The Impact of the Spread of Major Infectious Diseases on Internet Public Opnion

Zongjing Liang[1,a], Qiuling Li[1,b], Zhijie Li[1,c], Yun Kuang[2,d,*]

[1]*School of Economics and Management, Guangxi Normal University, Guilin, Guangxi, 541001, China*
[2]*Library, Guilin Normal University, Guilin, Guangxi, 541199, China*
[a]*zjliang@mailbox.gxnu.edu.cn,* [b]*2287804357@qq.com,* [c]*812811205@qq.com,*
[d]*kyun@mail.glnc.edu.cn*
*Corresponding author

Abstract: This paper uses the autoregressive distributed lag model (i.e., ARDL model) to study the impact of the COVID-19 pandemic on changes in online public opinion. The research objects of this paper are four types of online public opinion platforms (Weibo, short videos, electronic newspapers or magazines, and online news) and the number of new COVID-19 infections per day during the same period. The research method is the autoregressive distributed lag model (i.e., ARDL model). Research conclusions: (1) There is a long-term equilibrium relationship between the epidemic and public opinion. (2) Public opinion prediction is achieved based on the epidemic, and the prediction accuracy is high. (3) Cross-platform performance comparison of public opinion is achieved. This study can achieve innovations in the theory and practice of public opinion dissemination. The research conclusions can provide real-time public opinion reference materials for the prevention and control of major infectious diseases that may occur again in the future.

1. Introduction

In the past decade, major infectious disease outbreaks caused by global climate change and other reasons have occurred from time to time, such as dengue fever, SARS, monkeypox, and COVID-19. In particular, the COVID-19 outbreak that began in 2019 has become a major infectious disease outbreak in humans in recent years. The occurrence and spread of this epidemic has had a serious impact on the global society, economy, and culture. In order to effectively respond to major infectious disease outbreaks that may occur again in the future and minimize the impact of the spread of the epidemic on humans, in-depth research on the impact of the outbreak and spread of the epidemic on public opinion has important practical significance. Because by studying the impact of the epidemic on public opinion (especially online public opinion), by revealing the mechanism of information's influence on behavior, we can optimize the risk communication management strategies of governments in various countries and provide a social psychological basis for scientific decision-

making. This paper focuses on the impact of the COVID-19 epidemic on online public opinion as the research target, and quantitatively studies its impact mechanism and its inspiration for practice.

1.1. Impact of the epidemic on public opinion

The outbreak of various major infectious diseases can immediately generate responses in online public opinion. According to psychological theory, when people encounter a major infectious disease outbreak, they will have various psychological reactions, which will induce them to search, query, and understand the relevant situation online. With the rise of various new public opinion platforms such as Weibo , short videos , electronic newspapers or magazines , and online news media in the past decade , they have become the main platform for the general public to obtain information and interact with information, thus forming the phenomenon of "attention digital epidemic".

1.2. Limitations of existing research

In order to effectively deal with the phenomenon of "attention digital epidemic", existing research on current online public opinion platforms (Weibo , short videos , electronic newspapers or magazines , online news) mainly focuses on qualitative analysis of a single platform, and there are few research results that conduct quantitative comparisons of multiple platforms, especially lacking research on the causal relationship between the epidemic and the spread of public opinion and its long-term impact. It is very important to conduct objective quantitative research on the long-term relationship between the epidemic and public opinion in order to truly understand social emotions and effectively formulate prevention and control policies.

1.3. Research Questions

Current research on online public opinion often regards different platforms as simple information dissemination and interaction platforms, without in-depth understanding of the essential differences between different platforms. For example, Weibo has the characteristics of real-time information, while short videos emphasize image dissemination. Different media forms have different mechanisms and functions in spreading epidemic information and forming public opinion, there is heterogeneity among media.

1.4. Research objectives

In order to solve the above problems, this paper intends to take the impact of the spread of COVID-19 epidemic on online public opinion as the research object. By constructing a database consisting of epidemic data and public opinion data, according to the basic principles of communication theory and agenda setting theory, the time series statistical model autoregressive distributed lag model (ie, ARDL model) is creatively applied to the study of the impact of the epidemic on public opinion, and quantitatively analyzes the causal and long-term cointegration relationship between the epidemic and four forms of online public opinion (ie, Weibo , short videos , electronic newspapers or magazines , and online news media). This study has important theoretical and practical value. First, it can understand the response differences of different media platforms from a quantitative perspective. Secondly, it can realize public opinion prediction and contribute to the effective monitoring of social public opinion.

2. Literature review

Research on epidemic crisis communication. A review of the development of crisis communication research [1]. Using the "epidemic rumor-busting" function of Sina Weibo mobile client as a sampling source, this paper adopts a mixed research method to examine the production, circulation, and dissemination of epidemic rumors on domestic social media [2]. The extension of news platforms and the expansion of news boundaries have led to a new development trend in the current public crisis communication in terms of content, platform, and boundaries [3].

Research on the heterogeneity of media platforms. Diverse online social platforms provide rich channels for information sharing [4]. An analysis is conducted on the performance of "Beijing News" in WeChat video accounts, with a focus on the content characteristics of traditional media outlets [5]. Analysis of short video reports on current affairs news in mainstream media from the perspective of the audience [6]. From the analysis of the title word frequency and content of the sample, "Today's Minhang" reflects the colour of epidemic prevention during the epidemic [7]. In today's increasingly fierce media competition, news links from the Internet are also used by newspapers [8].

Research on network public opinion model. The evolution behaviour of network crisis public opinion in the crisis communication environment was analyse using the Multi-Agent model method [9]. A network public opinion propagation model based on a double-layer network was proposed, focusing on the impact of public opinion field interaction on the propagation of network public opinion [10]. The study analyses Weibo data based on the social safety valve theory and constructed an early warning model in combination with the EGM grey system theory [11].

Review of current research. Current research on new media platforms on the Internet mainly focuses on qualitative analysis and quantitative analysis of a single platform, and the research methods mainly focus on text analysis and simple statistical analysis. Current research lacks causal quantitative research on the impact of the epidemic on public opinion, and the current public opinion model perspective is single. In view of the shortcomings of current research, this paper constructs an autoregressive distributed lag model (ARDL model) of public opinion, studies the impact response and long-term impact of the epidemic outbreak on different platforms of online public opinion, and realizes the theoretical expansion and practical innovation of public opinion dissemination.

3. Research Methods

This paper studies the impact of the spread of the COVID-19 pandemic on Internet public opinion. First, we extract the public opinion data of Weibo, short videos, electronic newspapers or magazines, and online news with the keyword "COVID-19". Secondly, we collect the daily data of newly confirmed cases in the same period from the official website of the National Health Commission, and use the autoregressive distributed lag model (ARDL) to analyse the impact of the spread of the COVID-19 pandemic on Internet public opinion. The specific research steps are as follows.

3.1. Data Source

The Internet public opinion data with the keyword "COVID-19" collected in this paper comes from the Goonie Internet Public Opinion Monitoring System (http://www. goonie.com. cn/2018/08/14/gooniespider/), and the number of newly confirmed cases comes from the official website of the National Health Commission (http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml). The data collection time range is from January 16, 2020 to April 1, 2020. Among them, the Internet public opinion data on "COVID-19" on four new media platforms, including Weibo, short videos, electronic newspapers or magazines, and online news, were obtained in the Goonie Internet Public Opinion Monitoring System.

3.2. Variable Definition

According to the research purpose of this paper, the ARDL model is proposed to be used for research, in which the corresponding independent variables and dependent variables are set. The specific definitions are shown in Table 1.

Table 1: Variable description.

Variable Types	Variable Name	Variable Symbols
Dependent Variable	Weibo public opinion data	WB
	Short video public opinion data	SP
	Electronic newspaper or magazine public opinion data	PM
	Online news and public opinion data	XW
Independent Variable	New confirmed cases per day	XZ

3.3. Model construction

The autoregressive distributed lag model (ARDL model) was proposed by Pesaran and Shin in 1998 [12]. The ARDL model consists of a regression analysis model with lagged variables and can be considered as an extension of the ordinary regression model. The ARDL model can incorporate Phillips and Hansen's estimation of the cointegration relationship between variables into the regression model, thereby achieving a description of the long-term cointegration relationship between variables [13]. The ARDL model has the following advantages: it is suitable for time series data with a small number of samples, and does not require the independent variable to be of the same order as the dependent variable, as long as the dependent variable or the independent variable is 0th order or 1st order stationary. The calculation steps of the ARDL model include: stationarity test of the original data, model operation, boundary cointegration test, etc. [13].

3.4. Model estimation steps

3.4.1. Unit root test

According to the ARDL model requirements, the time series data of each variable needs to be tested for unit roots. The first condition for building an autoregressive distributed lag model (ARDL) is that the data series is stationary. Therefore, before building the model for analysis, the time series data of each variable needs to be logarithmized and tested for unit roots. The logarithm method is mainly used to reduce or eliminate the heteroscedasticity of the data. Empirical tests show that the logarithmic form of the original data is a stationary data sequence, so it can be considered that the data passes the stationarity test.

3.4.2. Model calculation

After determining that the time series data of each variable is a stationary series, the data of each variable can be used to build a model. The Internet public opinion data of four new media platforms, namely, Weibo, short videos, electronic newspapers or magazines, and online news, are used as dependent variables, and newly confirmed cases are used as independent variables. The autoregressive distributed lag model with a lagged variable is constructed, where the lag order of the model is determined according to the AIC criterion. The ARDL model estimation is performed on the Internet public opinion data of these four new media platforms, and the corresponding models are defined as Model 1, Model 2, Model 3, and Model 4, respectively. The modelling time range of this

study is from January 16, 2020 to March 25, 2020.The model calculation results are shown in Table 2.

Table 2: ARDL model results.

Model Name	Dependent Variable	Independent Variable	ARDL Model	R^2	P-value
Model 1	LNWB	LNXZ	ARDL (4,1)	0.73	0.00
Model 2	LNSP	LNXZ	ARDL (6,9)	0.98	0.00
Model 3	LNPM	LNXZ	ARDL (9,7)	0.88	0.00
Model 4	LNXW	LNXZ	ARDL (3,0)	0.92	0.00

Note: LN in the table means the variables are logarithmically processed.

The following conclusions can be drawn from the model estimation results given in Table 2: (1) The specific form of the ARDL (p, q) model is determined. The p and q lag orders are determined by the AIC criterion. The calculation results show that model 1 is ARDL (4, 1), model 2 is ARDL (6, 9), model 3 is ARDL (9, 7), and model 4 is ARDL (3, 0). (2) The model fits well. R^2 is used to judge the fit of the model. From Table 2, it can be seen that the fit (R^2) of the four models are 0.73, 0.98, 0.88, and 0.92, respectively. Therefore, the fitting effects of these four models are all good, and the P value of the independent variable coefficient is about 0. Therefore, it can be considered that the impact of newly confirmed cases on the four Internet public opinion data of Weibo, short videos, electronic newspapers or magazines, and online news is very significant.

3.4.3. Long-term cointegration relationship

According to the theory of autoregressive distributed lag model, the model estimation results must be subjected to boundary test. Boundary test can test whether there is a cointegration relationship between variables. Boundary test can be judged by the F statistic calculated by the model. The calculation results show that the F statistic values of models 1 to 4 all pass the boundary test at the 1% statistical level corresponding to the critical value. Therefore, it can be considered that there is a cointegration relationship between the Internet public opinion data and the newly confirmed cases on the four platforms of Weibo, short videos, electronic newspapers or magazines, and online news. The long-term cointegration equation is shown in Table 3.

Table 3: Long-term cointegration equation.

Model Name	Dependent Variable	Independent Variable	Long-run cointegration equation
Model 1	LNWB	LNXZ	LNWB=0.34×LNXZ+10.99
Model 2	LNSP	LNXZ	LNSP = 0.6 5 ×LNXZ + 3.32
Model 3	LNPM	LNXZ	LNPM=0.24×LNXZ+5.7 8
Model 4	LNXW	LNXZ	LNXW=0.18 ×LNXZ+10.81

3.4.4. Internet public opinion prediction

Based on the four ARDL models constructed, this paper uses newly confirmed cases to predict the Internet public opinion of each new media platform. The Internet public opinion data from March 26, 2020 to April 1, 2020, a total of 7 days, is used as verification data for out-of-sample prediction. The mean absolute error (MAE) and mean absolute percentage error (MAPE) are used to evaluate the accuracy of public opinion prediction. The smaller the MAE and MAPE values, the closer the prediction model prediction value is to the actual situation. The prediction results are shown in Table 4.

As can be seen from Table 4, the MAE and MAPE values of Model 1, Model 2, Model 3, and Model 4 are all less than 10, so it is believed that the prediction ability of the model is good.

Table 4: Comparison of public opinion prediction accuracy.

Indicator name	Model 1	Model 2	Model 3	Model 4
MAE	0.78	0.22	0.30	0.40
MAPE	6.80	4.10	5.04	3.64

4. Research Results

The following conclusions were drawn from this study:

(1) There is a long-term equilibrium relationship between the epidemic and public opinion. The ARDL model estimation shows that the independent variable coefficients between the public opinion variables and the number of confirmed cases on the four types of platforms studied in this paper are all positive and statistically significant. Among them, the independent variable coefficient of the short video platform model (Model 2) is the largest, indicating that the increase in the number of newly confirmed cases per day has driven the growth of public activities on the short video platform. The independent variable coefficient of Model 1 ranks second in value, that is, as the epidemic worsens, the traffic of Weibo public opinion also increases. The other two models, namely the changes in public opinion on electronic newspapers or magazines and online news, are also positively correlated with the severity of the epidemic, but their response levels are relatively smaller than those of the first two platforms.

(2) Comparison of prediction accuracy and model performance. According to the comparison of model error accuracy indicators, it can be seen that the model prediction accuracy is ranked as follows: Model 2 has the highest prediction accuracy, followed by Model 3, Model 4, and Model 1 has the worst prediction accuracy. That is, the prediction accuracy is ranked from high to low as follows: short video public opinion, electronic newspaper or magazine public opinion, online news public opinion, and Weibo public opinion.

(3) Different responses of multiple platforms to the spread of the epidemic. By observing the long-term cointegration equation, we can find differences in the responses of different online platforms to the epidemic. The four public opinion platforms can be divided into two categories based on the size of the coefficient of the independent variable. The first category is the highly responsive platform. The data from Weibo and short video platforms have a greater response to the epidemic and a high degree of response, that is, the same epidemic data has a higher impact on these two platforms than on the other two platforms. The second category is the medium-responsive platform. The response intensity of electronic newspapers or magazines and online news platforms is less than that of Weibo and short video platforms, indicating that the epidemic has little impact on the public opinion of electronic newspapers or magazines and online news platforms.

5. Conclusion and Implications

(1) Research findings. This paper constructs a study on the response of the COVID-19 pandemic to online public opinion platforms based on the autoregressive distributed lag model (i.e., ARDL model). The research objects are four types of online media platforms, namely, Weibo, short videos, electronic newspapers or magazines, and online news. The research results show that there is a positive long-term cointegration relationship between changes in the epidemic and changes in media public opinion. Among them, they can be arranged according to the strength of the response as follows: short videos, Weibo, electronic newspapers or magazines, and online news. The public opinion prediction accuracy is arranged from high to low: short video public opinion, electronic newspaper or magazine public opinion, online news public opinion, and Weibo public opinion.

(2) Theoretical contribution. This study has expanded the theory of crisis public opinion

communication and enriched the connotation of agenda setting theory. The ARDL model has been introduced into the study of public opinion communication, adding quantitative analysis tools to public opinion research and realizing the differentiated study of the response of online public opinion platforms to the epidemic.

(3) Practical implications. The practical guidance value of this article lies in providing a real-time monitoring tool for public opinion during the outbreak of a major infectious disease, providing real-time reference data for effectively responding to the spread of the epidemic and managing society. At the same time, this study also provides differentiated reference materials for managing different online public opinion platforms.

In summary, through the empirical research in this article, real-time reference data can be provided for the monitoring and prediction of various major infectious diseases that may break out again in the future, and a factual basis can be provided for the effective management of different online public opinion platforms.

Acknowledgements

References

[1] Wang Wei. A review of the development of crisis communication research[J]. Youth Journalist, 2018, 1(6): 29-30.

[2] Chen Yuefen. The presentation, evolution and governance of COVID-19 rumors in cyberspace[J]. Journal of Shantou University: Humanities and Social Sciences, 2022, 38(1): 75-82.

[3] Liu Yongjian. New characteristics and improvement strategies of public crisis communication in local mainstream newspapers: A case study of Hubei newspapers' response to the "new coronavirus pneumonia" incident[J]. Media, 2020, 1(7): 3.

[4] Kim M, Newth D, Christen P. Modeling dynamics of diffusion across heterogeneous social networks: News diffusion in social media[J]. Entropy, 2013, 15(10): 4215-4242.

[5] Li He, Wang Shuang. Research on the content characteristics of traditional media WeChat video accounts: Taking the Beijing News' "We Video" as an example [J]. Journal of Journalism Research, 2021, 012(001): 66-67.

[6] Cui Xiaoxiao. Analysis of short video reports on current political news in mainstream media from the audience perspective: Taking "Anchors Talking about News Broadcast" as an example [J]. News Culture Construction, 2021, 000(010): P.142-143.

[7] Cui Songge. Research on the content strategy of district-level integrated media WeChat public accounts: Taking "Today's Minhang" as an example [J]. News World, 2023, 1(6): 17-21.

[8] Wen Weiping. Research on newspaper "news links" and online "news links"[J]. Journal of Hunan Mass Media Vocational and Technical College, 2005, 5(4): 3.

[9] Zong Liyong, Gu Baoyan. Research on Multi-Agent Modeling of Network Public Opinion Evolution in Crisis Communication Environment[J]. Information Science, 2010, (9): 7.

[10] Xu Huijie, Jin Hua. Modeling and simulation of online public opinion propagation model for public opinion field[J]. Computer Simulation, 2024, 000(7): 7.

[11] Zhang Liubin, Peng Cheng. Research on the public opinion warning model of local government release based on social safety valve[J]. Information Science, 2020, 38(4): 7.

[12] RAMZAN M, IQBAL HA, USMAN M, et al. Environmental pollution and agricultural productivity in Pakistan: new insights from ARDL and wavelet coherence approaches [J]. Environmental Science and Pollution Research, 2022, 29(19): 28749-28768.

[13] PESARAN MH, SHIN Y. An autoregressive distributed lag modeling approach to cointegration analysis [M]. Department of Applied Economics, University of Cambridge Cambridge, UK, 1995.