Advances in Educational Technology and Psychology (2020) 4: 52-67
Clausius Scientific Press, Canada
DOI: 10.23977/aetp.2020.41010
ISSN 2371-9400

Using Network Literature to Improve Mood during Episodes of Air Pollution: An Empirical Study of Online Comments

Zongyue Xue[1,a], Yutao Gong[2,*]

[1]*School of Public Economics and Administration, Shanghai University of Finance and Economics, Shanghai, P.R. China;*
[2]*School of Finance, Shanghai University of Finance and Economics, Shanghai, P.R. China*
[a]*2017111090@live.sufe.edu.cn*
**Correspondence: 2018111405@live.sufe.edu.cn*

Keywords: mood; air quality; network literature; emotion mining

Abstract: Understanding of the complex physiological and psychological effects of air pollution has grown in recent years. Based on the comments on Jinjiang Literature City, a large Chinese network literature website, Baidu online search index and air quality data in 323 Chinese prefecture-level cities from January 1, 2018 to January 20, 2020, this research reveals that the sentiment of comments improve significantly as air quality deteriorates, while reading behavior also increase significantly. Moreover, this effect is found in positive style works of literature rather than negative style works of literature. These empirical results indicate that network literature has a significant “relieving effect” on residents’ moods, relieving the negative emotions associated with air pollution. Our research demonstrates that changes in air quality can have broader effects on residents’ behaviors and moods over short periods of time than people originally thought. Therefore, efforts to prevent and control air pollution should be strengthened. Subsequent tests of the dependent and independent variables show that the results of this study are relatively robust.

1. Introduction

Following a period of economic reform and opening up, China’s economy has undergone rapid development, resulting in a serious air pollution problem. The 2020 Environmental Performance Index ranked China’s air quality 137th among 180 countries [1]. Multiple studies have shown that air pollution affects residents both physically and psychologically. Ebenstein found that because of severe air pollution, people in the north of China have a life expectancy reduction of 5.5 years caused by cardiopulmonary disease [2]. In addition to the physical effects of air pollution, the long-term impact on psychology and mood is also significant. Based on national data from the United States, Levinson found that concentrations of PM10 and other air pollutants are significantly negatively related to well-being [3]. Meanwhile, research on urban air pollution and the search index of some specific keywords concluded that aggravation from air pollution increases perceptions of government corruption, significantly reduces trust in the government, and makes people more selfish and repressive, and more eager for legal justice [4].

Meanwhile, with the rapid development of global communication technology, the Internet has increasingly become an indispensable part of our lives and is also an important place for people to express their emotions. In this environment, a new literary form, the network literature, as well as the network literature websites that support this form, have emerged. Network literature websites are platforms that provide innovative publishing platforms for professional and amateur writers, and access to online novels for readers. Some novels are free to read or cost a very small proportion of website membership fees. Some writers choose to release some chapters for free, but hide remaining chapters behind a paywall, which can only be accessed through additional payment [5]. These websites encourage authors to write long stories of several million words by paying authors on the basis of word count [6]. As authors receive instant reader feedback, network literature tends to be reader oriented, and authors usually modify their stories according to positive feedback from readers [7].

The network literature industry has undergone rapid development. In 2019, the 44th Statistical Report on Internet Development in China reported approximately 455 million network literature users [8]. In addition, the market revenue of network literature in China was 12.92 billion yuan in 2018, with more than 14 million online writers and 16.47 million online works of literature [9]. Unlike traditional literature formats, the comment sections of network novels allow readers to express their opinions, and to make complaints and suggestions about novel plotlines. These comments are often emotionally driven. At the same time, research has shown that behaviors on the Internet, such as searching practices, often reflect the impact of air quality factors on mood [10]. Therefore, a relationship may exist between readers' online comments and air quality at the time when the comments were posted.

Existing research on the relationship between emotion and air pollution, as reflected in web comments or search behavior, has focused on analysis of key words related to air quality, such as "corruption" [4] or "haze" [11]. There has, thus far, been no research exploring the relationship between the emotions reflected in network novel comments and air pollution.

Network novel comments focus on an evaluation of both novel and author, but do not directly discuss air quality. Rather than processing data according to key words related to air pollution, this research took the innovative approach of exploring emotional behavior, in the form of the tone and sentiment of online comments, under varying conditions of air pollution. In this way, the potential wide-ranging impacts of air pollution could be examined, thus providing further evidence for the necessity of preventing and controlling air pollution.

2. Literature review

A number of studies have found that air pollution has a significant negative impact on mood, mental health, and emotions. As early as 1987, Evans et al. found that exposure to a heavily polluted environment exacerbates depression, anxiety, and tension [12]. Using big data from Weibo (China's version of Twitter), Zheng et al. found that air pollution reduces residents' subjective well-being [13]. In the United States, increases in atmospheric concentrations of PM10 was been found to lead to a significant decline in well-being [3]. A 2020 study by Gao et al. used online data from dianping.com in Beijing to explore the relationship between air pollution and mood, and found that increased air pollution reduces dining satisfaction [14]. Fonken et al. found that haze has a negative impact on the function and structure of the human brain [15], while Zhang et al. found that haze is not only harmful to the body, but may also have a serious negative impact on mental health and cognitive ability, resulting in poor mental condition and decreased subjective well-being [16].

The negative impact of poor air quality on mood can exert an exogenous effect on behavior, often causing individuals to deviate from expectations. In a summary of existing research, Zhang et al. proposed the concepts of air quality cost effect and compensation effect [10]. Cost effect refers to the decline in an individual's information processing ability and the generation of negative moods under air pollution conditions, while the compensation effect refers to the tendency to alleviate the cost effect through shopping and entertainment. A study of the stock market by Levy and Yagil found that air pollution affects investors' stocks and securities investment decisions, and has a significant negative correlation with stock market performance [17]. Using daily Chinese air pollution data from 2005 to 2014, Li and Peng found that air pollution had a same-day negative impact on stock returns, and a two-day lagged positive relationship with stock returns [18]. In addition, studies have shown that air quality significantly reduces labor productivity [19, 20] and classroom attendance [21]. Air quality, therefore, has a multi-dimensional impact on individual behavior.

Internet big data is increasingly used to measure attitudes and behaviors. This research sought to extract data that represented attitudes and behaviors in the traces people leave on the Internet. A rich literature already exists on the use of big data to conduct economic research, which provides a useful reference for the current study. A growing body of literature uses Google or Baidu data to investigate individual behaviors and preferences. Tefft, for example, used Google Trends to investigate the trends in keywords such as depression and anxiety [22], and Da et al. used Google search frequency to measure investors' attention [23]. Goel et al. and Choi and Varian have used Google Trends to predict economic output [24,25]. The use of new Internet data to study environmental problems is a growing field. For example, Zhang and Mu have used online shopping data from Tmall and Taobao to study the impact of haze on the consumption of masks, and creatively analyzed individual environmental protection behavior [26]. Qin and Zhu used search preferences on the Baidu Index to analyze the impact of haze on migration tendencies and found a greater number of migration search terms on hazy days; in other words, haze stimulates individual migration tendencies [27]. Using key word data in Weibo comments, Mei et al. found that on hazy days, individuals are more inclined to complain about haze and convey negative emotions on the Internet [11].

Literature has also been found to be closely related to mood. On the one hand, reading can exert a significant influence on an individual's mood. A meta-analysis by The Reading Agency found that reading can relieve depression and make readers happier [28]. Djikic et al. found that reading literature can subvert biological emotional disengagement of "avoidantly attached individuals" [29]. On the other hand, moods can also have an effect on reading behavior. Mar et al. found that mood can influence which book an individual chooses to read, based partly on whether the reader's goal is to change or maintain their current emotional state [30].

All of the above studies have laid the foundation for this current research and provided theoretical and logical support. The purpose of this study, therefore, was to link air quality, readers' moods, and reading behaviors to explore the impact and influence of air quality on the behaviors and emotions of network literature readers, with the aim of furthering our understanding of the extensive impact of air pollution.

3. Materials and Methods

3.1. Data

3.1.1. Jinjiang Literature City

China's biggest original literature websites are Qidian China, Jinjiang Literature City (JLC), Zongheng China, and Xiaoxiang Academy. Their websites are qidian.com, jjwxc.net, zongheng.com and xxsy.net, respectively. JLC, the focus of this research, was founded in 2013, and currently contains more than 3.44 million network novels and hosts more than 1.46 million registered authors. As of April

2020, the number of registered users exceeded 40 million, reflecting the site’s extensive reader base, heavy web traffic and quantity of comments [31]. This site is representative of the current state of the network novel industry. Of relevance to this research, the Internet Protocol (IP) address of most comments is reserved in the source code of the webpage, so that the city where the comments were made can be easily determined, which allowed for an exploration of the relationship between post sentiment and air quality. In addition, JLC divides novels into categories of different perspectives, allowing for the study of readers’ genre preferences, including “comic”(*Bao Xiao* in Chinese), “relaxed”(*Qin Song* in Chinese), “serious”(*Zhen Ju* in Chinese), “tragic”(*Bei Ju* in Chinese) and “dark”(*An Hei* in Chinese) under air pollution conditions. Compared to JLC, the number of registered users and visitors on similar websites is often lower, leading to a lack of comments, and determining the geographic location of commenters can be difficult as most websites don’t reserve the IP address of comments. Moreover, some websites, such as Qidian China, directly close the comment area and data cannot be obtained. Therefore, JLC was chosen as the most appropriate channel to obtain data for the purposes of this research.

3.1.2. Collection and processing of posts

Data collection. Using a Python multi-threaded crawler, all posts from January 1, 2018 to January 20, 2020 in JLC were crawled. After invalid posts, including those lacking IP addresses or those with IP addresses that could not be connected to a specific city, were excluded, the total number of posts was 51,568,581.

Text Segmentation. Jieba, a widely acclaimed third-party library of Chinese text segmentation, was used to conduct text segmentation. The comments on JLC are often short, and readers always briefly express whether or not they liked the novel they are commenting on and express their opinion regarding the novel. The language styles of review texts are diverse, using various styles of praise, criticism, sarcasm, irony, and ridicule, as well as exhibiting strong emotional sentiments. The use of emoticons and kaomojis is common. These have become important references for judging the sentimental polarity of sentences. For example, the following were posted by two users, A and B:

A: Not enough〒▽〒

B: Not enough↖(^ω^)↗

The first post is negative, as A uses a “crying” kaomoji to express dissatisfaction and frustration with author updates perceived to be too slow. Meanwhile, B’s written “not enough” is accompanied by a pleasant kaomoji, suggesting that B was very impressed by the author's article, which has left B wanting more of the same, so this should be judged as a positive post. If all the special symbols in these posts were deleted as noise data, and only “not enough” was retained, a semantic judgment would classify these two opposing posts in the same category, thus increasing the possibility of misjudging sentimental polarity. Therefore, the identification of kaomoji and other non-linguistic symbols during the text segmentation process was essential to reduce the likelihood of misclassification.

In addition, JLC uses a large number of network terms, with abbreviations and vocabulary unique to the JLC community. For example, “be” means “bad ending”, that is, the novel ends tragically; “ooc” means “out of character”, that is, the characters in the novel employ words and deeds that do not conform to the character's usual settings; and “da call” means cheering. Therefore, commonly used emoticons and network vocabulary of the comments were counted and included in a custom list of commonly used words, so they could be recognized by jieba during text segmentation. Since the jieba library does not recognize special symbols, such as punctuation, by default, its source code was improved in order to recognize special characters.

Convert to word vector. Word2vec was used for unsupervised training. As a word bag model open sourced by Google in 2013, Word2vec uses a double-layer neural network to generate word vectors, which can map each word to a vector to achieve the generation of word vectors. Unlike One-Hot Vector, it assumes that the meaning of a word can be derived from the context. While simplifying the dimension of word vectors and avoiding dimensional catastrophe, it considers the co-occurrence between words, which can reflect the interrelationship of words in a sentence. The word vectors of synonyms are relatively close, and this paves the way for the subsequent training of semantic recognition models. On the basis of text segmentation, the word frequency of each word in the overall posts was counted, and words with a word frequency of less than 10 were eliminated to simplify the workload. Python's Word2vec module was used to convert each word into a 200-dimensional vector to achieve text preprocessing.

Semantic recognition. A random sample of 20,000 posts was conducted and these were manually scored by the authors. Scores of −1, 0, or 1, were awarded, where −1 represented negative sentiment and 0 represented neutral sentiment, which contained many meaningless posts, and 1 represented positive sentiment. Next, machine learning was conducted using a convolutional neural network (CNN). After dividing the training set and the test set according to a ratio of 8:2, the accuracy of the CNN on the test set reached 89.34%. Perhaps because of the simplicity and repetition of sentences, JLC posts are not particularly complex, so this result may be higher than that in similar studies. Other machine learning classification algorithms were also applied to sentiment analysis. These included Gaussian and Bernoulli naive Bayes as well as random forest and support vector machine classifiers. The accuracy of these algorithms on the sentiment of comments on JLC was found to be significantly lower than that of CNN.

Data aggregation. The comment data were aggregated by city and date, and the total number of posts, positive posts, and negative posts sent by each city each day were counted. Each day in each city was recorded as city*day, and the Sentiment Index per city*day was calculated accordingly. There are two reasons for this setting. First, if the score of the neural network was directly included as a dependent variable in the linear regression equation, because it was a discrete variable, explaining the meaning of the regression result would prove difficult. Second, the aggregation of comment data by city*day greatly reduced the number of dependent variables and facilitated calculation. After the data were aggregated, the number of city*days totaled 234,962.

Build index. In order to facilitate an investigation of the impact of air quality on network literature sentiments, this research drew on the work of Antweiler and Frank to construct an index to measure network literature sentiment, which is referred to as the Sentiment Index (SI) [32]. The formula for calculating SI is as follows:

$$SI_{cd} = \ln[\frac{1+\mathrm{Num}_{cd}^{POS}}{1+\mathrm{Num}_{cd}^{NEG}}] \quad (1)$$

where Num^{POS} represents the number of positive sentimental posts, Num^{NEG} represents the number of negative sentimental posts, and c and d represent the city and date, respectively. This index allowed for a value change within a smaller range and reduced the influence of extreme values. In addition, the value of the Sentiment Index constructed by this method approximately represents the ratio of positive posts to negative posts minus 1, which has certain practical significance. According to this method, the Sentiment Index can comprehensively measure the overall sentiment of comment posts on network literature.

3.1.3. Air quality data

The air quality data in this paper is originated from the website of China’s Ministry of Ecology and Environment (MEE), which was originally named the Ministry of Environmental Protection of China until September 1st, 2018. Multiple previous studies that focused on air quality used the same air quality data sourse [4,26,27,33]. The daily air quality index (AQI) is used by MEE to comprehensively measure air quality of a particular city. Since 2012, the AQI index has been published in a number of Chinese cities to replace the original API index. The calculation of the AQI index is based on six pollutants: fine particulate matter (PM2.5), respirable particulate matter (PM10), nitrogen dioxide (NO2), carbon monoxide (CO), sulfur dioxide (SO2), and ozone (O3). Compared with former API, the current AQI included PM2.5 and O_3 as new criteria pollutants, and the frequency of reporting pollution also increased from daily to hourly. Therefore, the AQI provides a more objective and representative account of the degree of urban pollution. The AQI value ranges from 0 to 500, and a higher value indicates poorer air quality. The AQI index is divided into six graded categories, AQI $\geq$ 50 (excellent), 51 $\leq$ AQI $\leq$ 100 (good), 101 $\leq$ AQI $\leq$ 150 (lightly polluted), 151 $\leq$ AQI $\leq$ 200 (moderately polluted), 201 $\leq$ AQI $\leq$ 300 (heavily polluted), and 301 $\leq$ AQI $\leq$ 500 (severely polluted). For specific standards and calculation methods, please refer to the Ambient Air Quality Standard (GB3095-2012) issued by the Ministry of Environmental Protection of China.

From January 1, 2018 to January 20, 2020, the average AQI values for 323 Chinese cities was 64.7, which can be categorized as "good". Days with different air quality levels in 323 Chinese cities over the same period are shown in Figure 1. Among 242250 city*days, 43.14% of them were categorized as "excellent", while about 1.8% of city*days fell within the range of "heavily polluted" or "severely polluted", with an AQI over 200, showing significant improvement of air quality compared with previous studies a few years ago [4, 26].

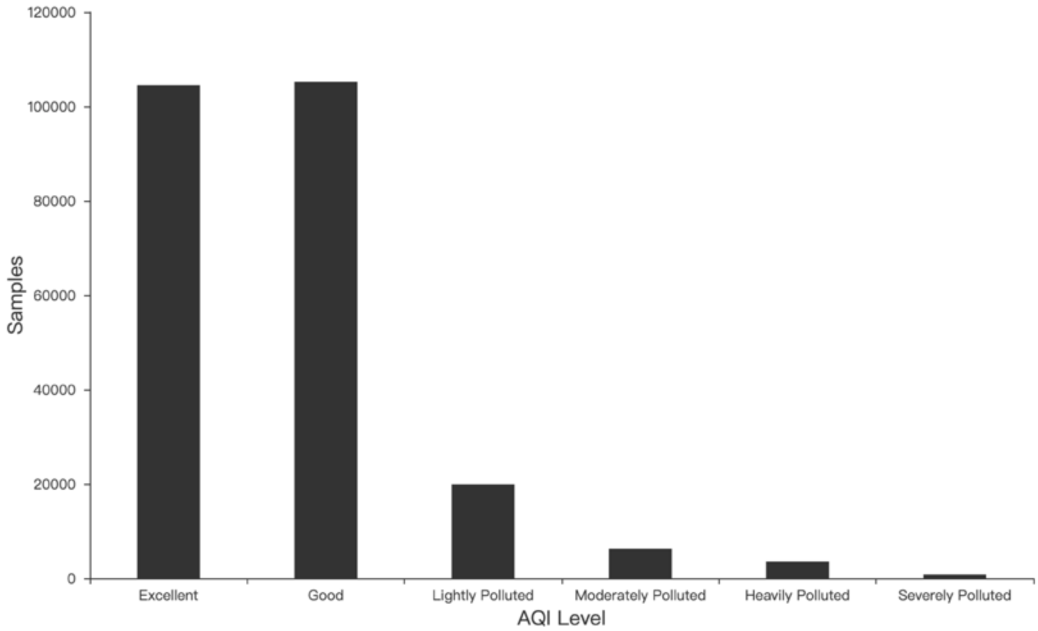

Figure 1. Days with different AQI levels in 323 Chinese cities from January 1, 2018 to January 20, 2020

Data Source: Ministry of Ecology and Environment of China

3.1.4. Weather data

Multiple studies have revealed that weather can affect people's behaviors and moods in various ways. For example, higher temperature or barometric pressure was generally related to better mood [34], while fewer sunshine hours and higher humidity may lead to "lower moods" [35]. Hirshleifer and Shumway analyzed the influence of weather factors on the stock market, finding that investors' moods are more positive and the stock market performs better on sunny days [36]. Moreover, it is also considered that local weather may considerably affect air quality [26,33]. Therefore, to avoid the potential biased estimations caused by weather conditions, a list of weather variables are included in our regressions. Specifically, weather data, including daily temperature, humidity, rainfall, and wind speed, were extracted from the data of 800 national meteorological monitoring stations stored on weather.com.cn, which is run by the China Meteorological Administration, providing the public with daily weather data at the monitor station level. Data from the meteorological monitoring station closest to each city was regarded as the weather data for that city.

3.2. Statistical Analysis

The post came from 323 prefecture-level cities across China. Of the 51,568,570 posts, the number of sentiments identified as positive was 29,533,312; negative was 11,629,334; and neutral was 10,407,924. Table 1 shows the descriptive statistics of the data aggregated by city and date. To save space, only the most important variables are included in Table 1. The table shows that the mean value of the Sentiment Index was 0.856, which is greater than 0, indicating an overall positive sentiment. The average number of posts per city*day was 212.873, and the standard deviation was greater than 500, indicating that the post number of the city*day varied greatly. In addition, the standard deviation of AQI and PM2.5 was large, reflecting the uneven distribution of air pollution in time and space.

Table 1. Descriptive statistics of main variables

Variables	Data Size	Mean Value	Standard Deviation	Minimum Value	Maximum Value
SI	234961	0.856	0.477	−2.398	4.684
Num	242250	212.873	513.056	0.000	9185.000
Num^{POS}	242250	140.098	334.259	0.000	6468.000
Num^{NEG}	242250	0.996	3.501	0.000	155.000
AQI	240290	64.707	42.881	7.667	500.000
$PM_{2.5}$	240285	38.415	33.511	1.600	1787.583
Temp	240884	14.132	11.493	−37.600	40.600
Rain	228766	2.589	9.344	−2.900	363.600
Humidity	240885	67.180	19.475	2.620	100.000
Wind speed	240833	2.472	1.354	0.000	20.570

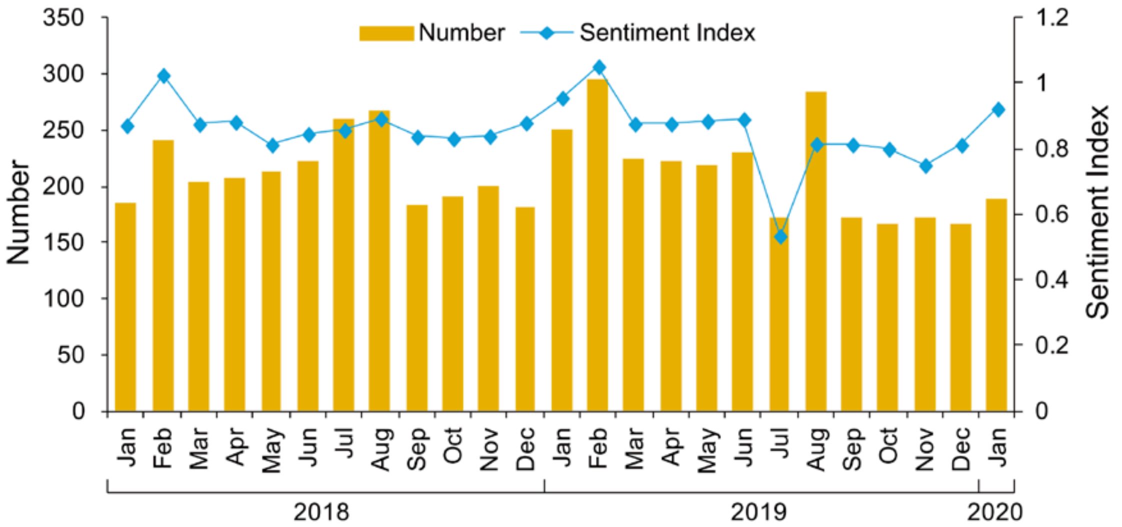

Figure 2. The monthly distribution of average post Sentiment Index and average post number

Figure 2 shows the monthly distribution of average post Sentiment Index and average post number. The distribution shows a certain degree of fluctuation in sentiment and post number from month to month. Therefore, the fixed effect of year-week were strictly controlled as fixed effects. The Sentiment Index also showed a certain degree of consistency with the trend of the post number.

3.3. Research design

Based on the above dependent and independent variable settings, the following measurement equation was used:

$$SI_{cd} = \beta_0 + \beta_1 \ln\left(AQI_{cd}\right) + \lambda X_{cd} + \delta_c + \mu_d + \varepsilon_{cd} \quad (2)$$

where the subscript c represents the city, and the subscript d represents the date (year, month, day); SI_{cd} represents the overall Sentiment Index of a network post on a city*day, and $\ln(AQI)_{cd}$ represents the air quality index. In addition, X_{cd} represents a list of weather variables, including temperature, humidity, rainfall, and wind speed of the day in question, in order to control the potential influence of weather conditions on sentiment. δ_c represents the dummy variable of prefecture-level city to control the fixed effect of the city. μ_d represents the fixed effect of time. Specifically, the fixed effect of year-week, public holiday (dumb variable) and day of week (dumb variable) on sentiment were controlled for. ε_{cd} is a random disturbance. Using this model, we not only studied the changes in post sentiment according to air quality change at different times in the same city, we also compared the changes in sentiment caused by different air quality in different cities at the same time. A focus on the coefficient β_1 of $\ln(AQI)_{cd}$ reflected the direction and magnitude of sentiment change with air quality change. To better study the relationship with the dependent variable and eliminate the problem of heteroscedasticity that may have existed in the data, the natural logarithm of AQI was unified as the independent variable. Since taking the logarithm of the Sentiment Index had no practical meaning , the original value was retained and introduced into the model.

4. Results

4.1. Basic regression

The Sentiment Index was first used as the dependent variable to perform the basic regression. The results are shown in Table 2.

In the basic regression, fixed effects are added step by step. Column (1) in Table 2 directly regresses ln(AQI) and the Sentiment Index. Column (2) adds variables representing weather, including temperature, rainfall, humidity and wind speed. Column (3) adds the dummy variable of the city based on column (2) to absorb the factors that can’t be observed in spatial dimensions. Column (4) continues to add the dummy variable of the year-week to capture the temporal and spatial unobservable factors. Column(5) adds the rest of the time variable, including the day of the week and public holidays, providing evidence obtained with the strictest regression. The empirical results show that the increase of the control variable has no significant effect on the significance of the independent variable ln(AQI) coefficient in the regression results, which are significantly positive at the 1% level. These results unanimously show that when the AQI level increased, overall post sentiments on the JLC website increased significantly.

These results seem to contradict the conclusions of previous research that air quality decline leads to an increase in depression. However, Zhang et al. found that there may be a compensation effect when air quality declines, as residents mitigate their negative emotions through the consumption of shopping and entertainment [10]. Other research has found that reading releases feelings of depression and improves mood [28]. Therefore, it is possible that ln(AQI) generated in Table 1 is positively correlated with the Sentiment Index because as air quality deteriorates, people may have negative emotions including depression, anxiety, etc. , so they turn to network literature to entertain themselves. Therefore, the act of reading network literature improves mood when air quality is low, having an overall positive effect on post sentiment. If this is the case, then network literature can be considered to have a “relieving effect” on the negative emotions generated by deteriorating air quality. If this assumption is true, then it is logical to deduce that people are likely to read more network literature, as such a behavior can alleviate the negative emotions that arise when air quality deteriorates.

Table 2. Results of basic regression

	(1) **SI**	(2) **SI**	(3) **SI**	(4) **SI**	(5) **SI**
ln(AQI)	0.060***	0.068***	0.032***	0.013***	0.011***
	(0.009)	(0.010)	(0.003)	(0.003)	(0.003)
Temperature	No	−0.001**	−0.003***	0.000	0.001
		(0.000)	(0.000)	(0.000)	(0.000)
Rainfall	No	0.000*	0.001***	0.000	0.000*
		(0.000)	(0.000)	(0.000)	(0.000)
Humidity	No	0.001***	−0.001***	-0.000	-0.000
		(0.000)	(0.000)	(0.000)	(0.000)
Wind speed	No	0.003	−0.001	-0.000	0.000
		(0.003)	(0.001)	(0.001)	(0.001)
City FE	No	No	Yes	Yes	Yes
Year-week FE	No	No	No	Yes	Yes
Holiday FE	No	No	No	No	Yes
Day-of-week FE	No	No	No	No	Yes
N	234950	221624	221624	221624	221624
Adjusted R^2	0.00456	0.00705	0.0776	0.142	0.150

Robust standard errors, clustered at the city level, are presented in parentheses; ***, ** and * indicate the significance levels of 1%, 5% and 10%, respectively.

4.2. The impact of air quality on page views of JLC

In order to further verify the assumption that network literature has a “relieving effect” as air quality deteriorates, the Sentiment Index in the basic equation in Column (4) of Table 2 was replaced with the total number of posts as the new dependent variable to estimate whether reading behavior increases under air pollution. However, a significant increase in post numbers during periods of deteriorated air quality does not necessarily represent a change in the number of visits to the website, as it may simply reflect an increased desire for public expression under such air quality conditions. Therefore, the Baidu online search index of keywords related to Jinjiang Literature City was employed as the second variable to verify, from different perspectives, whether during periods of air pollution people read more on the JLC site. Baidu is the top search engine in China, and the Baidu Index contains data generated by Baidu based on the behavior of a large number of Baidu users, which can be used to analyze the search trends of specific keywords in specific periods of time. Qin and Zhu have found that the Baidu index and actual search volume are highly likely to have linear correlation, making the Index a good proxy for real search behavior [27]. One function of the Baidu Index is that different keywords can be added together. Therefore, eight Baidu Index keywords were selected based on the principle of comprehensiveness and accuracy. These keywords were "Jinjiang Literature" (*Jin Jiang Wen Xue* in Chinese), "Jinjiang Literature City" (*Jin Jiang Wen Xue Cheng* in Chinese), "Jinjiang Novel" (*Jin Jiang Xiao Shuo* in Chinese), "Jinjiang Novel Network" (*Jin Jiang*

Xiao Shuo Wang in Chinese), "Jinjiang Original Network" (*Jin Jiang Yuan Chuang Wang* in Chinese), "Jinjiang Literature Network" (*Jin Jiang Wen Xue Wang* in Chinese), "Jinjiang Literature City Mobile Version" (*Jin Jiang Wen Xue Cheng Shou Ji Ban* in Chinese), and "Gallery of Jinjiang Literature City" (*Jin Jiang Wen Xue Cheng Zuo Pin Ku* in Chinese). The Baidu Index of these eight keywords in the same city on the same day was added together to achieve a comprehensive search index. These Baidu keywords lead directly to the official JLC website. Other related terms were excluded from the search. The keyword "Jinjiang" (*Jin Jiang* in Chinese), for example, had a high Baidu Index, but could also refer to "Jinjiang City" (*Jin Jiang Shi* in Chinese), leading to ambiguity. Other JLC-related keywords in the Baidu Index, including "Jinjiang Literature City Computer Version" (*Jin Jiang Wen Xue Cheng Dian Nao Ban* in Chinese) and "Jinjiang Literature City Mobile" (*Jin Jiang Wen Xue Cheng Shou Ji* in Chinese), were newly included after January 1, 2018, the date marking the start of the research time frame, and were, therefore, excluded.

In addition, when the Sentiment Index was used as the dependent variable, if the total post number of a city*day was 0, calculating the Sentiment Index of posts was meaningless. Therefore, the regression data in Table 1 do not retain city*days with zero posts. However, when conducting regression on post number and the search index, zero posts were meaningful because the number of posts increasing from zero could reflect the impact of air quality on behavior. For this reason, all city*days with a post number of 0 were retained when the post number or the search index were dependent variables. To avoid the situation where the result has no mathematical significance, 1 was added to the value before taking the logarithm of the post amount and the search index. This is also the case for tables 4, 6 and 7 that refer to post number or search index for the same reason. Meanwhile, this measure doesn’t have significant impact on the empirical results as our further analysis found the inclusion and exclusion of post numbering 0 led to similar results (results not shown).

The regression analysis of post number and search index as dependent variables is shown in Column (1) and Column (2) of Table 3.

Table 3. The impact of air quality on page views of JLC

	(1) **ln(Num)**	(2) **ln(Search)**
ln(AQI)	0.023***	0.015**
	(0.005)	(0.006)
Weather	Yes	Yes
City FE	Yes	Yes
Year-week FE	Yes	Yes
Holiday FE	Yes	Yes
Day-of-week FE	Yes	Yes
N	226660	226660
Adjusted R^2	0.941	0.832

Robust standard errors, clustered at the city level, are presented in parentheses; ***, ** and * indicate the significance levels of 1%, 5% and 10%, respectively.

Table 3 shows that the total post-logarithmic post and search index were significantly positively correlated with ln(AQI), showing that when air quality deteriorates, the search for keywords related to JLC in Baidu increased and individuals wrote more posts. This confirms that, from different perspectives, there is a significant increase in JLC novel reading behavior under conditions of poor air quality.

4.3. The effect of different style novels

JLC novels are tagged according to five styles: "comic", "relaxed", "serious", "tragic" and "dark", as well as a sixth, "unknown", tag. Each novel is assigned one of the tags, and changes in novel preference under different air quality conditions were explored. For the purposes of this research, "comic" and "relaxed" styles were defined as "positive", and "tragic" and "dark" were defined as "negative". The few novels of unknown style were eliminated from the study. Columns (1) and (2) in Table 4 show the impact of changes in air quality to the number of posts relating to positive and negative style novels. Columns (3) and (4) show how post sentiment changed in positive style novels and negative style novels, respectively, when air quality changed. The results show that, for positive novels, both post number and Sentiment Index increased significantly when air quality deteriorated, with no difference to the regression result for the entire database. However, there was no significant change to post number or Sentiment Index for negative novels when air quality changed. This suggests that negative style novels are rarely read during periods of deteriorating air quality, but positive style novels are read more, and the post sentiment of positive style novels also improves significantly. This difference between positive and negative style novels continues to confirm the assumption that network literature can have the “relieving effect” during episodes of air pullution because the comic and relaxed style novels have been written to make people feel happier, helping to relieve people’s negative emotions. Naturally, it makes sense that as people have depression or anxiety that originates from air pollution, they consciously or subconsciously choose positive style novels to relieve their symptoms. Therefore, the results in Table 4, combined with Table 2 and 3, provide preliminary confirmation of the hypothesis that network literature has a relieving effect when air quality is poor.

Table 4. Changes in post number and sentiment of novels of different styles under air pollution

	(1)	(2)	(3)	(4)
	ln(Num_Pos)	**SI_Pos**	**ln(Num_Neg)**	**SI_Neg**
ln(AQI)	0.027***	0.010***	0.004	0.008
	(0.005)	(0.003)	(0.004)	(0.010)
Weather	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
Year-week FE	Yes	Yes	Yes	Yes
Holiday FE	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes
N	221624	219945	226660	63740
Adjusted R^2	0.921	0.146	0.517	0.0468

Robust standard errors, clustered at the city level, are presented in parentheses; ***, ** and * indicate the significance levels of 1%, 5% and 10%, respectively. Num_Pos, number of posts relating to positive style novels; Num_Neg, number of posts relating to negative style novels; SI_Pos, post sentiment in positive style novels; SI_Neg, post sentiment in negative style novels.

4.4. Robustness test

4.4.1 Robustness of AQI

Because AQI is a comprehensive evaluation of air quality, it is the index followed by most people. However, over the years, PM2.5 has gradually become a key indicator of air pollution and smog severity. PM2.5 is a particulate matter with a diameter less than or equal to 2.5 microns in air, which can float for an extended period of time and over long distances. The high concentration of PM2.5 in the atmosphere not only reduces visibility, it also greatly affects normal production and individual life [37] and is a major cause of respiratory and cardiovascular disease [38]. PM2.5 has, therefore, received wide attention and was considered another important basis for measuring air quality in this research. Therefore, AQI was replaced with PM2.5 concentration data, and the logarithm of PM2.5 data was used as an independent variable as a robustness test. The basic regression results of the robustness test are shown in Table 5.

Table 5. Robustness Test for AQI

	(1) **SI**	(2) **ln(Num)**	(3) **ln(Search)**	(4) **ln(Num_Pos)**	(5) **SI_Pos**	(6) **ln(Num_Neg)**	(7) **SI_Neg**
$\ln(PM_{2.5})$	0.008^{***}	0.019^{***}	0.012^{**}	0.022^{***}	0.007^{**}	0.004	0.008
	(0.003)	(0.004)	(0.005)	(0.004)	(0.003)	(0.003)	(0.008)
Weather	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Holiday FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	221622	226655	226655	221622	219943	226655	63740
Adjusted R^2	0.150	0.941	0.832	0.921	0.146	0.517	0.0468

Robust standard errors, clustered at the city level, are presented in parentheses; ***, ** and * indicate the significance levels of 1%, 5% and 10%, respectively. Num_Pos, number of posts relating to positive style novels; Num_Neg, number of posts relating to negative style novels; SI_Pos, post sentiment in positive style novels; SI_Neg, post sentiment in negative style novels.

Table 5 shows the calculations for the Sentiment Index, the total number of posts related to novels, the comprehensive Baidu Index, and the post number and the Sentiment Index of different style novels. No significant difference was found between the results of Table 5 and those of Tables 2, 3 and 4. The relieving effect of reading network literature is significant for both AQI data and PM2.5 data, illustrating that the regression results obtained are relatively reliable.

4.4.2 Robustness of the Sentiment Index

As the key dependent variable for measuring changes in post sentiment, it was necessary to fully test the robustness of the Sentiment Index. We further referred to two other methods, used by Antweiler and Frank [32], to construct the Sentiment Index: Sentiment Index 2 (SI2), floating between −1 and 1, and Sentiment Index 3 (SI3), following formula (3) and formula (4), respectively. The results of the robustness test are shown in Table 6.

$$SI2_{cd} = \frac{Post_{cd}^{POS} - Post_{cd}^{NEG}}{Post_{cd}^{POS} + Post_{cd}^{NEG}} \quad (3)$$

$$SI3_{cd} = SI2 \times \ln(1 + Post_{cd}^{POS}) \quad (4)$$

Table 6. Robustness Test for Sentiment Index

	(1) **SI2**	(2) **SI3**	(3) **SI2_Pos**	(4) **SI3_Pos**	(5) **SI2_Neg**	(6) **SI3_Neg**
ln(AQI)	0.004^{**}	0.026^{***}	0.010^{***}	0.023^{***}	0.008	0.011
	(0.002)	(0.006)	(0.003)	(0.005)	(0.010)	(0.012)
Weather	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-week FE	Yes	Yes	Yes	Yes	Yes	Yes
Holiday FE	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes	Yes	Yes
N	221624	221624	219945	219945	63740	63740
Adjusted R^2	0.0839	0.490	0.146	0.485	0.0468	0.0470

Robust standard errors, clustered at the city level, are presented in parentheses; ***, ** and * indicate the significance levels of 1%, 5% and 10%, respectively. SI2_Pos, the Sentiment Index 2 in positive style novels; SI3_Pos, the Sentiment Index 3 in positive style novels; SI2_Neg, the Sentiment Index 2 in negative style novels; SI3_Neg, the Sentiment Index 3 in negative style novels.

Table 6 shows that the results of the newly constructed Sentiment Index 2 and Sentiment Index 3 are not significantly different from the original Sentiment Index. The Sentiment Index 2 and 3 of both the whole database and comments on positive style novels have a positive relationship with air pollution, while such a relationship is not found in negative style novels. This proves that the construction method of the Sentiment Index has no significant effect on the regression results.

In addition, in our further analysis, given that the distribution of post number per city*day was very uneven, city*days with less than 5 posts were excluded from analysis to avoid randomness in the results shown in Tables 2, 3 and 4. In total, 16,807 city*days were excluded, leading to no significant change in the regression results (results not shown).

5. Discussion

The empirical results of this research shed new light on the effects of literature under conditions of air pollution. The results of Table 1 suggest that the sentiment of online comments in JLC is more positive on air-polluted days. Network literature is a form of entertainment similar to TV series or music. As the quality of network literature itself is not likely to be correlated with air pollution, the greater positivity of comment sentiments is most likely caused by changes in readers' emotional status. This means that, on days of increased air-pollution, network literature offers readers an more enjoyable form of entertainment than usual. However, as comments are usually posted while in the process of or upon completion of reading a novel, changes in sentiment mean that the reader's mood is improved while he or she is reading or when he or she has just completed reading a novel. How long this effect lasts, and whether a reader's improved mood can be maintained for the remainder of an air-polluted day, requires further investigation.

Our empirical results suggest the existence of the phenomenon that people tend to read more network literature on days of higher air pollution, and the sentiment of posts is also more positive. It seems that this phenomenon may be caused by different mechanisms. It is probable that the use of online literature sites is a result of people spending more time indoors on days of higher air pollution, as they are less likely to be engaged in outside activities, such as exercising or socializing, and may turn to online forms of entertainment as a substitution for other free time activities. Therefore, reading behavior may be a compensation for not being able to get out of doors, and improvements in sentiment are perhaps a by-product of this behavior. Meanwhile, the relieving effect of network literature may also be the result of people actively seeking a release for the negative emotions of depression [12] and dissatisfaction [14], or the decline in well-being [3,13,16] that has previously been associated with air pollution.

However, the empirical results in Table 4 revealed that positive novels have a significant relieving effect on emotions, while negative novels have little effect under conditions of air pollution. As we found no proof that the amount of time spent indoors has an impact on the type of books people read, if the relieving effect is caused by people spending more time indoors, it is hard to explain why negative style novels would have no such effect. Thus, such a phenomenon is highly likely due to the depression experienced under conditions of air pollution and the needs to relieve that depression. This also helps to explain why people are more likely to read "comic" and "relaxed" books under conditions of air pollution, as these books have been written to make people feel happier and more relaxed. Therefore, our analysis in Table 4 suggests that the relieving effect of network literature under conditions of air pollution is probably caused by an increased need to relieve feelings of depression.

6. Conclusions

Our understanding of the impact of air pollution on health, behavior, and sentiment has grown in the past decade. Based on previous research, using panel data and air pollution data linked to posts on JLC from residents in 323 cities in China from January 1, 2018 to January 20, 2020, this research explored the impact of air pollution on behavior and sentiment. Several conclusions can be drawn from the findings.

First, network literature has a significant "relieving effect" on sentiment when air quality deteriorates. The empirical results show that the sentiment of network literature-related posts improved significantly under air pollution conditions, while reading behavior also increased significantly. Moreover, this effect is found in positive style works of literature rather than negative style works of literature. Combined with existing research on the negative impact of air pollution on emotions and extensive compensation effects during smog events, this confirms the hypothesis of the "relieving effect" of network literature. In other words, when air pollution causes individuals' moods to drop, they will spend more time reading network literature, which may relieve and soothe their negative emotions. This understanding of the positive significance of network literature on psychology and sentiment should be the basis for promoting the healthy development of this industry. In addition, for countries or regions with severe air pollution, the creation of relaxing and positive cultural works of literature should be encouraged, and network literature websites should try to promote such works, making them easier to access, thus helping to mitigate the negative impact of air pollution on emotions.

Second, previous research on the impact of air quality on behavior and sentiment has tended to focus on online search behavior [4], shopping behavior [26], and other activities related to air quality or smog. However, the research in this paper indicates that network literature, which would appear to be unrelated to air quality, is actually inextricably linked to it. This suggests that the impact of air pollution on residents is likely to be more extensive and profound than previously thought. Therefore, many countries, including China, should pay greater attention to the prevention and control of air pollution as they work toward greater economic development. The establishment of scientific and long-term mechanisms to defend the blue sky should be prioritized, in order to protect the physical and mental health of individuals, and to effectively increase happiness.

Finally, some limitations of this research should be considered. First, this research was based on online comments, and as the majority of Internet users are younger members of society, the current research may have an age bias. Second, we explored the relationship between mood, literature, and air pollution using comment data aggregated at the city level rather than direct investigation, such as questionnaires or experiments at the individual level. This restricted our ability to more comprehensively analyze this relationship. Third, as mentioned in the Discussion, comments are usually posted when readers are reading or immediately after they finish reading; thus, further exploration is required to understand the extent to which the relieving effect of literature under air pollution conditions can last when not engaged in online literature sites. Last but not least, as our research found that network literature, as a form of entertainment, has a "relieving effect" during periods of air pollution, whether other forms of entertainment, such as films, TV series or music have this effect as air quality deteriorates may also require further exploration in the future.

References

[1] Wendling, Z.A.; Emerson, J.W.; de Sherbinin, A.; Esty, D.C.; et al. 2020 Environmental Performance Index. 2020. Available online: https://epi.yale.edu/ (accessed on 15 June 2020)

[2] Ebenstein, A.; Fan, M.; Greenstone, M.; He, G.; Zhou, M. New evidence on the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy. Proc. Natl. Acad. Sci. U. S. A. 2017, 114,10384-10389.

[3] Levinson, A. Valuing Public Goods Using Happiness Data: The Case of Air Quality, J. Public Econ. 2012, 96, 869-880.

[4] Shi, Q.; Guo, F. Do people have a negative impression of government on polluted days? Evidence from Chinese cities. J. Environ. Plan. Manag. 2019, 62, 797-817.

[5] Wang, Y. Globalization of Chinese Online Literature: Understanding Transnational Reading of Chinese Xuanhuan Novels Among English Readers. Available online: http://www.inquiriesjournal.com/a?id=1716 (accessed on 2 June 2020).

[6] Lugg, A. Chinese Online Fiction: Taste Publics, Entertainment, and Candle in the Tomb. Chin J Commun. 2011, 4, 121-136.

[7] Tian, X.; Adorjan, M. Fandom and Coercive Empowerment: The Commissioned Production of Chinese Online Literature. Media Cult. Soc. 2016, 38, 881-900.
[8] China Internet Information Center. The 44th Statistical Report on Internet Development in China. Available online: http://www.cac.gov.cn/2019-08/30/c_1124938750.htm (accessed on 5 April 2020).
[9] Tao, D.; He, Y. Report on the Development of Chinese network literature is Released. Available online: http://culture.people.com.cn/n1/2019/0810/c429145-31287235.html (accessed on 1 June 2020)
[10] Zhang, Z.; Wei, F.; Han, J.; Chen, W. Impact of air quality on consumer Internet search behavior. J. Man. Sci. 2018, 31, 16-29.
[11] Mei, S.; Li, H.; Fan, J.; Zhu, X.; Dyer, C.R. Inferring air pollution by sniffing social media, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), Beijing, China, 17-20 Aug 2014, pp.534-539.
[12] Evans, G.W.; Jacobs, S.V.; Dooley, D.; Catalano, R. The interaction of stressful life events and chronic strains on community mental health, Am. J. Community Psychol. 1987, 15, 23-34.
[13] Zheng, S.; Wang, J.; Sun, C.; Zhang, X.; Kahn M. Air pollution lowers Chinese urbanites' expressed happiness on social media. Nat. Hum. Behav. 2019, 3, 237-243.
[14] Gao, R.; Ma, H.; Ma, H.; Li, J. Impacts of Different Air Pollutants on Dining-Out Activities and Satisfaction of Urban and Suburban Residents. Sustainability 2020, 12, 2746:1-2746:13.
[15] Fonken, L.K.; Xu, X.; Weil, Z.M.; Chen, G.; Sun, Q.; Rajagopalan, S.; Nelson, R.J. Air pollution impairs cognition, provokes depressive-like behaviors and alters hippocampal cytokinen and morphology, Mol. Psychiatry. 2011, 16, 987-995.
[16] Zhang, X.; Zhang, X.; Chen, X. Happiness in the air: How does a dirty sky affect mental health and subjective well-being? J. Environ. Econ. Manage. 2017, 85, 81-94.
[17] Levy, T.; Yagil, J. Air pollution and stock returns in the US. J. Econ. Psychol. 2011, 32, 374-383.
[18] Li, Q.; Peng, C.H. (2016) The stock market effect of air pollution: evidence from China, Appl. Econ. 2016, 48, 3442-3461.
[19] Li, T.; Liu, H.; Salvo, A. Severe Air Pollution and Labor Productivity: Evidence from Industrial Towns in China. Am. Econ. J.-Appl. Econ. 2019, 11, 173-201.
[20] Graff Zivin, J.; Neidell, M. The impact of pollution on worker productivity. Am. Econ. Rev. 2012, 102, 3652–3673.
[21] Currie, J.; Hanushek, E.A.; Kahn, E.M.; Neidell, M.; Rivkin, S.G. Does pollution increase school absences? Rev. Econ. Stat. 2009, 91, 682–694.
[22] Tefft, N. Insights on unemployment, unemployment insurance, and mental health. J. Health Econ. 2011, 30, 258-264.
[23] Da, Z.; Engelberg, J.; Gao, P.J. In search of attention. J. Financ. 2011, 66, 1461-1499.
[24] Goel, S.; Hofman, J.M.; Lahaie, S.; Pennock, D.M.; Watts, D.J. Predicting consumer behavior with web search. Proc. Natl. Acad. Sci. USA. 2010, 107, 17486-17490.
[25] Choi, H.; Varian, H. Predicting the present with Google Trends. Econ Rec. 2012, 88, 2-9.
[26] Zhang, J.; Mu, Q. Air pollution and defensive expenditures: Evidence from particulate-filtering facemasks. J.Environ.Econ.Manage. 2018, 92, 517-536.
[27] Qin, Y.; Zhu, H. Run away? Air pollution and emigration interests in China. J. Popul. Econ., 2018, 31, 235-266.
[28] The Reading Agency. Literature Review: The impact of reading for pleasure and empowerment. Available online: https://readingagency.org.uk/news/The%20Impact%20of%20Reading%20for%20Pleasure%20and%20Empowerment.pdf (accessed on 9 May 2020).
[29] Djikic, M.; Oatley, K.; Zoeterman, S.; Peterson, J.B. Defenseless against art? Impact of reading fiction on sentiment in avoidantly attached individuals. J. Res. Personal. 2009, 43, 14-17.
[30] Mar, R.A.; Oatley, K.; Djikic, M.; Mullin, J. Sentiment and narrative fiction: Interactive influences before, during, and after reading. Cogn. Emot. 2011, 25, 818-833.
[31] Jinjiang Literature City official website. Available online: http://www.jjwxc.net/aboutus (accessed on 10 May 2020).
[32] Antweiler, W.; Frank, M.Z. Is all that talk just noise? The information content of Internet stock message boards. J. Financ. 2004, 59, 1259-1294.
[33] Viard, V. B.; Fu, S. The effect of beijing's driving restrictions on pollution and economic activity. J. Public Econ. 2015, 125, 98-115.
[34] Keller, M. C.; Fredrickson, B. L.; Ybarra, O., Côté, S.; Johnson, K.; Mikels, J.; Conway, A.; Wager, T. A Warm Heart and a Clear Head: The Contingent Effects of Weather on Mood and Cognition. Psychol. Sci. 2005, 16, 724–731.
[35] Persinger, M.A. Lag responses in mood reports to changes in the weather matrix. Int. J. Biometeorol. 1975, 19, 108–114.
[36] Hirshleifer, D.; Shumway, T. Good day sunshine: Stock returns and the weather. J. Financ. 2003, 58, 1009-1032.
[37] Hanna, R.; Oliva, P. The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City, J.

Public Econ. 2015, 122, 68-79.
[38] Polichetti, G.; Cocco, S.; Spinali, A.; Trimarco, V.; Nunziata, A. Effects of particulate matter (PM_{10}, $PM_{2.5}$ and PM_1) on the cardiovascular system. Toxicology, 2009, 261, 1-8.