787.0068, 8.5220)$ 。将 \hat{y}_{38} 的期望值作为预测值, 计算得到预测值为 $E[\hat{y}_{38}] = 787.0068$ 。因此预计 2023 年 1 月 29 日尼日利亚猴痘累计确诊病例为 787 人。进一步得到 95% 的置信区间为

$$787.0068 \pm \frac{8.5220 \sqrt{3}}{\pi} \ln \frac{1+0.95}{1-0.95},$$

即 $787.0068 \pm 17.2130 = [769.7938, 804.2197] \approx [770, 804]$, 2023 年 1 月 29 日尼日利亚猴痘累计确诊病例人数 95% 的可能性介于 770~804 人之间。同理可以得到 2023 年 2 月 5 日到 3 月 5 日的预测值和置信区间, 详见图 3 黄色星号部分和表 1 数值。

为了对模型可行性进行评估, 借助 MATLAB 的“cvpartition”函数, 采用 5 折交叉验证的方法, 得到模型的平均绝对误差为 7.0319, 均方根误差为 8.2619。尼日利亚猴痘累计确诊病例数据的范围为 3~789, 在这个数据范围下, 上述误差较小, 说明模型性能较好。此外, 比较表 1 的观测值、预测值及置信区间, 可以发现观测值均位于置信区间内, 从而进一步说明了模型的可行性。

表 1 1 月 29 日到 3 月 5 日的预测值和置信区间

Table 1 Forecast values and confidence intervals from January 29 to March 5

日期	观测值	预测值	置信区间
1 月 29 日	789	787	[770, 804]
2 月 5 日	796	791	[774, 808]
2 月 12 日	800	794	[777, 811]
2 月 19 日	805	796	[779, 814]
2 月 26 日	807	799	[781, 816]
3 月 5 日	812	800	[783, 817]

4 结论

本文采用不确定回归分析对尼日利亚猴痘累计确诊病例进行分析研究。通过交叉验证以及直接对比观测值和预测值的方法, 说明估计的不确定逻辑增长模型可以很好地拟合尼日利亚猴痘累计确诊病例数目。进一步对确诊病例人数进行预测并给出置信区间。另外, 通过分析相关残差的特征, 本文给出使用不确定统计量代替概率统计量的原因。

未来, 还将结合最新数据, 对猴痘传播风险、峰值、达峰时间、传染规模等疫情技术指标进行预测分析, 并对模型进行进一步优化。

参考文献:

- [1] KUMAR N, ACHARYA A, GENDELMAN H E, et al. The 2022 outbreak and the pathobiology of the monkeypox virus[J]. *Journal of autoimmunity*, 2022, 131:102855.
- [2] 张钰倩,张太雷. 具有复发效应的 SEAIR 模型及在新冠肺炎传染病中的应用[J]. *山东大学学报(理学版)*, 2022, 57(1):56-68.
ZHANG Yuqian, ZHANG Tailei. An SEAIR model with relapse effect and its application in COVID-19 transmission[J]. *Journal of Shandong University (Natural Science)*, 2022, 57(1):56-68.
- [3] 李春喜,姜丽娜,邵云等. 生物统计学[M]. 北京:科学出版社, 2013:163-166.
LI Chunxi, JIANG Lina, SHAO Yun, et al. *Biostatistics*[M]. Beijing: Science Press, 2013:163-166.
- [4] 赵军圣,赵雪静,孙宗耀. 随机非线性系统的有限时间命令滤波输出反馈跟踪控制[J]. *聊城大学学报(自然科学版)*, 2023, 36(6):9-17.
ZHAO Junsheng, ZHAO Xuejing, SUN Zongyao. Finite-time command filtering output feedback tracking control of stochastic nonlinear systems[J]. *Journal of Liaocheng University (Natural Science)*, 2023, 36(6):9-17.
- [5] DRAPER N R, SMITH H. *Applied regression analysis*[M]. John Wiley & Sons, 1998:135-148.
- [6] LIU B D. *Uncertainty Theory*[M]. 2nd ed. Berlin: Springer-Verlag, 2007.
- [7] LIU B D. Some research problems in uncertainty theory[J]. *Journal of Uncertain Systems*, 2009, 3(1):3-10.
- [8] 温艳清,刘宝亮,师海燕,等. 基于不确定理论的多部件竞争失效系统的可靠性评估[J]. *山东大学学报(理学版)*, 2023, 58(5):76-83.
WEN Yanqing, LIU Baoliang, SHI Haiyan, et al. Reliability evaluation for multi-competing failure system based on uncertainty theory[J]. *Journal of Shandong University (Natural Science)*, 2023, 58(5):76-83.
- [9] LIU B D. *Uncertainty theory: A branch of mathematics for modelling human uncertainty*[M]. Berlin: Springer-Verlag, 2010.
- [10] YAO K, LIU B D. Uncertain regression analysis: an approach for imprecise observations[J]. *Soft Computing*, 2018(22):5579-5582.
- [11] LIO W C, LIU B D. Residual and confidence interval for uncertain regression model with imprecise observations[J]. *Journal of Intelligent & Fuzzy Systems*, 2018, 35(2):2573-2583.
- [12] LIO W C, LIU B D. Uncertain maximum likelihood estimation with application to uncertain regression analysis[J]. *Soft Computing*, 2020(24):9351-9360.
- [13] LIU Y, LIU B D. A modified uncertain maximum likelihood estimation with applications in uncertain statistics [J]. *Communications in Statistics - Theory and Methods*, 2023, 53(18):6649-6670.
- [14] LIU Y, LIU B D. Estimation of uncertainty distribution function by the principle of least squares[J]. *Communications in Statistics-Theory and Methods*, 2023, 53(21):7624-7641.
- [15] YE T Q, LIU B D. Uncertain hypothesis test with application to uncertain regression analysis[J]. *Fuzzy Optimization and Decision Making*, 2022, 21(2):157-174.
- [16] YE T Q, LIU B D. Uncertain significance test for regression coefficients with application to regional economic analysis[J]. *Communications in Statistics-Theory and Methods*, 2023, 52(20):7271-7288.
- [17] 朱佳璇,刘利英. 基于不确定时间序列的外汇储量智能预测[J/OL]. *聊城大学学报(自然科学版)*, 2024. <https://doi.org/10.19728/j.issn1672-6634.2024010010>.
ZHU Jiaxuan, LIU Liying. Intelligent prediction for foreign exchange reserves based on uncertain time series[J]. *Journal of Liaocheng University (Natural Science)*, 2024. <https://doi.org/10.19728/j.issn1672-6634.2024010010>.
- [18] LIU Z. Uncertain growth model for the cumulative number of COVID-19 infections in China[J]. *Fuzzy Optimization and Decision Making*, 2021, 20(2):229-242.
- [19] LIU Y. Moment estimation for uncertain regression model with application to factors analysis of grain yield [J]. *Communications in Statistics-Simulation and Computation*, 2022, 53(10):4936-4946.

(编辑:胡春燕)