

基于时间序列分析韩国重大公共事件及预测

李素蕊,王玲书

(河北经贸大学,石家庄 050061)

摘 要 2019 年 12 月突发一起重大公共事件,新型冠状病毒以惊人速度传播开来。此传染病已成为全球性公共卫生问题,世界各国不仅要了解本地情况,还应对其他国家的扩散情况进行掌握。该文选取 2022 年 7 月 1 日至 2022 年 9 月 30 日的韩国累计确诊数据,建立了 ARIMA(0,2,2)模型,并对 2022 年 10 月 1 日至 2022 年 10 月 5 日韩国累计确诊人数进行预测。结果表明,利用该模型预测的结果与实际结果基本吻合。该文结果对此次重大公共事件的预测和防控提供理论支持。

关键词 重大公共事件;时间序列分析;ARIMA 模型;发展现状;预测

中图分类号:R563.1

文献标志码:A

文章编号:2095-2945(2023)11-0006-04

Abstract: A major public incident broke out in December 2019, and Covid-19 virus spread at an alarming rate. This infectious disease has become a global public health problem. Countries around the world should not only understand the local situation, but also grasp the spread of other countries. In this paper, based on the cumulative confirmed data of Korea from July 1 to September 30, 2022, an ARIMA model is established, and the cumulative number of confirmed cases in Korea from October 1 to October 5, 2022 is predicted. The results show that the predicted results are basically consistent with the actual results. The results of this paper provide theoretical support for the prediction, prevention and control of this major public event.

Keywords: major public events; time series analysis; ARIMA model; development status; prediction

2019 年 12 月,正当中国人民喜迎新春之际,湖北省武汉市接连查出多起新型冠状病毒感染(以下简称“新冠感染”)病例,该病毒具有传染速度快且发病率高的特性。随后,在中国对此开展防控治疗时,其他各国也陆续出现确诊案例。该传染病迅速演变成全球范围内的公共卫生问题,新型冠状病毒感染在各国的扩散阶段不同,扩散趋势也有显著不同^[1]。因此,世界各国不仅要研究本国当前情况,还要抓紧了解其他地区的情况,实现信息共享,对病毒传播情况有更加全面地了解才能在预防与控制中占据上风。

本文采用模型分析法,运用 RStudio 对韩国累计确诊数据进行时间序列分析。首先根据数据绘制时序图、自相关图,以此判断数据的平稳性。由于获取的韩国累计确诊数据具有非平稳性的特点,对非平稳序列^[2]进行差分处理,序列平稳后选取合适的 ARIMA 模型进行拟合,并检验模型的有效性,检验通过后进行趋势预测、趋势图绘制,从而了解国外在此次重大公共卫生事件中病毒扩散的趋势。我国便能及时掌握国外新冠感

染发展程度,应对将来可能出现的危机与挑战。

1 方法

1.1 数据来源

韩国病例数据来源是新型冠状病毒感染全球疫情实时播报,从中选取出 2022 年 7 月 1 日至 2022 年 9 月 30 日韩国的累计确诊数据。

1.2 研究方法

时间序列分析是将一组不同时间上的数据按照时间先后顺序记录下来,通过研究序列的统计规律来分析生活中的难题,也可将这些规律与信息用于预测。根据序列的统计特征可分为平稳时间序列和非平稳时间序列。在进行时间序列分析时,简便的 ARIMA 模型总是备受青睐,其也是最常见的使用方法之一。ARIMA 模型简单,只依靠内生的变量,但需注意的是该模型依赖平稳序列,当序列为非平稳时无法提取到规律,需进行差分使其平稳化。ARIMA 模型相当于 AR 模型与 MA 模型的组合,包括 p 、 d 、 q 3 个参数。 p 是构造的模型中使用的时序数据自身的滞后数, d 是差分阶数, q 是

第一作者简介:李素蕊(1999-),女,硕士研究生。研究方向为应用统计学。

构造的模型中使用的预测误差的滞后数,进一步可看出该模型无须依赖其他的外生变量,当 q 为 0 时,ARIMA 模型变为 AR 模型 p 为 0 时,ARIMA 模型变为 MA 模型^[3]。ARIMA(p d q)模型结构如下

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, Var = \sigma^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases}$$

式中: $\nabla^d = (1-B)^d$, $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ 代表自回归系数多项式, $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ 代表移动平滑系数多项式, x_t 代表非平稳序列, ε_t 代表零均值白噪声序列。

1.3 建模过程

1)数据平稳化 根据时序图观察序列的平稳性,当序列呈现周期性变动或有剧烈波动趋势时,为非平稳序列。随后用自相关图做进一步辅助识别,平稳序列的自相关系数呈现迅速衰减趋势,非平稳序列中自相关系数则是逐渐降到 0,在 0 处波动一段时间后,自相关系数又变为负。当序列非平稳时做差分处理,使其平稳化^[4]。

2)白噪声检验 对经过差分后得到的平稳序列进行纯随机性检验,在各阶延迟下 p 值小于 $\alpha(0.05)$,则认为序列是非纯随机的,包含着可提取的动态规律^[5]。

3)通过自相关图和偏自相关图找出合适模型,当存在多个合适的模型时,由 AIC 准则判定最合适的模型, AIC 的值越小,模型拟合越好。

4)模型检验 模型检验包括 2 部分。第一部分是模型的显著性检验,查看模型筛选的信息是否充分^[6]。拟合效果越好,序列中样本的相关信息获取越充分,此时残差序列项为纯随机序列;第二部分是参数的显著性检验,用于检验模型中每一个参数是否显著非 0,不显著非 0 的参数视为不合格,可以舍弃^[7]。

5)模型预测 模型通过检验后展开预测。

2 对重大公共事件下韩国现状分析及预测

2.1 平稳性检验

图 1 是韩国累计确诊序列的时序图,从图中可发现明显的上升趋势,呈现出非平稳序列的特征,通过自相关图进一步辅助识别,如图 2 所示。累计确诊序列在很长的延迟时间里,自相关系数从正数递减到 0 后,又持续是负数,因此序列是非平稳序列^[1]。图 3 中序列经二阶差分后一直在一个固定的范围内随机波动,变得平稳。接下来对数据进行纯随机性检验,确认该序列是

否有继续研究的价值。

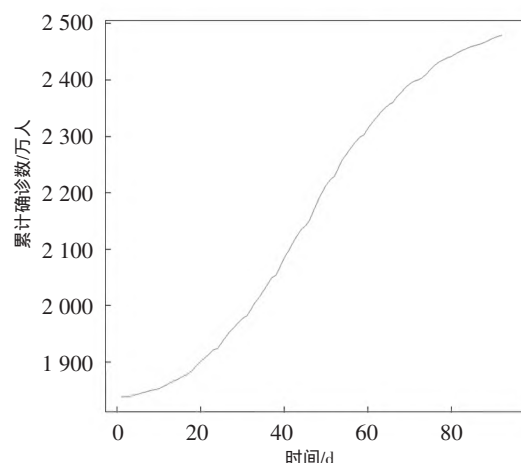


图 1 韩国 2022 年 7 月 1 日—2022 年 9 月 30 日累计确诊时序

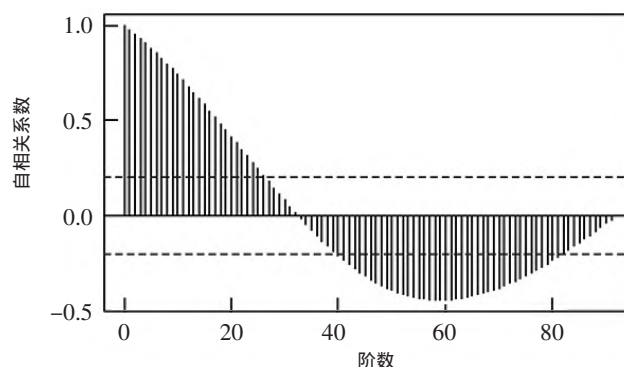


图 2 韩国累计确诊序列自相关图

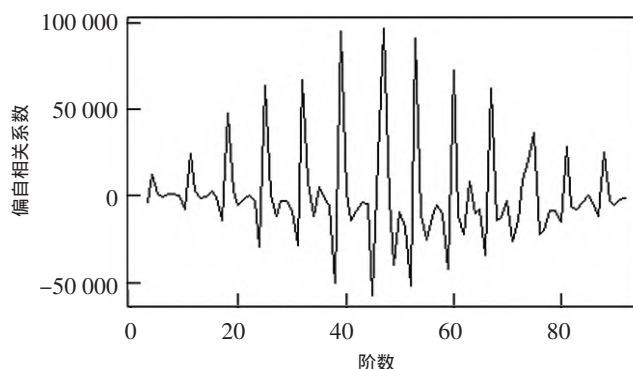


图 3 累计确诊二阶差分图

2.2 白噪声检验

通过 RStudio 对 3 个观察值序列进行白噪声检验,检验结果显示,在 6 阶、12 阶延迟下 p 值均小于显著性水平 $\alpha(0.05)$,有很大把握认为韩国累计治愈序列是非纯随机的,即为非白噪声序列,存在研究价值。

2.3 建立模型

图 4、图 5 是差分处理后的自相关图和偏自相关图。由图 4、5 可知,自相关系数 2 阶截尾,偏自相关系数衰减过程非常缓慢,偏自相关系数拖尾。因此确定

ARIMA(p 2 q)模型中参数 p 为 0 q 为 2 ,因此选取 ARIMA(0 2 2)为最优模型 ,ARIMA(0 2 2)的 $AIC=2$ 083.9。拟合模型的口径为

$$x_t = \varepsilon_t - 0.3965\varepsilon_{t-1} - 0.2761\varepsilon_{t-2},$$

$$Var(\varepsilon_t) = 619333966。$$

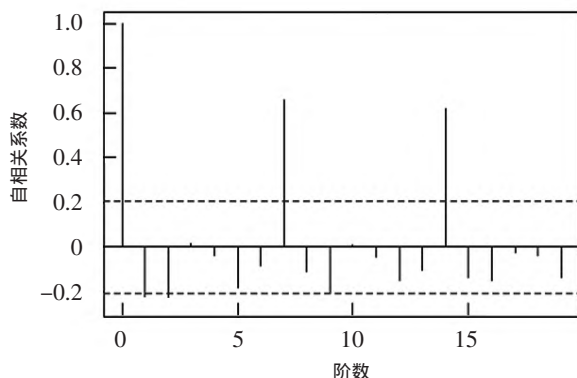


图 4 二阶差分的自相关图

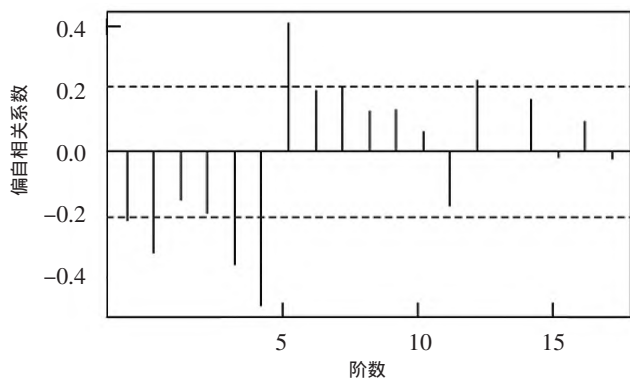


图 5 二阶差分的偏自相关图

2.4 模型检验

进行模型检验 结果见表 1。6 阶延迟下模型残差通过了检验 ,可认为 ARIMA(0 2 2)模型拟合恰当。从表 2 可以看出 ,韩国累计确诊序列中 2 个参数均显著非 0 ,ARIMA(0 2 2)模型 2 个系数都通过了 t 检验 ,拟合效果很好。

2.5 模型预测

本文通过 RStudio 对韩国此次重大公共事件中的累计确诊数据进行时间序列分析 ,展开预测。图 6 是累计确诊序列的趋势预测图 ,预测了 2022 年 10 月 1 日后的发展趋势 ,在 2022 年 10 月 1 日到 2022 年 10 月 5 日韩国的累计确诊序列呈现平稳上升趋势 ,确诊人数的增速并未放缓 ,在未来一段时间内感染者人数仍会继续上升 ,事态发展与防控的形势不容乐观。韩国此次事件的累计确诊序列预测结果与实际结果见表 3 ,两者对比进一步展现出模型预测有效程度 ,预测结果接近真实值 ,保障了预测的准确性。

表 1 ARIMA(0,2,2)模型白噪声检验表

卡方	自由度	P 值
9.6320	6	0.9367
39.473	12	<0.0500

表 2 参数估计结果

参数	系数值	标准差	t 值	P 值
MA(1)	-0.3965	0.1067	3.7160	<0.05
MA(2)	-0.2761	0.0971	2.8435	<0.05

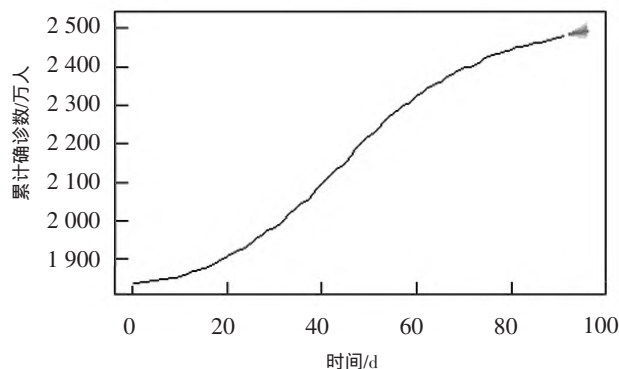


图 6 韩国累计确诊序列预测图

表 3 韩国预测与实际对照表

日期	估计值	真实值
10 月 1 日	24825107	24819611
10 月 2 日	24855088	24831761
10 月 3 日	24885069	24848184
10 月 4 日	24915050	24882894
10 月 5 日	24945031	24911497

3 结束语

自 2019 年检测出第一个确诊病例后 ,大部分国家迅速被笼罩在新冠感染的阴霾之下 ,这场传染病夺去了无数人的生命 ,造成的经济损失也不可估量 ,深入研究并尽快解决此次重大公共卫生传染病问题成为迫切要求。本文从新冠感染全球疫情实时播报中选取 2022 年 7 月 1 日至 2022 年 9 月 30 日韩国累计确诊数据 ,建立了 ARIMA(0 2 2)模型 ,该模型对序列进行拟合得到了显著的效果。最后 ,利用模型预测 2022 年 10 月 1 日到 10 月 5 日韩国累计确诊人数 ,韩国确诊人数依然持续上升 ,事态控制效果不明显 ,传染病发展和防控形势严峻。预测结果与真实值吻合 ,预测的韩国累计确诊病例数据总体呈现上升趋势 ,与真实数据的整体变化趋势也大致相同 ,保障了预测力度 ,而随时间推移 ,预测力度呈递减趋势。通过对韩国现阶段的形势进行

(下转 12 页)

表 1 回采方案优缺点对比表			
内容	方案一 干式回采	方案二 湿式回采	方案三 干式+湿式联合开采
回采效率	稍高	较低	较高
安全性	机械设备运输安全性较高，组织得当安全性有保障	专业化管理要求标准高，仅采用该单一方案易造成坝体浸润线提升，最终使坝体安全性降低	组织安全可靠，浸润线高度可根据实际工况进行控制
调整能力	具有灵活的调整能力，能针对不同情况采取不同的应对措施	设备运行过程中不具有可调整性	湿式回采设备较少，可灵活调整，针对不同工况的适应性较强
经济性	一般	一般	高
优点	经济性一般，但效率较可靠、具有较强的可调整性	运输沿线无需管理，无需开挖道路	经济性好且效率高，可调整性高，安全性高
不足	大范围的地基整平，项目费用高	工艺复杂，经济性不高	对管理要求高
比选结果	不推荐	不推荐	推荐

4 结束语

尾矿库尾砂回采施工作业能够有效降低甚至清除尾矿库闭库后的潜在不安全因素，尾矿库尾砂回采可以保持生态平衡、减少尾矿量及提高尾砂综合利用率。不仅如此，该工程还能在一定程度上产生较高的经济收益及社会效益。为保障作业过程中的库坝安全及人身安全，本研究以某尾矿库实体工程为例，对常见的尾砂回采方案进行分析，根据各自的优缺点最终选择了最优方案。同时，研究中还对各回采方案的技术特点进行了具体的说明与分析，本研究一方面可为将来同类型的工程施工提供有益的借鉴，另一方面也可为设计人员在今后的设计工作中提供一定的参考与经验。

参考文献：

[1] 国家矿山安全监察局.“十四五”矿山安全生产规划(征求意见稿)

(上接 8 页)

追踪，可以掌握本次重大公共卫生传染病发展的情况，深入了解该传染病的传播特点，为我国提供有效参考，在本国实际情况基础之上进行精准施策，如制定严格、准确和及时的防控措施等，从而遏制传染病的发展。

参考文献：

[1] 白云, 钱箴, 孙玉莹, 等. 基于综合集成预测方法的新冠肺炎疫情预测[J]. 系统工程理论与实践, 2022, 42(6): 1678-1693.
[2] 裴韬, 王席, 宋辞, 等. COVID-19 疫情时空分析与建模研究进展[J]. 地球信息科学学报, 2021, 23(2): 188-210.

稿)[Z].2021.

[2] 吴涛. 红泥坡铜矿地下开采采矿方法研究[D]. 昆明: 昆明理工大学, 2021.
[3] 岑建, 郑学鑫, 宋会彬. GB 39496—2020《尾矿库安全规程》解读[J]. 劳动保护, 2021(8): 72-73.
[4] 李金梁, 韩西鹏. 某矿山尾矿回采实践[C]//智慧矿山 绿色发展: 第二十六届十省金属学会冶金矿业学术交流会论文集, 2019: 82-84.
[5] 马壮, 钱晨阳, 胡贵生. 尾矿库再开采利用的项目分析与思考[J]. 中国资源综合利用, 2020, 38(4): 99-101.
[6] 郑灿辉, 张殿合, 谭晓飞, 等. 河南栾川某尾矿库回采综合回收白钨矿试验研究与生产实践[J]. 有色金属(选矿部分), 2018(6): 46-50.
[7] 董霄. 某尾矿库尾矿回采技术方案及稳定性分析[J]. 黄金, 2020, 41(7): 44-48.
[3] 王燕. 时间序列分析——基于 R[M]. 北京: 中国人民大学出版社, 2015.
[4] 徐娟年, 马冉, 赵中睿, 等. 山东省新型冠状病毒肺炎疫情流行趋势初步预测[J]. 重庆师范大学学报(自然科学版), 2020, 37(2): 101-106.
[5] ILSU C, HO L D, KUK K Y. Effects of timely control intervention on the spread of middle east respiratory syndrome coronavirus infection [J]. Osong Public Health and Research Perspectives, 2017, 8(6): 373-376.
[6] 黄志辉, 黄波, 汪炯骅. 人口流动对新冠肺炎传播影响的时空分析——以美国为例[J]. 地理与地理信息科学, 2022, 38(5): 65-71.
[7] 万时雨, 刘珏, 刘民. 新型冠状病毒基本再生数研究进展[J]. 科学通报, 2020, 65(22): 2334-2341.