KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

Economic impact of air pollution

Dissertation presented to obtain the degree of Doctor in KU Leuven and Tsinghua University

by

Qian Wang

Economic impact of air pollution

Dissertation Submitted to

KU Leuven and Tsinghua

in partial fulfillment of the requirement

for the degree of

Doctor of Philosophy

in

Business Economics

by

Wang Qian

Dissertation Supervisor : Jan Dhaene

May, 2019

ACKNOWLEDGEMENTS

First of all, I would like to thank my father, Wang Shiqiang, who helps me to foster in the outlook on the world and life, face life honestly and bravely, and cultivate the habit of thinking. I would also like to thank my mother, Chen Zhuping, for bringing me into this world, for letting me know the world from books, and for helping me develop my study habits. Without what they taught me, I would not have had the opportunity to come to Tsinghua to complete my doctoral studies. They help me all the time, and I love them.

I sincerely thank my promotors, Professor Jan Dhaene of KU Leuven and Professor Gaofeng of Tsinghua University for their careful guidance. Their words and deeds will benefit me all my life. Professor Jan Dhaene is a smart and interesting person. He is always very good to students and gives us academic guidance and help. It is very important and precious for me to have him point out the direction of actuarial research and give me sufficient guidance. Thanks to Jan for all his help, I spent a happy year in Belgium. I still miss my days there and hope to have the opportunity to go back to study and do research in the future. Thank you, Professor Jan Dhaene of the University of Leuven, Belgium, for your kind guidance and help, as well as Daniel, Karim and Hamaza of the research group.

Associate Professor Gao Feng always answers all my questions carefully and patiently, and patiently gives me a lot of guidance, so that I slowly develop the thinking and ability of academic research. Under the careful cultivation of the instructors, I was at a loss from the beginning and gradually built up my research self-confidence.

Thank the friends from the insurance company for their help, so that this article can be successfully completed. Thank you for your help. Let me never have any background in economics mathematics students, step by step familiar with economic research, and finally formed this article.

Thank you to Tsinghua University and KU Leuven for their five-year doctoral years. I met many interesting friends, elder students and teachers. All these are colorful imprints in my life. Every section of the road is a kind of insight.　Thank you for coming. Everything is the best arrangement for me.

I used to see the passages of the award-winning speech on the Internet are all the sentences at the beginning of thanks.　Although I am not award-winning, in retrospect of all the past, too many people need to thank. Everyone's life has its own ups and downs of the plot, I am grateful to be able to get this script, because I met you in life.

Wang Qian

May 20th 2019

Abstract

Air pollution is a serious problem in China, which reflects the contradiction between economic development and environmental protection. It is of great practical significance to study the impacts of air pollution. The term, "beautiful China", appeared three times in the report of the Nineteenth National Congress of the Communist Party of China, which aroused wide public debates. Studying the hazards of air pollution will help us to realize the seriousness of the problem and take appropriate measures to reduce or eliminate the hazards of air pollution. Using the insurance policy data of a large domestic insurance company and air pollution data of cities above Prefecture level, this dissertation empirically studies the effects of air pollution on human physical health and mental health, insurance purchase decision-making, and the work efficiency and competencies of analysts, respectively. The results of empirical regression were used to estimate the excess prevalence rate and excess life expectancy loss caused by air pollution. The related direct economic loss was estimated by welfare loss, and the economic loss of air pollution impairing cognitive ability and reducing work efficiency was also estimated. Estimating the economic losses caused by pollution problems can provide references for the formulation of environmental policy.

Firstly, this thesis uses a cohort study of more than 20 million people to explore the impacts of air pollution on human health. The results of COX proportional hazard regression showed that PM2.5 pollution resulted in 1.6 years of life expectancy loss per capita. The direct economic loss of PM 2.5 caused by excessive illness amounted to 100 billion yuan, accounting for at least 0.3% of GDP. The probability of serious diseases, malignant tumors and lung cancer increased by 12%, 19% and 75% respectively with the change of PM2.5 concentration of 10 $\mu g/m^3$. The problem of air pollution should be paid great attention by relevant departments. From the perspective of public health, environmental policies should be formulated to ensure air quality and people’s life and health. Secondly, air pollution can lead to anxiety, depression and even increase the risk of suicide. For change in every standard deviation of the air quality index (AQI), the suicide rate increased by 26%, and about 34,000 people committed suicide because of air pollution. More attention should be paid to the psychological impacts of air pollution and its harm should be actively prevented. Thirdly, using COX proportional risk regression and likelihood unrelated regression, this thesis studies the influencial factors of insurance purchase decisions, the insured's risk level and the city's air pollution level, which both affect the insurance purchase decisions. Empirical results show that long-term air pollution will increase insurance purchases. Finally, the impacts of our air pollution on brain-workers. The results show that the air pollution level increases the possibility of errors in analysts' reports and decreases the accuracy of forecasts.

Serious air pollution affects human health level. A wide range of air pollution problems throughout the country will have an economic impact on the macro market through the economic activities of various economies. Economic development should not be at the expense of environmental pollution, but realizing green development. There is a long way to go to solve the existing environmental problems and build ecological civilization in China.

Key words: air pollution; cancer; mental health; decision-making; analyst prediction

CONTENTS

Chapter 1: INTRODUCTION 8
1. 1 Research background and problems 8
1. 2 Research Significance... 11
1. 3 Research methods 12
1. 3. 1 Cohort study method 12
1. 3. 2 Other research methods 13
1. 4 Innovation and Research Contribution 14
1. 5 Article structure 16
CHAPTER 2: The Effects of PM2.5 Concentration on the Morbidity of Lung Cancer 19
2.1 Introduction 19
2.2 Data Description 23
2.2.1 Data Source 23
2.2.2 Descriptive statistics 26
2.2.3 Correlation coefficient matrix 27
2.3 Empirical analysis.. 28
2.3.1 Regression model.. 28
2.3.2 Empirical results of lung cancer 29
2.3.3 Poisson regression... 35
2.3.4 Robustness test 37
2.3.5 Evaluation of the welfare loss 40
2.4 Conclusion... 45
Chapter 3: The effect of air pollution on suicide rate............................. 46
3.1 Introduction 46
3.2 Data and descriptive statistics 49
3.2.1 Data sources 49
3.2.2 Descriptive statistics 52
3.2.3 Matrix of correlation coefficient...................................... 52
3.3 Empirical Analysis... 53
3.3.1 Logistic regression model 53
3.3.2 Regression results 54

3.4 Conclusion 61

Chapter 4: Air Pollution and Insurance Purchase Decision 62

4.1 Introduction 62

4.2 Data and descriptive statistics 66

4.2.1 data sources 66

4.2.2 descriptive statistics 70

4.2.3 Correlation matrix 71

4.3 Empirical Analysis 72

4.3.1 COX proportional hazard model 72

4.3.2 regression model 73

4.3.3 Empirical results 74

4.3.4 Seemingly Unrelated Regression and Empirical Results 85

4.3.5 The Regression of Urban Policy Purchase 88

4.4 Conclusion 90

Chapter 5: Buy, Hold or Sell? Air Quality and Financial Analyst Reports 92

5.1 Introduction 92

5.2 Description of Data 94

5.3 Results 103

5.3.1 Prediction accuracy regression results 103

5.3.2 Analysis Report Error Regression Results 111

5.3.3 Operational risk and loss 115

5.4 Conclusions 117

BIBLIOGRAPHY 118

Chapter 1: INTRODUCTION

1. 1 Research background and problems

Many of the world's heavily polluted cities are in developing countries. These cities bear a heavy burden of manufacturing economic activities, and use relatively old technologies in heating systems and other modern necessities, so the air pollution problem in these cities is more serious. Due to the development of the world economy and the increasing dependence on the economic impetus of developing countries, it is of great economic significance to study the impact of air pollution.

This dissertation studies the effects of air pollution on people's physical health and behavioral decision-making. By discussing the various effects of air pollution, we can evaluate the hazards of air pollution more objectively and comprehensively. Estimating various welfare losses caused by air pollution from an economic point of view is helpful to the formulation of environmental policies and to the establishment of environmental supervision and management system. The problem of air pollution should arouse the attention of relevant departments. Considering from many angles, such as public health and economic efficiency, we should not develop our economy at the expense of environment and people's health. Instead, we should promote the healthy and sustainable development of China's economy on the basis of new driving force and new economic growth point.

Literature shows that PM2.5 concentration is significantly correlated with lung cancer, cardiovascular mortality, and total mortality (Ma et al., 2011; Xie et al., 2015; Schwartz et al., 2002; Yang et al., 2012; Kasiscovick and Miller, 2007; Naess et al., 2007; Turner et al., 2011), and reducing PM2.5 concentration can reduce mortality risk (Laden et al., 2006). Relevant studies of natural experiments also show that air pollution will increase mortality, hospitalization, cancer mortality and cardiovascular mortality (Anderson and Michael, 2015; Chen et al., 2013; Rd and Pope, 1989; 3Rd and Pope, 1996). Because the air pollution concentration in China is higher than that in other countries, the long-term exposure reaction at high concentration may not be consistent with that at low concentration, but the related research is limited. In Chapter 2, we studied the effects of long-term exposure to high concentration PM2.5 on the incidence of malignant tumors, critical and severe diseases, nervous system, respiratory system, endocrine system, immune system and other diseases. The risk ratio of PM2.5 concentration is estimated, and the economic loss caused by air pollution is estimated by the risk ratio of regression.

Air pollution affects not only physical health but also mental health. When air pollution was serious, the number of first aid calls for psychiatric patients in the United States would increase (Rotton and Frey, 1984; Rotton and Frey, 1985). Air pollutants contain photochemical oxidants, which increase the number of people anxious (Gary et al., 1983; Cavanagh et al., 2003; Lim et al., 2012). Air pollutants are neurotoxic and can damage the nervous system (Brook et al., 2011). The threat of haze to mental health is still neglected in China. Chapter 2 also points out that air pollution increases the risk of neurological diseases. Studies have shown that high concentrations of particulate matter increase the risk of suicide (Fleehart et al., 2014; Jee et al., 2011; Ha et al., 2015; Bakian et al., 2015). Chapter 3 studies the relationship between air pollution and suicide in China, weather condition can affect people's emotions (Howarth and Hoffman, 1984), so we control meteorological data in our study.

The results of Chapter 2 show that air pollution increases the risk of various diseases. Exposure to heavily polluted air also raises concerns about the risk of illness, and may even lead to anxiety and depression (Arvin and Lew, 2012; Marques and Lima, 2011). People may cope with the risk of illness by purchasing commercial health insurance, and purchasing a guarantee can also alleviate anxiety and anxiety to a certain extent. Since the increased risk of illness increases the purchase of commercial health insurance, it is a risk but also a development opportunity for insurance companies. In Chapter 4, we study the impact of air pollution on insurance purchase decision-making. First, we study the impact of air pollution on whether to buy commercial insurance. Prospect theory shows that insurance consumption decision-making is directly related to the risk perceived by insurers, but the uncertainty perceived by consumers is not necessarily consistent with the objective probability. People mainly rely on intuition to judge and estimate the risk of things, but limited by memory and information, they usually generalize in a partial way. Risk perception is an individual's subjective understanding of external objective risks. The dissemination of negative news and information will increase people's fear and may increase individual risk perception. In prospect theory, people have a subjective evaluation of objective probability, and in decision-making, they will make the best choice based on the subjective risk. Air pollution existed in China many years ago, but according to the search volume of search engines, the search volume of air pollution peaked at the end of 2013, and has been higher since then. It can be seen that the public's real understanding and concern about haze and air pollution began at the end of 2013. Haze weather has aroused widespread concern in society. People begin to realize the haze's harm to health and worry about its impact on health. Health risk of air pollution has existed objectively for many years, and subjective risk perception has changed. When the level of individual risk perception rises, the subjective probability evaluation of a certain risk rises, so the demand for insurance increases. Chapter 4 confirms the prospect theory with the data of China's commercial health insurance. This Chapter studies the subjective risk perception and objective risk, and their impact on purchase decision-making respectively. In addition to studying the relationship between risk perception and purchase decision-making, the fifth chapter also uses the posterior claim status as the basis to judge the risk degree of the insured. By investigating the insurance purchase decision-making of high-risk groups, the Chapter empirically examines the problem of information asymmetry in the insurance market. It proves that adverse selection exists in China's health insurance market, and provides reference for product design of insurance companies.

How environmental condition affects decision-making and market operation is an important issue. Some papers have explored the impact of pollution on the labor market. Particulate matter concentrations in the air (Chang, Graff Zivin, Gross and Neidell, 2016; Lichter, Pestel and Sommer, 2017; He, Liu and Salvo, 2019), and increased ozone concentrations (Graff Zivin and Neidell, 2012) will reduce the productivity of manual workers. There are few studies on the impact of air pollution on non-manual labor. Air pollution can reduce typing efficiency, increase cognitive difficulty, increase errors (Lagercrantz et al., 2000), and also reduce work efficiency (Chang et al., 2019). In Chapter 2, we found that air pollution increases the risk of neurological diseases. Other studies have confirmed that air pollution is neurotoxic, impairs the nervous system (Brook et al., 2011) and impairs cognitive function (Zhang et al., 2018). In today's economy and society, more and more work is non-manual. The majority of work requires cognitive ability. The potential harm of air pollution to human cognitive ability will be reflected to the market and macroeconomic level through economic activities, and its impact should not be underestimated. The majority of work in the financial industry is cognitive work. Chapter 5 studies the impact of air pollution on the efficiency and ability of market agents through the analysis of the accuracy of analysts' forecasts and reporting errors.

Similar effects may also exist on other non-manual workers. In Chapter 5, the potential market impact is assessed by calculating the operational risk based on the estimated relationship between air pollution and error rate.

1. 2 Research Significance

The air pollution problem in China is severe, which has aroused widespread concern of the whole society. It is of great practical significance to study the impact of air pollution. Serious air pollution affects health level and further affects production and living activities. Large-scale air pollution problems throughout the country will have an economic impact on the macro-market through the economic activities of various economies. Over the past decades, China's economy has developed rapidly, even at the expense of the environment to achieve economic benefits. Estimating the economic losses caused by pollution can provide reference for the formulation of environmental policies. Only fully aware of the impact of air pollution, can we take appropriate measurement to reduce or eliminate the harm of air pollution.

Firstly, it has been widely accepted that air pollution will do harm to health, and PM2.5 has been hotly discussed. It is essential to study the effect of long-term exposure to high concentration PM2.5 on the incidence of malignant tumors and critical diseases. In the past 30 years, lung cancer mortality in China has increased by 465%, making it the most dangerous cancer in China (Lee, 2010). The high incidence of cancer, especially the dramatic increase of lung cancer mortality, makes it more important to study the correlation between air pollution and morbidity. In this chapter, we study the pathogenic risk of PM2.5. The correlation between air pollution and incidence is studied from respiratory system, cardiovascular and cerebrovascular system, immune system and nervous system.

Secondly, according to the China Death Cause Surveillance Data Set published by the China Center for Disease Control and Prevention, it can be estimated that the number of suicides in China is about 78,000. Based on the conclusion that air pollution has neurotoxicity, air pollution may make depression worse and increase the probability of suicide. In this chapter we study the relationship between air pollution and suicide, so that people can take appropriate measures to prevent suicide and save lives. This study confirms the impact of air pollution on mental health. Relevant departments can provide appropriate early warning actions to prevent people from risk. At the same time, patients with mental illness should seek professional help in time in the haze.

Third, air pollution will objectively increase the risk of critical diseases and malignant tumors. Subjectively, people do not recognize the risk and will not affect the objective existence of the risk. From the perspective of insurance purchase, this chapter explores how subjective cognition and objective facts affect behavioral decision-making. Based on this, the decision-making mechanism can be deeply understood. Everyone is facing a large number of economic decisions every day. Every decision is a part of the social and economic operation. Understanding the decision-making mechanism of human beings has significant economic value. In addition to the decision-making mechanism, in this chapter we also discuss the information asymmetry in the insurance market, which provides references for the supervision of commercial health insurance market and the development strategy of insurance companies in China.

Fourth, air pollution can also damage the nerve system, brain function and cognitive function (Brook et al., 2011). Nowadays, with the social division of labor environment, more and more work needs to be done with cognitive ability. Large-scale air pollution will inevitably affect the cognitive ability of numerous workers. Studying the impact of air pollution on labor productivity is of great significance to the study of air pollution on social economy. Air pollution will affect the cognitive ability of mental workers, thus reducing the efficiency of their complex work. Starting from the financial market, this chapter evaluates the quality of analysts' reports and studies the impact of air pollution on labor productivity. The research expands the existing research on air pollution. The economic loss of air pollution is estimated from the perspective of operational risk.

1. 3 Research methods

1. 3. 1 Cohort study method

In this study, long-term exposure to PM2.5 was studied in a cohort. COX proportional risk model and Poisson model were used for regression. Among them, COX proportional hazard model assumes hazard function or immediate death probability. $\lambda(\text{survival time})$It can be decomposed into a benchmark hazard function.

$\lambda_0(survival\ time)$And proportional risk functionΦThe product ofΦLogarithmic linear functions that can be expressed as related variables:

$$\frac{\lambda(\text{survival time})}{\lambda_0(\text{survival time})} = \phi(PM2.5, personal, control, city)$$

In addition to COX proportional hazard regression, Poisson regression is often used in cohort studies. In Chapter 2, Poisson regression method is used to ensure the reliability of the conclusion.

1. 3. 2 Other research methods

In addition to the two regression methods commonly used in the above two cohort studies, this dissertation mainly uses the following five methods for empirical research. (1) Similar unrelated regression method is used to study the product selection of the insured in Chapter 4. Because policyholders need to decide the insurance amount, premium rate and insurance period at the same time, these three decisions are not independent of each other, so the similar unrelated regression method is used to estimate three explanatory variables at the same time. (2) Fixed-effect regression method, different cities have different levels of air pollution, in addition, there are many differences, but can not find the corresponding variables for these differences. In order to solve the problem of missing variables, urban fixed effect and time fixed effect are added to the regression. (3) The instrumental variable method, because the impact of air pollution on human health is endogenous, the regression coefficient may be biased. Therefore, meteorological data is used as a tool variable for regression to alleviate the endogenous problem in order to obtain unbiased regression results. (4) Logistic regression method, Chapter 3 and Chapter 5 all have the regression of dependent variable 0-1, so logistic regression is adopted. (5) OLS regression method.

In addition, Jiang Qinglang's life table algorithm is used to calculate life expectancy and cause-of-death life expectancy. The loss of labor force is also estimated from the perspective of operational risk.

1. 4 Innovation and Research Contribution

The air pollution problem in China is severe. Seven of the ten cities with the worst air quality in the world are in China. Studying the hazards of air pollution will help us to realize the seriousness of the problem and provide a reference for policy making. It is of great practical significance to study the impact of air pollution. From the perspective of mental health and behavioral decision-making, in this dissertation we make a more comprehensive study of the impact of air pollution on individuals, and makes up for the shortcomings of existing research. In addition to affecting physical health, air pollution has an impact on the whole social and economic activities. The losses caused by air pollution are enormous and more attention should be paid. It is urgent to formulate environmental protection policies. Next, we will elaborate on the innovations and research contributions of each chapter.

There have been a lot of studies abroad confirming that air pollution can affect human health, but air pollution in China is more serious. PM2.5 concentration is much higher than that in foreign countries. Long-term exposure under high PM2.5 concentration still needs to be further explored. Natural experiments on the dividing line of heating policy of the Huaihe River in Qinling Mountains were carried out at the urban level, with the population as the research object, ignoring individual differences (Chen et al., 2013; Ebenstein A et al., 2013). There are few studies on the impact of air pollution on health in China. Due to the lack of PM2.5 detection data, early related variable statistics and major disease incidence data, epidemiological and health effects research is limited, and related research is scarce. Specific innovations are as follows: First, the impact of air pollution on human health from the personal level, control the impact of personal characteristics. Secondly, we studied critical and severe diseases, malignant tumors, lung cancer and other cancers. More over we studied a wide range of diseases and assessed health risks from multiple latitudes. Thirdly, because the incidence of chronic diseases is difficult to obtain, the CDC, the Health Planning Commission and other departments do not have statistical data on the incidence of chronic diseases. The existing studies are all about mortality. Because of the improvement of medical level, only studying mortality rather than morbidity will underestimate the harm to a certain extent. This chapter uses the national insurance policy data, and contains nearly 80% of the national commercial health insurance policies. The data set are large and cover a wide range. The results are more credible than before with eight years observation period. Using the claim data of the commercial health insurance, the incidence of chronic diseases is studied, and the harm of air pollution is estimated more accurately. Fourthly, China's early air quality monitoring standards and systems are not perfect, and there is no long-term air-monitoring data. Especially the concentration of PM2.5 has not been monitored until 2014, but foreign studies show that PM2.5 has a greater health risk than other pollutants. In this chapter, the PM2.5 concentration data retrieved from satellite data is used, which avoids the risk of artificial manipulation and covers a wide range. This is the first study to examine the health risk level of long-term exposure to high concentrations of PM2.5. Fifthly, the excess morbidity and life expectancy loss were estimated. From the perspective of economics, the loss of health and welfare caused by air pollution is estimated at the macro level, which provides a quantitative reference for the formulation of environmental policy.

Air pollution affects not only people's health, but also their mental health. European and American countries have a clear understanding of mental illness, and will make reasonable suggestions for this weather in weather forecasting. There is little attention paid to mental health in China. There is almost no research on the relationship between air pollution and mental health in China, and corresponding early warning measures have not been set up. But according to the online health survey of domestic portals, haze can make people fear, anxious and depressed. Some counseling centers in China claim that patients in haze days will increase by about 10% than usual. People with mental illness will feel unhappy because of the haze. There have been some studies on the correlation between air pollution and suicide abroad. High concentration of particulate matter can increase the risk of suicide. Some studies only consider the correlation of peak values and do not control the influence of other factors. Domestic research in this area is almost blank, and there is little mention of psychological intervention in fog and haze weather. In this chapter, meteorological data and other factors are controlled to reduce the problem of missing variables. Studying the relationship between air pollution and mental health fills the gap in related research fields in China.

Regarding the impact of air pollution on policy purchase, relevant studies (Chang et al., 2018) used the hourly data of urban PM2.5 to regress from the urban level, confirming that air pollution would increase the purchase of health insurance on the same day in cities. This study neglects personal characteristics, and the data sample only comes from three cities, which is not representative enough. The explanatory variable is the purchase amount of insurance policies, without considering the different insurance coverage levels of different cities. The innovations of this chapter are as follows: Firstly, we study the relationship between air pollution and insurance purchase from the personal level, controlling personal characteristics. Secondly, subjective risk is distinguished from objective risk, and the impact of air pollution concern on health insurance purchase decision-making is analysed. It is found that subjective risk perception will affect purchase decision, but objective risk will not.

Although objective risk will not affect the health insurance purchase decision-making, it will affect people's cognitive ability and work efficiency. This potential harm has not been paid attention to, but affects every individual in social and economic activities, thus causing widespread impact. Starting from the financial market, this chapter studies the impact of air pollution on agents in the market. Analysts are the main information transmitters in the security market. The research shows that the accuracy of analysts' forecasts is affected by the variables such as the length of the forecast interval, the personal characteristics of analysts, the characteristics of security dealers and the characteristics of listed companies studied. The innovation of this chapter is as follows: Focusing on the impact of air pollution on work efficiency, this chapter uses the errors of the analysis report to measure the cognitive ability and the degree of caution of the analysts, and takes the quality of work as the measure index of work efficiency. The cross-validation of error rate and prediction accuracy further confirms the negative impact of air pollution on analysts' work efficiency. At present, there are few related studies. And this chapter makes up for this gap. From the perspective of operational risk, the loss caused by the increase of error rate is estimated.

1. 5 Article structure

The structure of this dissertation is as follows: Chapter 1 is an introduction, which introduces the background, significance, research methods and contributions of air pollution related research, and finally introduces the structure and research logic of this dissertation.

In Chapter 2, the long-term exposure response of PM2.5 was studied by survival analysis. The results showed that PM2.5 pollution increased the risk of malignant tumors such as lung cancer. The impact of air pollution on human health has been basically confirmed. There are few related studies in China. Because air pollution in China is more serious than in other countries, it is of great significance to study the impact of high concentration PM2.5 on human health. The COX proportional hazard model and Poisson regression model were used to study the added value of the risk of serious illness and death for every 10 μg/m3 increase in air pollutant concentration. Furthermore, the author estimated the life expectancy loss and welfare loss caused by air pollution with regression results.

Chapters 3, 4 and 5, using the data of air pollution index AQI, discuss the influence of short-term air pollution on people's behavior, mood, mental health, decision-making and work efficiency. Chapter 3 introduces the impact of air pollution on mental health, using AQI daily value data of cities and daily suicide data. To study the correlation between air pollution level and suicide rate on the same day.

The fourth chapter introduces the relationship between air pollution index and insurance purchase, using AQI daily value data of the city published by the Ministry of Environmental Protection and insurance policy data of a large domestic insurance company. The COX proportional risk regression model was used to analyze the cohort, and the influencing factors of the policy-holder's insurance purchase decision were studied from the individual level. Subjective risk perception will increase the purchase demand of health insurance, while objective risk will not affect the purchase decision. Similar unrelated regression model is used to further verify the conclusion. It is found that air pollution also has an impact on the selection of insurance amount, insurance period and premium rate.

Chapter 5 mainly introduces the impact of air pollution on analysts' prediction. Similarly, using AQI daily data, the relationship between air pollution index and analysts' prediction accuracy and reporting error rate is studied. It was found that the air pollution level in the days before the report was released increased the misalignment rate of the report. Although analysts did not write the report abnormally, the quality of their work declined. It shows that air pollution can reduce productivity and thus have an impact on the economy.

In the four chapters, the annual data of PM2.5 concentration and the daily data of air quality index are used to study the effects of air pollution from both long-term and short-term perspectives. The long-term impact is mainly the relationship between PM2.5 concentration and health risk. The short-term impact studies the impact of air pollution from mental health, risk perception, cognitive ability, productivity and other aspects.

Overall, this dissertation studies the effects of air pollution on physiology and behavior. The structure of this dissertation is shown in Figure 1. The first part is about the effects of air pollution on physiology, which is mainly divided into two aspects: physical health and mental health, including chapter 2 and 3. The second part is the impact of air pollution on behavior. As the first part points out, air pollution can endanger people's physical and mental health. Subjective perception of the risk of air pollution may affect people's behavior. Based on this hypothesis, we study the impact of air pollution level on insurance purchase decision-making in Chapter 4. Subjective risk perception will affect people's risk attitude and increase insurance demand. In addition, existing studies have found that air pollution can reduce people's cognitive ability and labor productivity. The first part of the study also pointed out that air pollution can lead to critical and severe diseases, so that people lose their labor force, and air pollution can affect psychology, but also affect work efficiency. On this basis, Chapter 5 studies the impact of air pollution on market agents from the perspective of financial markets. That is to say, to study the influence of air pollution on the accuracy of analysts' prediction and the error rate of reporting. This dissertation discusses the impact of air pollution on individuals, which will have a far-reaching impact on the market and macroeconomic level through everybody's daily economic activities.

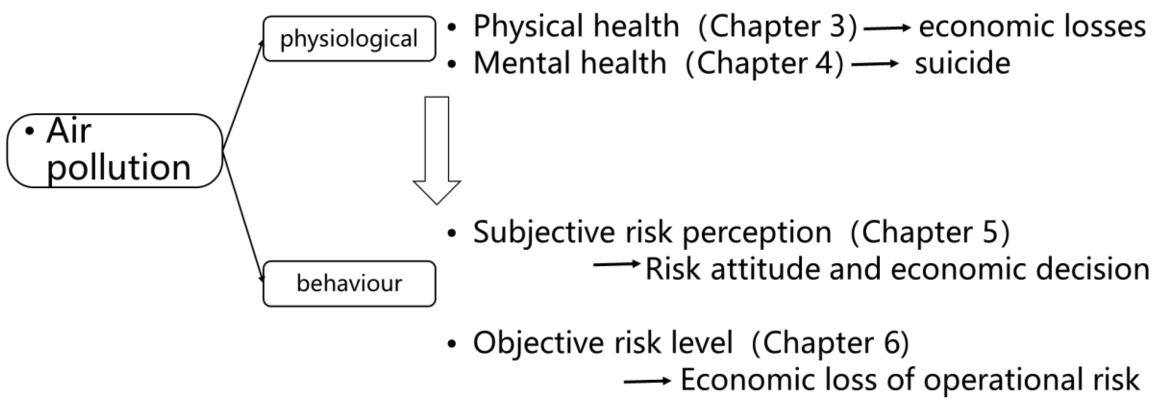

Figure 1 dissertation framework

CHAPTER 2: The Effects of PM2.5 Concentration on the Morbidity of Lung Cancer

Abstract: air pollution is a serious problem in China, which reflects the contradiction between economic development and environmental protection. It is of great practical significance to study the impacts of air pollution. Studying the hazards of air pollution will help us to realize the seriousness of the problem and take appropriate measures to reduce or eliminate the hazards of air pollution. This paper uses a cohort study of more than 20 million people to explore the impacts of air pollution on human health. The results of COX proportional hazard regression showed that PM2.5 pollution resulted in 1.6 years of life expectancy loss per capita. The direct economic loss of PM 2.5 caused by excessive illness amounted to 200 billion yuan, accounting for at least 0.3% of GDP. The probability of serious diseases, malignant tumors and lung cancer increased by 12%, 19% and 75% respectively with the change of PM2.5 concentration of 10 $\mu g/m^3$. Serious air pollution affects human health level. The problem of air pollution should be paid great attention by relevant departments. From the perspective of public health, environmental policies should be formulated to ensure air quality and people's life and health. There is a long way to go to solve the existing environmental problems and build ecological civilization in China.

2.1 Introduction

In recent years, as the public becomes more aware of environmental protection, air pollution is attracting significant attention in China. However, the degree to which air pollution affects human health is unknown. Many foreign studies have investigated the effects of air pollution on human health and its mechanisms (Kampa et al. ,2008); however, the conclusions may not apply in China because of heavier air pollution.

Studies have shown that air pollution affects health, and the effect is related to the type and concentration of pollutant, individual susceptibility, and exposure length (Franklin et al. ,2015). Air pollutants may enter the human body via expiration, ingestion, or skin contact; these toxic substances circulate in the bloodstream and eventually deposit in different tissues. Air pollution may affect respiratory, cardiovascular, nervous, urinary, and digestive systems and impair organ function, thus affecting all aspects of human health (Pope et al. ,2004; Brook RD et al. ,2010).

Studies on the effects of short-term exposure to air pollution focus on certain unexpected short-term events. For example, during the Olympics, the host country usually takes measures to ensure the air quality. The short-term improvement in air quality provides a natural experiment to investigate the effects of short-term exposure to air pollution on health by analyzing the morbidity, admission rate, and mortality before, during, and after the Olympics (Friedman et al. ,2001; Kipen et al. ,2010; Su et al.,2016; Su et al. ,2015; Rich et al. ,2012; Rich et al. ,2015). Many foreign studies have shown that PM2.5 is significantly correlated with short-term morbidity and mortality (Schwartz et al. ,1996; Schwartz et al. ,2000; Petersen and Floyd, 2000; Anderson et al. ,2001; Brook et al. ,2011; Ai et al. ,2014; Zanobetti and Schwartz,2009; Ostro et al. ,2006). In recent years, studies on the effects of short-term exposure to air pollution in China (Zhang et al. ,2008; Chen et al. ,2011; Bank et al. ,2007) also show that PM2.5 concentration is significantly correlated with mortality (Ma et al. ,2011; Xie et al. ,2015; Schwartz et al. ,2002; Yang et al. ,2012), which is consistent with the findings of similar foreign studies. However, the correlation of short-term exposure is more about the harvesting effect of susceptible populations and cannot reflect the long-term cumulative effects of pollution on health. Given the "harvesting effect" and "cumulative effect", the wealth of current findings on the health effects of short-term exposure (Lu et al. ,2015) is not necessarily applicable to long-term exposure, nor can the results of short-term exposure be used as the upper or lower limits of long-term exposure.

Many countries have introduced policies to improve the environment, but it takes a long time to see the effects. Thus, studies on long-term exposure are more valuable references than those on short-term exposure for developing policies. However, for studies on long-term exposure, it is more difficult to assure no change in other factors associated with a health effect. Moreover, it is difficult to prove the causation between air pollution and mortality and morbidity. In addition, the effects of long-term exposure on the human body cannot be inferred from the results of short-term exposure. If air pollution has only a short-term health effect, then the effect may vary greatly, with no correlation between long-term air pollution and health. The effect of long-term exposure can be proven only if air pollution is correlated with mortality with low variation.

The Harvard six city study is the most classic and influential cohort study on this topic. The study showed that smoking was the variable most correlated with mortality. After controlling for smoking and other risk factors, mortality was significantly correlated with air pollution. Air pollution is positively correlated with mortality due to lung cancer and cardiopulmonary diseases, but it is unrelated to mortality caused by other diseases. Moreover, mortality is mostly correlated with fine particles. The study first proposed that long-term exposure is associated with increased mortality from cardiovascular diseases. The mortality is 26% higher in the most polluted city than in the least polluted city (Dockery et al. ,1993).

Since then, many researchers have expanded the Harvard six city study. A 2002 cohort study covered nearly all major cities in the US and involved nearly half a million participants. The study used a method similar to that of the Harvard six city study to analyze the 16-year follow-up data and controlled for active smoking. The results showed that long-term exposure to PM2.5 produced by combustion is an important environmental risk factor for cardiovascular- and lung cancer-related deaths. After controlling for covariates such as diet and pollutants, the cardiovascular mortality was shown to increase by 6% with every increase of 10 μg/m³ in the annual average PM2.5 concentration (Iii et al. ,2002). After controlling for active smoking, the PM2.5 concentration is positively correlated with overall mortality, cardiovascular mortality (Kasiscovick and Miller, 2007), and lung cancer-related mortality. Reducing the PM2.5 concentration reduces the risk of death (Laden et al. ,2006). Data without controlling for a history of smoking also showed that PM2.5 is positively correlated with lung cancer-related mortality (Naess et al. ,2007).

In 2013, an European cohort study of 300 000 individuals showed that the risk of lung cancer increases by 1.18-fold with every increase of 5 μg/m³ in PM2.5 concentration (Raaschou et al. ,2013). A follow-up study (average follow-up time: six years) in 60 000 postmenopausal women from 36 metropolitan areas in the US showed that for every increase of 10 μg/m³ in PM2.5 concentration, the risks of cardiovascular events, cerebrovascular events, and cardiovascular mortality increased by 24%, 35%, and 76%. In a cohort study sponsored by the American Cancer Society, 1.2 million U.S. adults were followed up for 26 years (1982-2008). The results showed that for every increase of 10 μg/m³ in PM2.5 concentration in the air, lung cancer-related mortality increased by 15% to 27% and is more pronounced in patients with chronic lung diseases (Turner et al. ,2011). Michelle et al screened 1.2 million participants to select 188 699 never-smokers and showed that for every increase of 10 μg/m³ in long-term PM2.5 concentration, lung cancer-related mortality increased by 19% to 30%(Turner et al. ,2011). The results showed that for every increase of 10 μg/m³ in the average 10-year PM2.5 concentration in urban areas, the risk of lung cancer was increased by 80%, the overall mortality was increased by 6%, and the cardiovascular mortality was increased by 11% (Hoek G et al. ,2013).

Cohort studies show that air pollution is related to health, and natural experiments further confirm the causation between air pollution and health. Few residents at Utah Valley smoke thanks to their religious beliefs; as a result, the local steel mill is the main source of PM10. During the 14 months between 1985 and 1988, the large steel mill was shut down because of worker strikes. Analyses of air pollution and related health data during the period showed that the PM10 concentration was highly correlated with admission rate, with a significant effect on the admission rate of diseases related to asthma and bronchitis (Rd and Pope, 1989). During the shutdown, the particulate concentration, admission rate of respiratory diseases, and mortality were significantly reduced. After the dispute was resolved, all relevant indicators returned to pre-strike levels (Rd and Pope, 1996). The evidence from Utah Valley shows that exposure to fine particles is closely correlated with mortality and morbidity. Moreover, the correlation is observed after a delay of several months, indicating that air pollution has a cumulative effect on human health.

Given the effect of wind (downwind and upwind) on the concentration of pollutants along the highway, some researchers investigated mortality associated with downwind versus upwind along the highways of Los Angeles, USA. The results showed that increased exposure to air pollution increases mortality (Anderson and Michael ,2015).

Some researchers conducted a natural experiment according to different heating policies along the Qinling Huaihe line in China. Given the spatial discontinuity of PM concentration, a breakpoint regression method was used and showed that PM pollution reduces the life expectancy by 5.5 years. Moreover, cardiovascular mortality shows concordant spatial discontinuity with PM concentration, indicating that increased cardiovascular mortality is a major cause of shorter life span (Chen et al.,2013). Some studies have been conducted to investigate the relationship between air pollution and lung cancer in China. However, most studies investigated population health in regions with different concentrations of pollutants without controlling for other variables. Moreover, the population was analyzed as a whole with no regard for individual variation.

After controlling for certain risk factors for lung cancer, such as smoking and indoor pollution, the morbidity of lung cancer was shown to be 1.3 times higher in areas with heavy air pollution than in areas with less heavy pollution. Some researchers performed a grey correlation analysis of air pollution data and lung cancer morbidity and mortality data and concluded that PM2.5 induces lung cancer after a latent period of approximately eight years. Some researchers used confirmed hospital deaths and population composition to calculate the mortality in three districts in Fujian and found that lung cancer-related mortality is positively correlated with total suspended particles (TSPs). However, the researchers did not control for individual or other factors, and thus the results may not be applicable to other cities.

Some researchers have used the Poisson regression model and geographic weighted regression model to investigate the risk factors for lung cancer. The primary explanatory variables were meteorological and air pollution data, and the results showed that air pollution is a risk factor for lung cancer. However, some control variables in the study lack long-term monitoring data and thus cannot reflect the long-term effects of pollution.

In this study, we performed a survival analysis to analyze the effect of PM2.5. To investigate the long-term effects of air pollution on human health, we analyzed the data of tens of thousands of policies on critical and severe diseases and claims in the past decade from a major insurance company.

The sections below will detail the following information: Section II reviews current research in this field; Section III introduces data and data processing methods; Section IV introduces empirical analysis methods and regression models; Section V describes the results of the empirical analysis and provides a robust analysis of the results; and Section VI provides a conclusion and policy recommendations.

2.2 Data Description

2.2.1 Data Source

2.2.1.1 Policy data

In this study, the individual disease-related data were obtained from anonymous policy data provided by a Chinese insurance company. The policy was sold from 1999 through 2009. Long-term valid policies covering critical and severe diseases during the study period (January 1, 2006 - December 31, 2013) were collected, including 24 716 981 policies and 957 388 claims.

According to the policy data, we determined the policy holder's age, gender, place of residence, and occupation. The occupations were classified into eight categories: I: head of administrative agencies, enterprises, and non-profit organizations (no field work); II: specialized and technical personnel; III: clerk and related personnel; IV: commercial and service personnel; V: agriculture, forestry, animal husbandry, fisheries, and water conservancy personnel; VI: manufacturing and transportation equipment operators and related personnel; VII: military personnel; and VIII: other (such as unemployed).

From the claims data, particularly the details of the reason for claims and the judgment of the insurance company, we collected the illness and death information of the policy holders. If claims were paid for a policy, the reason was coded as "1" for lung cancer and "0" for all other reasons and all policies without claims. For policies with claims, the date of payment was regarded as the end of the study period. For policies without claims, the end of the study period was December 31, 2013. The time from the date of birth to the end of the study period was the "survival time" of the policy holder. In our model, the end of the study period could be either December 31, 2013, in which case no additional lung cancer cases were recorded after December 31, 2013 (censored data), or missing if the payment was made for other reasons (censored data). The date of payment for claims, in which case the data were before December 31, 2013(complete data) and if the payment was made for lung cancer (end-point event). The survival time was defined as the time from the date of birth to the end of follow-up or the time of the end-point event.

2.2.1.2 Pollution data

This study investigated the effects of air pollution on human health, and the primary explanatory variable was PM2.5 concentration. Previous studies used air quality indices or high-altitude PM2.5 concentration data from the National Aeronautics and Space Administration (NASA). We did not use air quality indices because of issues with monitoring period and coverage. We also did not use high-altitude PM2.5 concentration data because the data cannot accurately reflect the exposure level close to earth, where most human activities occur. Finally, we used ground-level PM2.5 concentration data from Martin et al to calculate the ground-level PM2.5 concentration in each prefecture-level and above city in China from 1998 to 2014(VAN et al., 2019; VAN et al., 2015).

2.2.1.3 Other social and economic variables

Other control variables were obtained from the China Urban Statistical Yearbook, including the factors for public health, such as health care, economic development, and population density. In this study, we used three variables: the number of medical beds per 10 000 individuals, GDP per capita, and population density in each prefecture-level and above city. The definition and data source of the variables used in this study are shown in Table 1.

Table 1 definition and data source of the variables

Policy variables	Definition of the variables	Source
claim_not	"1" for complete data and "0" for all censored data	Health insurance
buy_age	policy holder's age	
survival time	the difference between the day the begin of observed and the day the stop of observed	
amount	policy holder's	
non_smoke	Smoker, 0 if the insurer is smoker, 1 if not	
gender	Policy holder's Gender of the insurer, 1 for man, 0 for woman	
Pollution variables		
L(PM2.5)	PM2.5 concentration for the city where the policy holder is located among the previous 1 year before the end day	Atomospheric composition analysis group
PM2.5_meanX	The mean PM2.5 concentration during the X years before the year of study observation	
Control variables		
dens_meanX	The mean population density among the previous X years before the end day	URBAN STATISTICAL YEARBOOK OF CHINA
perGDP_meanX	The mean GDP per capita among the previous X years before the end day	
Medical_meanX	The mean number of medical beds per 10 000 individuals among the previous X years before the end day	
indus_water	industrial wastewater discharge	
Indus_dust	industrial fume emission	

Based on the city where the policy was located and the year when the study period ended, the policy data were matched with the PM2.5 concentration and public health data to obtain 21 980 162 data points for the empirical analysis. The data were obtained from 245 prefecture-level and above cities in China.

2.2.2 Descriptive statistics

Descriptive statistics of the variables are shown in Table 2. Because susceptible populations had stronger tractions to PM2.5 pollution, we excluded individuals under 10 or over 65 years of age. Moreover, the results from non-susceptible populations are more readily applicable to the general population. Thus, the data set included individuals aged 10 to 65. We focused on the effects of long-term exposure to high PM2.5 concentrations on human health, and thus, we defined the exposure level to air pollution as the mean PM2.5 concentration over a period of time before the year of study observation (long-term mean concentration). For example, PM2.5_mean5 represented the mean PM2.5 concentration during the five years before the year of study observation (excluding the year of study observation). PM2.5 concentration was expressed in 10 μg/m³. Given the distribution and value of the variables, the logarithmic values were used, including the number of medical beds per 10 000 individuals, mean population density, GDP per capita (x 10 000 yuan), industrial wastewater discharge, industrial fume emission, and policy coverage (x 10 000 yuan).

Table2 Descriptive statistics

Variable	Max	Mean	Min	Medium	Stander error
survival_time	65	41.75462	10	43	11.37847
gender	1	0.483874	0	0	0.49974
amount	1.79176	0.205627	-1.60944	0	0.515776
non_smoke	1	0.991865	0	1	0.089825
L(PM2.5)	113.4	55.34183	4.43292	55.4813	24.21945
PM2.5_mean5	96.92	54.48841	6.340338	54.05	23.4965
PM2.5_mean10	87.54289	50.54567	9.670706	50.47778	22.10809
PM2.5_mean15	84.85079	48.61386	8.335933	48.63261	21.23516
perGDP_mean10	2.215615	0.588459	-0.81362	0.532633	0.699312
Medical_mean10	4.395045	3.387091	2.480495	3.361198	0.416334

dens_mean10	7.653464	5.919877	3.408173	6.129808	0.814632
indus_water	2.063513	-0.26936	-2.90151	-0.33699	1.023681
Indus_dust	2.797706	0.979138	-1.32388	0.985191	0.856303

2.2.3 Correlation coefficient matrix

Table 3 shows the correlation coefficient matrix of some primary explanatory variables. The correlation coefficient was 0.73 between population density and PM2.5 concentration and 0.75 between GDP per capita and the number of medical beds per 10 000 individuals. These indicated a certain level of correlation between past air pollution and economic growth.

Table 3 correlation coefficient matrix of explanatory variables

	PM2.5_10	perGDP_10	Medical_10	dens_10	indus_water	Indus_dust
PM2.5_mean10	1	0.1	-0.18	0.73	0.49	0.27
perGDP_mean10	0.1	1	0.75	0.34	0.44	0.15
Medical_mean10	-0.18	0.75	1	0	0.19	0.24
dens_mean10	0.73	0.34	0	1	0.55	0.17
indus_water	0.49	0.44	0.19	0.55	1	0.49
Indus_dust	0.27	0.15	0.24	0.17	0.49	1

Air pollution may be natural pollution or man-made pollution. This study focused on man-made pollution and thus excluded the effects of sea salt and dust < 2.5 μm. Table 4 shows the PM2.5 concentration resulting from man-made pollution and all other sources and the correlation coefficient matrix of PM2.5 concentrations resulting from man-made pollution over different time spans. The results indicate high correlations between the concentrations.

Table 4 correlation coefficient matrix of PM2.5 concentrations

	L(PM2.5)	PM2.5_5	PM2.5_10	PM2.5_15	PM2.5_10all
L(PM2.5)	1	0.97	0.95	0.97	0.93
PM2.5_mean5	0.97	1	0.99	0.99	0.97
PM2.5_mean10	0.95	0.99	1	1	0.98
PM2.5_mean15	0.97	0.99	1	1	0.98
PM2.5_mean10all	0.93	0.97	0.98	0.98	1

2.3 Empirical analysis

2.3.1 Regression model

We used a proportional-hazard model to estimate the health effects of PM2.5 concentration, assuming that the hazard function or immediate death probability $\lambda(\text{survival time})$ was the product of baseline hazard function $\lambda_0(survival\ time)$ and proportional risk function ϕ, where ϕ could be expressed as the logarithmic linear function of the relevant variables such as the mean PM2.5 concentration:

$$\frac{\lambda(\text{survival time})}{\lambda_0(\text{survival time})} = \phi(PM2.5, personal, control, city)$$

$$= \exp(\beta_0 \text{PM2.5} + \beta_1 \text{personal} + \beta_2 \text{control} + \beta_3 \text{city})$$

wherein *personal* is the personal information in the policy data, such as gender, smoking, occupation, and the amount of coverage purchased; *control* refers to several factors of public health, such as the GDP per capita and the number of medical beds per 10 000 individuals; *PM2.5* refers to the mean PM2.5 concentration over several years prior to the end of the study period in the city; and *city* represents the fixed effect of a prefecture-level city. Assuming that the morbidity varied in each city, *city* was used as a stratification variable in subsequent models.

Survival time (survival time) was defined as the time from the date of birth to the end of the study period, which may be the date of payment for claims or the end of the follow-up period.

The primary explanatory variable for the regression model was the PM2.5 concentration from satellite data, which was used to calculate the long-term mean PM2.5 concentration. Most of the current studies on the relationship between air pollution and morbidity and mortality use the monitoring data of air pollution to estimate the exposure level. However, humans spend more time indoors than outdoors. In developed countries, humans spend approximately 80% of their time indoors; as a result, the actual exposure level will differ from the monitoring data. The concentration of pollutants differs between indoors and outdoors; thus, it may not be suitable to use the outdoor concentration of pollutants to estimate the exposure level. Nonetheless, the indoor and outdoor concentrations of PM2.5 are related (Janssen et al.,2000). Studies have shown that outdoor concentrations of PM2.5, nitrogen oxides, sulfur oxides, and ozone can be used as proxies for individual exposure to PM2.5 (Sarnat et al.,2001). The change in the long-term mean outdoor concentration reflects the change in the individual long-term exposure level (Rijnders et al.,2001; Oglesby et al.,2000). Therefore, we used the long-term mean PM2.5 concentration derived from satellite data as the primary explanatory variable to investigate the correlation between long-term exposure and health.

2.3.2 Empirical results of lung cancer

We used the COX regression model described above and the city-related factors as control variables and obtained the regression results (Table 5). Model (1) used city as a stratification variable. The primary explanatory variable was the 10-year mean PM2.5 concentration, and all other control variables were personal information in the policy data. The coefficient of PM2.5 concentration was significantly positive. The regression results of model (1) showed that for every increase of one unit (10 μg/m³) in PM2.5 concentration, the risk of lung cancer in the population increased by 87% (= exp (0.628) -1). The coefficient of the sex variable was significantly positive, indicating that the incidence of lung cancer was higher in men than in women. In China, the incidences of all common malignant tumors (both men and women) are higher in men than in women. Moreover, cancer-related mortality is also higher in men than in women (Chen et al.,2017). Researchers in the US performed genetic analyses and showed that this is a result of the "Escape from X-Inactivation Tumor Suppressors" (also known as the "EXITS" gene) on the X chromosome in women (Dunford et al.,2015).

The coefficient of the amount of coverage was also significantly positive, indicating that high coverage was associated with higher morbidity, which may be related to information asymmetry in the commercial health insurance market in China. The coefficient of non-smokers was significantly negative, indicating that smoking was a risk factor for lung cancer. Smoking increased the risk of lung cancer by 77%.

Based on model (1), model (2) included the GDP per capita of each city as a control variable. GDP per capita reflects the level of economic development of a city. A high GDP per capita is associated with a high level of economic development. Model (3) included the number of medical beds per 10 000 individuals as a control variable, which reflects the level of health care in the city. In general, the level of health care is proportional to the level of health, and the coefficient of the number of medical beds per 10 000 individuals was significantly negative, indicating that the level of health care was not a risk factor for lung cancer. Models (4), (5), (6), and (7) included other variables that may affect morbidity, such as population density in the city, industrial wastewater discharge, industrial fume emission, and age at purchase. The coefficient of PM2.5 concentration remained significantly positive, indicating that the regression results were robust.

Table 5 PM2.5 concentration and lung cancer

Parameter	(1)	(2)	(3)	(4)	(5)	(6)
mean_pm_10	0.628***	0.572***	0.621***	0.623***	0.625***	0.558***
	(0.023)	(0.04)	(0.041)	(0.041)	(0.041)	(0.042)
gender	0.632***	0.632***	0.632***	0.632***	0.632***	0.632***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
amount	0.601***	0.6***	0.6***	0.6***	0.6***	0.6***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
non_smoke	-0.57***	-0.569***	-0.569***	-0.568***	-0.568***	-0.569***
	(0.053)	(0.053)	(0.053)	(0.053)	(0.053)	(0.053)
occupation1	0.124***	0.124***	0.123***	0.123***	0.123***	0.123***
	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)
occupation2	0.21***	0.209***	0.209***	0.209***	0.209***	0.209***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
occupation3	0.133***	0.133***	0.132***	0.132***	0.132***	0.133***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
occupation4	0.251***	0.25***	0.25***	0.25***	0.25***	0.25***
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
occupation5	0.234***	0.233***	0.233***	0.232***	0.232***	0.232***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
occupation6	0.297***	0.296***	0.296***	0.296***	0.296***	0.296***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
occupation7	-0.342	-0.343	-0.345	-0.345	-0.345	-0.346
	(0.317)	(0.317)	(0.317)	(0.317)	(0.317)	(0.317)
perGDP_mean10		0.097*	0.376***	0.359***	0.361***	0.447***
		(0.057)	(0.072)	(0.073)	(0.073)	(0.074)
Medical_mean10			-1.285***	-1.279***	-1.259***	-1.265***
			(0.201)	(0.201)	(0.201)	(0.201)
dens_mean10				0.32	0.336	0.292
				(0.245)	(0.245)	(0.243)
indus_water					-0.131*	-0.002
					(0.079)	(0.081)
indus_DUST						-0.37***
						(0.051)
City stratification	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes

Note: "*", "**", and "***" indicate statistical significance at 10%, 5%, and 1%, respectively.

In Table 5, the primary explanatory variable is the 10-year mean PM2.5 concentration. In Table 6, model (1) uses the mean PM2.5 concentrations of different time spans derived from the 10-year mean PM2.5 concentration as the primary explanatory variables. For simplicity, Table 6 omits the regression results of the occupation variable. Models (2), (3), and (4) represent the 15-year mean PM2.5 concentration, the 5-year mean PM2.5 concentration, and the 1-year mean PM2.5 concentration, respectively. The coefficients were all significantly positive. The coefficient of the 15-year mean PM2.5 concentration was the highest, indicating the cumulative effect of air pollution on the morbidity of lung cancer, the health end-point. For every increase of 10 $\mu g/m^3$ in the 1-year, 5-year, 10-year, or 15-year mean PM2.5 concentration, the risk of lung cancer increased by 1.78, 4.81, 2.17, or 1.12 times, respectively, which was consistent with the trend of the findings of literature reports, showing that with every increase of 10 $\mu g/m^3$ in PM2.5 concentration, the risk of lung cancer-related deaths was increased by 19% to 30%. The specific values were higher in this study, which occurred because this study used claims data that reflected morbidity, whereas literature reports focused on mortality.

Moreover, we used the mean PM2.5 concentrations of different time spans (1 to 15 years) to substitute the 10-year mean PM2.5 concentration for regression analysis. The results showed that all PM2.5 concentrations were significantly positively correlated with morbidity. To simplify the calculations, the 10-year mean PM2.5 concentration was used as the primary explanatory variable in subsequent regression analyses.

In this study, the PM2.5 concentration data were derived from the raster data of Martin et al. The group provided two PM2.5 concentrations: one was the overall concentration, and the other excluded the concentration of natural particles. The regression analyses described above used the PM2.5 concentration without natural particles. In model (5), the 10-year mean overall PM2.5 concentration was used as the primary explanatory variable, and the results were still significantly positive. Model (5) showed that with every change of one unit (10 $\mu g/m^3$) in overall PM2.5 concentration, the lnRR of relative risk changed by 1.63 units. For both models (1) and (5), the results were significantly positive, and the coefficient of model (1) was higher than that of model (5). Therefore, the PM2.5 concentration of man-made pollutants only (without natural particles) was associated with a higher relative risk, with a stronger correlation with the morbidity of lung cancer. In subsequent analyses, the PM2.5 concentration without natural particles was used as the primary explanatory variable.

Table 6 Different time spans of PM2.5 concentration and lung cancer

Parameter	(1)	(2)	(3)	(4)	(5)	(6)
mean_pm_10	0.578***					0.467***
	(0.042)					(0.043)
mean_pm_15		1.57***				
		(0.082)				
mean_pm_1			0.117***			
			(0.016)			
mean_pm_5				0.773***		
				(0.032)		
mean_pm_10all					0.488***	
					(0.038)	
gender	0.628***	0.628***	0.628***	0.628***	0.628***	0.628***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
amount	0.596***	0.596***	0.596***	0.596***	0.596***	0.594***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
non_smoke	-0.555***	-0.554***	-0.555***	-0.553***	-0.555***	-0.553***
	(0.053)	(0.053)	(0.053)	(0.053)	(0.053)	(0.053)
perGDP_mean10	0.393***	0.253***	1.02***	0.527***	0.451***	
	(0.072)	(0.071)	(0.062)	(0.065)	(0.072)	
Medical_mean10	-1.178***	-0.908***	-0.386**	0.179	-1.107***	
	(0.199)	(0.198)	(0.195)	(0.202)	(0.198)	
dens_mean10	0.261	0.306	0.145	0.581**	0.263	
	(0.242)	(0.242)	(0.241)	(0.248)	(0.242)	
perGDP_mean5						0.521***
						(0.051)
Medical_mean5						-1.064***
						(0.115)
dens_mean5						0.071
						(0.142)
indus_water	-0.015	-0.122	0.054	-0.197**	-0.025	-0.01
	(0.081)	(0.082)	(0.08)	(0.084)	(0.081)	(0.081)
indus_DUST	-0.361***	-0.309***	-0.501***	-0.303***	-0.377***	-0.389***
	(0.051)	(0.051)	(0.05)	(0.051)	(0.051)	(0.051)
City stratification	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes

The GDP per capita and the number of medical beds per 10 000 individuals in each city varied year to year. In models (1) to (5), the 10-year means of GDP per capita and the number of medical beds per 10 000 individuals were used as the control variables. In model (6), the 5-year means were used as the control variables, with consistent regression results. The conclusions remained robust with various control variables. The time span of control variables had no effect on the significance of the regression results.

The adverse effect of air pollution on the respiratory tract was visualized. Studies have shown that PM2.5 has complex effects on human health and increases the morbidity of many diseases such as cardiovascular diseases and stroke. We then selected the morbidity of malignant tumors at different sites in the human body as the end-point event, and the regression results are shown in Table 7. PM2.5 concentration was the risk factor for the diseases on the right. The regression results of the diseases on the left were insignificant and were sorted in ascending order of hazard ratio. The results showed that in addition to lung cancer, PM2.5 was a risk factor for some other malignant tumors.

Table 7 The hazard ratio of PM2.5 for each malignant tumors

Malignant Tumor	Estimate	StdErr	HazardRatio	Malignant Tumor	Estimate	StdErr	HazardRatio
Thyroid	0.101**	0.047	1.107	Skin	0.253	0.223	1.287
Lymphoma	0.177*	0.093	1.194	Adenocarcinoma	0.225	0.164	1.252
Leukemia	0.202**	0.091	1.223	Bladder	0.144	0.098	1.155
Cardiovascular	0.349***	0.06	1.418	Kidney	0.107	0.075	1.112
Biran	0.437***	0.104	1.548	Esophagus	0.079	0.061	1.082
Trachea	0.545**	0.264	1.725	Ovary	0.066	0.089	1.069
Lung	0.558***	0.042	1.748	Rectum	0.012	0.061	1.012
Liver	0.963***	0.046	2.62	Mammary	-0.007	0.029	0.993

The regression analyses per disease type showed that PM2.5 was harmful to the respiratory system and to other sites. Because of its complex components, which include heavy metals and toxic and hazardous substances, PM2.5 is harmful to the respiratory system and beyond. Next, we performed subgroup analyses per organ system, including the respiratory, endocrine, digestive, skeletal and muscular, urinary, reproductive, blood circulatory, nervous, and immune systems. The overall morbidity and mortality of malignant tumors of each organ system were used as health end-points to investigate the hazard ratio of PM2.5 for each organ system (Table 8).

Table 8 The hazard ratio of PM2.5 for each organ system

calim	Estimate	StdErr	HazardRatio
Respiratory system	0.501***	0.037	1.65
Immune system	0.497***	0.175	1.644
Nervous system	0.391***	0.026	1.478
Digestive system	0.34***	0.022	1.405
Exercise system	0.306***	0.078	1.358
Urinary system	0.159***	0.039	1.172
Endocrine system	0.142***	0.046	1.152
Blood circulation system	0.118***	0.023	1.125
Reproductive system	0.031	0.023	1.031

Table 9 Critical and severe diseases, malignant tumors, and lung cancer as the health end-points

Parameter	Critical and Severe Diseases	HazardRatio	Malignant Tumor	HazardRatio	Lung Cancer	HazardRatio
mean_pm_10	0.11***	1.116	0.176***	1.193	0.558***	1.748
	(0.01)		(0.013)		(0.042)	
female	0.198***	1.218	-0.101***	0.904	0.632***	1.881
	(0.004)		(0.004)		(0.013)	
amount	0.695***	2.004	0.647***	1.909	0.6***	1.822
	(0.004)		(0.005)		(0.015)	
non_smoke	-0.184***	0.832	-0.192***	0.825	-0.569***	0.566
	(0.019)		(0.023)		(0.053)	
perGDP_mean10	-0.075***	0.928	0.166***	1.181	0.447***	1.564
	(0.019)		(0.023)		(0.074)	
Medical_mean10	-0.561***	0.571	-0.785***	0.456	-1.265***	0.282
	(0.051)		(0.062)		(0.201)	
dens_mean10	0.083	1.087	0.181**	1.198	0.292	1.34
	(0.063)		(0.071)		(0.243)	
indus_water	0.062***	1.064	0.108***	1.114	-0.002	0.998
	(0.021)		(0.025)		(0.081)	
indus_DUST	-0.144***	0.866	-0.172***	0.842	-0.37***	0.69
	(0.013)		(0.016)		(0.051)	
City stratification	Yes		Yes		Yes	
Time fixed effect	Yes		Yes		Yes	

PM2.5 in the air enters the body primarily via the respiratory tract, which enters the bloodstream through the lungs and then each organ system via the bloodstream. As a result, PM2.5 may harm all organ systems. The regression results described above showed that PM2.5 had complex harmful effects on the human body with varying degrees of damage to each organ system, with particularly pronounced adverse effects on the respiratory system.

Next, we investigated the effects of PM2.5 on human health from a broader perspective and performed COX regression analyses using critical and severe diseases, malignant tumors, and lung cancer as the health end-points. The regression results are shown in Table 9.

The results showed that PM2.5 was significantly positively correlated with the risk of critical and severe diseases and malignant tumors. Specifically, with every increase of one unit (10 $\mu g/m^3$) in PM2.5 concentration, the risk of critical and severe diseases, malignant tumors, and lung cancer increased by 12%, 19%, and 75%. Thus, PM2.5 concentration was a risk factor for critical and severe diseases and malignant tumors, with a particularly pronounced effect on lung cancer.

To ensure the reliability of the results, we used different models and data subsets for the regression analysis to test the robustness of the results.

2.3.3 Poisson regression

In epidemiological studies, the COX model is frequently used in cohort studies to identify the risk factors for certain diseases. In addition, Poisson regression is commonly used to analyze cohort data, and some studies have compared these two methods. We used the Poisson model to analyze the cohort data described above. The policy holders were divided into 12 age groups to generate dummy variables, which were incorporated into the Poisson model. During data processing, the subjects were stratified per age into 5-year age groups (from 10 to 65). Claim was a binary variable and was coded as “1” if a health end-point event occurred or “0” if no health end-point event occurred. In addition, the city was incorporated into the Poisson model as a fixed effect, as follows:

$$claim = PM2.5 + personal + control + city$$

As described above, personal includes personal information in the policy data, such as smoking, the amount of coverage, gender, and occupation. Moreover, the age of the policy holder during the year of study was included in the Poisson model as a control variable. Control variables included certain city-level variables, such as GDP per capita, the number of medical beds per 10 000 individuals, industrial fume and sulfur dioxide emissions, and the city (as a fixed effect).

We used critical and severe diseases, malignant tumors, and lung cancer as health end-points for Poisson regression. The results are shown in Table 10. To simplify comparison, Table 10 omits the results with the city as a fixed effect and provides the upper/lower confidence values.

Table 10 Critical and severe diseases, malignant tumors, and lung cancer for Poisson regression

Parameter	Critical and Severe Diseases	Lower	Upper	Malignant Tumor	Lower	Upper	Lung Cancer	Lower	Upper
Intercept	-6.468*** (0.457)	-7.36	-5.57	-7.163*** (0.522)	-8.19	-6.14	-11.456*** (1.778)	-14.94	-7.97
mean_pm_10	0.218*** (0.01)	0.2	0.24	0.277*** (0.013)	0.25	0.3	0.067*** (0.004)	0.06	0.08
age1	-4.408*** (0.054)	-4.51	-4.3	-4.402*** (0.062)	-4.52	-4.28	-24.047 (2462.26)	-4849.99	4801.9
age2	-4.215*** (0.048)	-4.31	-4.12	-4.152*** (0.055)	-4.26	-4.04	-6.134*** (0.321)	-6.76	-5.5
age3	-3.771*** (0.037)	-3.84	-3.7	-3.82*** (0.044)	-3.91	-3.73	-6.382*** (0.321)	-7.01	-5.75
age4	-3.286*** (0.029)	-3.34	-3.23	-3.28*** (0.034)	-3.35	-3.21	-5.444*** (0.182)	-5.8	-5.09
age5	-2.719*** (0.023)	-2.76	-2.67	-2.644*** (0.026)	-2.7	-2.59	-4.449*** (0.097)	-4.64	-4.26
age6	-2.203*** (0.021)	-2.24	-2.16	-2.101*** (0.024)	-2.15	-2.05	-3.459*** (0.064)	-3.58	-3.33
age7	-1.696*** (0.02)	-1.73	-1.66	-1.621*** (0.023)	-1.67	-1.58	-2.753*** (0.056)	-2.86	-2.64
age8	-1.187*** (0.02)	-1.23	-1.15	-1.157*** (0.023)	-1.2	-1.11	-1.998*** (0.054)	-2.1	-1.89
age9	-0.765*** (0.02)	-0.8	-0.73	-0.785*** (0.023)	-0.83	-0.74	-1.299*** (0.053)	-1.4	-1.19
age10	-0.421*** (0.02)	-0.46	-0.38	-0.457*** (0.023)	-0.5	-0.41	-0.722*** (0.052)	-0.82	-0.62
female	0.143*** (0.004)	0.14	0.15	-0.148*** (0.004)	-0.16	-0.14	0.584*** (0.013)	0.56	0.61
amount	0.197*** (0.004)	0.19	0.21	0.135*** (0.005)	0.13	0.14	0.134*** (0.015)	0.1	0.16

non_smoke	-0.054***	-0.09	-0.02	-0.078***	-0.12	-0.03	-0.402***	-0.51	-0.3
	(0.019)			(0.023)			(0.053)		
perGDP_mean10	0.549***	0.51	0.59	0.81***	0.77	0.86	1.064***	0.92	1.21
	(0.019)			(0.023)			(0.074)		
Medical_mean10	-0.43***	-0.53	-0.33	-0.664***	-0.78	-0.54	-1.198***	-1.59	-0.81
	(0.051)			(0.061)			(0.198)		
dens_mean10	0.214***	0.09	0.34	0.272***	0.14	0.41	0.404*	-0.07	0.87
	(0.063)			(0.07)			(0.24)		
indus_water	0.02	-0.02	0.06	0.077***	0.03	0.13	-0.04	-0.2	0.12
	(0.021)			(0.025)			(0.08)		
indus_DUST	-0.15***	-0.18	-0.12	-0.176***	-0.21	-0.15	-0.364***	-0.46	-0.26
	(0.013)			(0.016)			(0.051)		
City stratification	Yes			Yes			Yes		
Time fixed effect	Yes			Yes			Yes		

With every change of one unit (10 μg/m³) in PM2.5 concentration, the coefficient of Cox regression changed 0.11-fold for critical and severe diseases and 0.18-fold for malignant tumors, and the coefficients of the Poisson regression were 0.22 and 0.28, respectively. With critical and severe diseases and malignant tumors as health end-points, the coefficient of the Poisson regression was higher than that of the Cox regression. Nonetheless, both methods showed that PM2.5 concentration was a risk factor for critical and severe diseases and malignant tumors. Thus, Poisson regression with the same health end-points as described above produced similar results, indicating that the results of this study were robust and reached the same conclusions with different models.

2.3.4 Robustness test

There are over 300 prefecture-level or above cities in China, of which less than 100 have air quality monitoring data. According to data from the Statistical Yearbook, the monitoring data before 2013 included the concentration of air pollutants, such as PM10, SO_2, and NO_2. The cities were divided into two groups based on whether the concentration of pollutants was monitored. The concentration data used in this study were derived from satellite data and thus were not affected by the distribution of surface data monitoring stations. As a result, we were able to obtain data of over 200 cities. If the city of the policy holder had pollutant monitoring stations, the policy data were included in dataset 1; otherwise, the policy data were included in dataset 2. The two data subsets were used for COX regression, and the results are shown in Table 11, where (1) - (3) are the regression results of the cities with monitoring data, and (4) - (6) are the regression results of dataset 2.

Table 11 Two data subsets were used for COX regression

	dataset 1			dataset 2		
Parameter	Critical and Severe Diseases	Malignant Tumor	Lung Cancer	Critical and Severe Diseases	Malignant Tumor	Lung Cancer
mean_pm_10	0.027	0.087***	0.53***	0.188***	0.258***	0.704***
	(0.021)	(0.026)	(0.086)	(0.012)	(0.015)	(0.051)
female	0.134***	-0.219***	0.602***	0.209***	-0.076***	0.632***
	(0.008)	(0.009)	(0.028)	(0.004)	(0.005)	(0.015)
amount	0.648***	0.638***	0.544***	0.747***	0.691***	0.646***
	(0.008)	(0.009)	(0.029)	(0.005)	(0.006)	(0.017)
smoke	-0.204***	-0.153***	-0.478***	-0.187***	-0.221***	-0.612***
	(0.034)	(0.044)	(0.104)	(0.022)	(0.027)	(0.062)
perGDP_mean10	0.03	0.256***	0.347	-0.063***	0.204***	0.511***
	(0.053)	(0.066)	(0.217)	(0.021)	(0.026)	(0.083)
Medical_mean10	-0.363**	-0.493***	-1.983***	-0.603***	-0.819***	-1.378***
	(0.144)	(0.174)	(0.573)	(0.057)	(0.069)	(0.224)
dens_mean10	-0.895***	-1.066***	-0.682	-0.173***	-0.096	-0.025
	(0.3)	(0.371)	(1.251)	(0.065)	(0.073)	(0.251)
indus_water	0.296***	0.271***	-0.452	0.13***	0.242***	0.141
	(0.069)	(0.083)	(0.28)	(0.023)	(0.028)	(0.092)
indus_DUST	-0.331***	-0.337***	-0.864***	-0.073***	-0.11***	-0.273***
	(0.042)	(0.05)	(0.163)	(0.014)	(0.017)	(0.055)
City stratification	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes

The regression results in Table 11 show that urban residents without air quality monitoring were at a higher risk of the harmful effects of air pollution than those with air quality monitoring. This may be because the residents in cities with air quality monitoring were more concerned about air pollution and were more likely to take appropriate self-protective measures such as using an air purifier and wearing a mask to reduce the risk of diseases. Conversely, the cities with air quality monitoring stations were mostly developed cities, with better public health and health care, which helped to reduce morbidity and improve the detection rate of diseases. We were unable to discern these effects in this study, and further research is necessary to investigate the causes. Nonetheless, the regression results of the two data subsets were consistent with those described above, demonstrating the robustness of the results of this study.

In order to solve the endogenous problem in this paper, we use exogenous meteorological data as instrument variables. According to the Box model (Jacobson M Z, Oppenheimer M, 2003), the diffusion of air pollutants is determined by two meteorological factors. The first is the wind speed, the greater the wind speed, the faster the diffusion of pollution in the horizontal direction; the second is the mixing height (also known as boundary layer height), which reflects the vertical diffusion ability of pollutants. The product of wind speed and mixing height is the ventilation coefficient used in this paper. The instrument variable is derived from exogenous meteorological data, which reflects the diffusion rate of pollutants in the atmosphere (horizontal and vertical directions).

Based on the Box model, Broner et al. used ERA-Interim meteorological data to construct instrument variables (Broner et al., 2012). Using the wind speed and boundary layer height at 10 meters above 75°×75°, combined with the map of administrative divisions in China, the average value of each city is obtained by calculating. The calculation method of the control variable refers to the existing literature (Cai X et al., 2016; Hering L and Poncet S, 2014).

Table 12 Instrument variable Regression Results

Parameter	重疾	malignant tumors	Lung Cancer
pmiv10	1.113***	0.977***	0.686*
	(0.25)	(0.281)	(0.395)
female	0.185***	-0.122***	0.621***
	(0.014)	(0.021)	(0.024)
amount	0.684***	0.641***	0.622***
	(0.013)	(0.012)	(0.02)
smoke	-0.059**	-0.058*	-0.394***
	(0.025)	(0.031)	(0.064)
log_mean_gdp_10	-0.403	-0.087	0.348
	(0.305)	(0.341)	(0.472)
log_mean_bed_10	-0.333***	-0.364***	-0.361**
	(0.095)	(0.109)	(0.179)
log_mean_dens_10	0.123	0.072	-0.001
	(0.155)	(0.194)	(0.233)

In the first stage, the PM2.5 concentration of each city is used as the explanatory variable, the instrumental variable as the explanatory variable, and the characteristic variable of the city was controlled for regression. The regression results were significant, and the instrument variables were highly correlated with the pollution level. The Cox proportional risk regression of the second stage is carried out with the first stage estimates as the main explanatory variables. The regression results are shown in Table 12.

Through the robustness test in this part, we can see that air pollution can indeed lead to an increase in the incidence of malignant tumors and major diseases. Air pollution is an important disease risk factor. Next, we will economically quantify these risks and assess the health and welfare losses caused by PM2.5 pollution.

2.3.5 Evaluation of the welfare loss

We further evaluated the loss of life expectancy and economic loss due to air pollution.

2.3.5.1 Effect on life expectancy

First, we will explain how to calculate excess morbidity from the actual morbidity and the relative risk of pollution. Assuming that the morbidity in clean air is X, which is increased by R with air pollution (i.e., relative risk [RR]-1), the actual morbidity Y should be X (1 + R). Given that X = Y/(1 + R), the excess morbidity AR = 1-X/Y = 1-1/(1 + R) = 1-1/RR.

According to the *Ambient Air Quality Standard* (GB3095-2012) in China, for particles ≤ 2.5 μm, the first-level limit is 15 μg/m³ for the annual mean concentration; the second-level limit is 35 μg/m³. According to the WHO Air Quality Guidelines (AQG), the annual mean guideline is 10 μg/m³ for PM2.5. In this study, we used the AQG for reference c0 to calculate relative risk $RR = HR^{c-c_0} = e^{\beta(c-c_0)}$, in which c is the observed concentration of pollutants, and c0 is the reference concentration.

We then combined population exposure and actual morbidity/mortality to calculate excess morbidity and excess mortality due to PM2.5 pollution. The equation for excess morbidity is as follows:

$$I = POP * m * AR \quad (1)$$

where I is excess morbidity, POP is population exposure, and m is actual morbidity. In this study, actual morbidity was the observed morbidity, and the baseline was the observed concentration of air pollutants. Table 12 shows the morbidity (n, %) during each year of the study period. According to the Statistical Yearbook of Health issued by the National Bureau of Statistics of China, the mortality of malignant tumors is approximately 0.0016, which is consistent with the overall mortality of malignant tumors shown in Table 12.

Table 12 Actual morbidity

year	Total person-years	Critical and Severe Diseases		Malignant Tumor		Lung Cancer	
		Number of illness	Morbidity	Number of illness	Morbidity	Number of illness	Morbidity
2006	15562366	19415	0.0012476	12189	0.0007832	537	3.451E-05
2007	19239179	26381	0.0013712	16879	0.0008773	925	4.808E-05
2008	22604307	35406	0.0015663	22693	0.0010039	1583	7.003E-05
2009	24794608	48287	0.0019475	31776	0.0012816	3175	0.0001281
2010	24746244	60377	0.0024398	44035	0.0017795	5475	0.0002212
2011	24648004	62283	0.0025269	46307	0.0018787	5741	0.0002329
2012	24549179	66382	0.002704	50002	0.0020368	6280	0.0002558
2013	24766675	73513	0.0029682	54382	0.0021958	6897	0.0002785
sum	180910562	392044	0.0021671	278263	0.0015381	30613	0.0001692

Table 12 shows the actual mean morbidity from all the cities. The excess morbidity is the product of population, actual morbidity, and AR. Because the value of AR is related to the actual concentration, the results of individual cities would be more accurate and reliable given the different concentrations of pollutants across the cities. Therefore, we analyzed each city before obtaining the sum. Specifically, for each city, POP is the population of the city, and the actual morbidity is calculated with raw data. Excess morbidity was 0 if the concentration of pollutants was below the baseline concentration; otherwise, excess morbidity was calculated with equation (1).

Finally, the overall excess morbidity is the sum of excess morbidity I_i of each city:

$$I = \sum_i I_i = \sum_i POP_i * m_i * \frac{e^{\beta(c_i - c_0)} - 1}{e^{\beta(c_i - c_0)}} \tag{2}$$

Table 13. Excess morbidity (x 10 000 individuals)

Excess Morbidity	Critical and Severe Diseases		Malignant Tumor		Lung Cancer	
	Zero Pollution	International Standard	Zero Pollution	International Standard	Zero Pollution	International Standard
All cities	227.7311	217.5814	164.9504	160.1279	18.93176	18.7144

Table 13 shows that if the PM2.5 concentration is reduced to 0 μg/m³ (zero pollution) and 10 μg/m³ (international standard), the total number of critical and severe diseases will be reduced by 2.2737 million and 2.1758 million cases, respectively; malignant tumors will be reduced by 1.6495 million and 1.6013 million cases, respectively; and lung cancer will be reduced 189 300 and 187 100 cases, respectively. The excess morbidity of lung cancer is consistent with the findings of literature reports (~180 000).

Air pollution not only increases morbidity, but it also increases the risk of premature death. Therefore, we calculated excess mortality due to pollution. Excess mortality D is calculated using a similar method as morbidity: $D = POP * m * AR$. Death is defined as a health end-point event. Assuming that individuals with critical and severe diseases die of their condition, the excess mortality will be the same as the excess mortality due to critical and severe diseases calculated above. We used the dataset containing susceptible populations and divided the subjects into 5-year age groups to obtain the POP for each age group and calculate the mean mortality of each group. The excess mortality of each age group is the product of the mean mortality and AR, which represents deaths due to PM2.5 pollution.

In Table 14, we referenced the Jiang Qinglang life expectancy table and constructed a simple actual life expectancy table using all sample observations. We then calculated the excess mortality of each age group due to air pollution using the method described above. We also constructed a life expectancy table by eliminating the effect of pollution. The loss of mean life expectancy as a result of PM2.5 pollution is the difference between life expectancy with no effect of pollution and simple life expectancy. The results are shown in Table 15.

Table 14. Simple life expectancy table

Age Group	Life Expectancy	Age Group	Life Expectancy
0	78.24	40	38.98
1	77.26	45	34.16
5	73.31	50	29.38
10	68.36	55	24.66
15	63.42	60	19.98
20	58.51	65	15.34
25	53.6	70	10.9
30	48.7	75	6.63
35	43.83	80	2.5

Table 15. Loss of life expectancy

Sample classification	Actual life expectancy	Life expectancy with no effect of pollution	Loss of life expectancy	Life expectancy with no effect of pollution	Loss of life expectancy
		Reference: zero pollution		Reference: AQG	
All cities	78.24	80.08	1.84	79.82	1.58

Table 15 shows that the loss of life expectancy is greater when using zero pollution than when using AQG as a reference point. Given that zero pollution does not exist in nature, the results using zero pollution as a reference point serves as a reference only for the upper limit. Taken together, the loss of life expectancy is approximately 1.58 years due to PM2.5 pollution.

2.3.5.2 Economic loss

Environmental pollution causes significant economic losses. In this study, we investigated the economic consequences of health loss due to pollution, including two types of health-related economic losses: disease-related healthcare expenses and mortality-related labor cost.

Air pollution increases morbidity. We calculated healthcare expenses associated with high morbidity. We used the data from hospital patients to calculate the recent mean healthcare expenses of malignant tumors, which was approximately 100 000 yuan in 2013 (baseline). The analysis above showed that the case number of excess morbidities due to PM2.5 pollution was approximately 1.601 million for malignant tumors and 187 000 for lung cancer. Thus, we used the case number and the mean healthcare expenses per patient to calculate the overall healthcare expenses due to PM2.5 pollution, which was approximately 160.1 billion yuan, of which approximately 18.7 billion yuan was related to lung cancer. Due to limited data, we provided a rough estimate of healthcare expenses for malignant tumors, but not for other critical and severe diseases.

Labor cost due to the loss of life expectancy was defined as the product of the mean wage, the mean annual number of employees (number of employees), and the mean loss of life expectancy. The equation is shown below:

$$L = \sum_i L_i = \sum_i wage_i * ET_i * wpop_i \tag{3}$$

where $wage_i$ is the mean wage of city i (fixed wage in 2006 [baseline]), ET represents the mean loss of life expectancy, and $wpop_i$ is the total number of employees in city i.

This method estimates only labor costs associated with the loss of life expectancy, where the mean wage is used as an approximate for labor value. The assumption is that the labor value is equal to the contribution to society (wage). As a result, this method fails to estimate the costs of premature death in individuals outside the workforce (minors and the elderly). Moreover, there are ethics concerns when measuring life value with wage in the working population. Therefore, some researchers have proposed an alternative method, in which the GDP per capita is used to measure the overall loss due to premature death, rather than the wage in the working population. The equation is shown below:

$$L = \sum_i \frac{GDP_i}{P_i} * ET_i * POP_i = \sum_i perGDP_i * ET_i * POP_i = perGDP * ET * POP \quad (4)$$

where ET is the mean loss of life expectancy, GDP per capita equals the total GDP divided by the total population (POP), and POP is the population exposure. The economic loss associated with premature death estimated with these two methods is shown in Table 16.

Table 16. Economic loss associated with premature death (x 0.1 billion yuan)

Pollution	Mean wage		GDP per capita	
	Zero pollution	International standard	Zero pollution	International standard
All cities	5.67	6.6	53.71	62.55
Cities with monitoring data	1.69	2.01	15.98	19.04
Cities with no monitoring data	8.44	9.55	79.89	90.43

In Table 16, the column on the left uses the mean wage to estimate labor costs associated with premature death. As a result, only the working population is included in the estimate, which is approximately 660 million yuan. The column on the right uses a modified method that includes the overall population, not just the working population, and the estimate is approximately 6.26 billion yuan.

The above analyses calculated the excess morbidity, excess mortality, and associated economic losses due to the PM2.5 pollution. The overall health-related cost due to air pollution is the sum of the health-related cost of all health end-points:

$$TV = \sum_j V_j = \sum_j v_j * E_j \quad (5)$$

where TV is the overall health-related loss due to air pollution, Vj is the health-related loss due to factor j, vj is the unit health-related loss (unit economic value), and Ej is the change in health risk due to the change in the concentration of air pollutants. This study measured two health end-points: death and malignant tumors. Health-related cost due to morbidity refers to healthcare expenses for excess morbidity, whereas health-related cost due to mortality refers to labor costs associated with premature death.

In 2006, the health-related cost due to death and malignant tumors was 90.08 billion yuan, accounting for approximately 0.4% of the GDP in 2006. In 2013, the overall economic loss was 172.05 billion yuan, accounting for approximately 0.3% of the GDP in 2013.

2.4 Conclusion

In this study, we used the data from tens of millions of health insurance policies across China to investigate the effects of long-term exposure to high PM2.5 concentrations on the morbidity of lung cancer and related diseases. The results showed that after controlling for the level of economic development, public health, and personal information, a high PM2.5 concentration was correlated with a high risk of lung cancer. The survival analysis showed that with every increase of 10 $\mu g/m^3$ in the 10-year mean PM2.5 concentration, the risk of lung cancer increased by 87%. Long-term exposure to high PM2.5 concentrations affected not only the respiratory system but also many other organs.

Furthermore, we investigated the loss of life expectancy and economic loss due to PM2.5 pollution. The results showed that the mean life expectancy was reduced by 1.6 years due to PM2.5 pollution, and direct healthcare expenses for pollution-induced diseases and direct economic loss associated with labor cost amounted to several hundred billion yuan. This study included the data from over 20 million policies on critical and severe diseases, where policy holders tended to have above-average income, which may mitigate some effects of pollution thanks to better health care. Nonetheless, indirect loss due to PM2.5 pollution is enormous, including the loss of work hours due to the need to care for family members and the loss of life expectancy and economic loss.

This study serves as a valuable reference because it included a large amount of detailed policy data across China, which sets it apart from traditional small epidemiological surveys and previous macroeconomic research. This study investigated the long-term health effects of PM2.5 and the effects of PM2.5 pollution on the morbidity of different diseases, thereby filling the gap in previous literature. Moreover, the results serve as a reference for debate of the loss of life expectancy and economic loss due to pollution and for the development of relevant environmental and economic policies.

Chapter 3: The effect of air pollution on suicide rate

Abstract: Air pollution not only has an impact on human health, but also has a negative impact on people's mood and psychology, leading to anxiety and depression, and even increases the risk of suicide. In this paper, the correlation between air quality and suicide incidents is studied by using the data of health insurance policies in China. Empirical results show that there is a significant positive correlation between air quality index AQI and suicide events, and men are more sensitive to air pollution.

3.1 Introduction

With the improvement of people's living standard, people begin to pay more attention to environmental pollution. Especially since 2013, after several times of heavy haze, the search volume of key words such as air pollution, PM2.5 and haze has increased dramatically, and the air pollution problem in China has become a concern of the whole people. The news media also reported many serious haze incidents, such as almost all parts of the Middle East after the winter of 2013. The air quality index of Tianjin, Hebei, Shandong, Jiangsu, Anhui, Henan, Zhejiang, Shanghai and other places has reached the sixth level of serious pollution.

Almost all people living in haze-covered areas are aware of its threat to health. However, relative to the physiological impact, the negative impact of haze on people's mood and psychology has been relatively neglected. The domestic research in this area is almost blank, and the related early warning measures rarely mention the psychological intervention under haze weather. In fact, some foreign studies have found that pollutants such as ozone, nitrogen dioxide, PM2.5 and PM10 also have effects on anxiety, depression and suicide risk. Canadian scholars have found that if the weather is sunny, it will have a positive impact on people's mood, more optimistic. If there is no sunshine in cloudy days, they will be in a bad mood and aggravate negative emotions such as depression (Howard and Hoffman, 1984). Modern medicine has found that the human brain has a gland called the pineal gland, which sensitively senses light. When sunlight is abundant, the activity of cells is low. While the light in the surrounding environment becomes dim, without sunlight, cells become active and inhibit the production of some hormones in the human body.

These include thyroxine and adrenaline, both of which are exciting hormones (Ye Baikuan and Guo Xiazhen, 2000). When the secretion of these two hormones decreases, people become depressed. If long-term haze, the state of depression is more obvious.

The report "Towards an Environment Sustainable Future: National Environmental Analysis of the People's Republic of China", issued jointly by the Asian Development Bank and Tsinghua University in 2013, points out that seven of the 10 most polluted cities in the world are in China. The World Pollution Map released by the World Health Organization (WHO) shows that China is a disaster-stricken area. In addition to the impact of air pollution on physical health, the impact on mental health should not be underestimated.

Air pollutants are neurotoxic and can damage the central and peripheral nervous systems, leading to cardiovascular diseases and brain and cognitive functions (Brook et al., 2011). Air-polluted particulate matter is an inflammatory agent that causes inflammation in the prefrontal cortex and further development of Vascular Depression, thereby increasing the risk of death (Wagner et al., 2014). Exposure to heavily polluted air, in addition to impairing cognitive and nervous systems, can also lead to increased anxiety and depression (Arvin and Lew, 2012; Marques and Lima, 2011).

The Department of Neuroscience of Ohio State University and Davis Institute of Cardiopulmonary Research conducted a mouse experiment. The results showed that the dendritic spines in the hippocampus of mice exposed to polluted air were less, dendrites were shorter and cell complexity decreased. The hippocampus of the brain controls learning and memory ability and depression. Changes in the hippocampus of mice can lead to reduced learning and memory ability. Proinflammatory cytokines are more active, causing inflammation, leading to health problems, including depression (Fonken et al., 2011).

For people with seasonal emotional disorder (SAD) and depression, bad weather is enough to make them emotionally uncontrolled, restless and even increase their risk of suicide. For SAD patients, everything is wonderful in summer, while the darkness in winter is painful and makes them suffer from depression. In addition to weather changes, air pollution can also aggravate the suffering of many patients with mental illness. Three researchers, Gary W. Evans of the University of California, Irvine and others, found that when there were more photochemical oxidants in the environment, more people showed anxiety symptoms (Lim et al., 2012). There are more photochemical oxidants in the air and more anxious people (Gary et al., 1983; Cavanagh et al., 2003).

Air pollution threatens people's psychology in all aspects. Ordinary people may feel anxious or depressed because of it. It only affects their mood, but the consequences are not serious. For people with more vulnerable psychology, haze can aggravate their depression and even lead to suicide. Some scholars have analyzed the number of telephone calls and the degree of air pollution for psychiatric first aid in the United States, and found that when air pollution is serious, the number of calls for mental illness will increase (Rotton and Frey, 1984; Rotton and Frey, 1985).

There is a positive correlation between air pollution and suicide rate in cities with different meteorological, geographical and cultural characteristics. Suicide only lagged behind the peak of fine particles and nitrogen dioxide for 2-3 days (Bakian et al., 2015). The cross-case method showed that PM2.5 and PM10 increased suicide risk. PM10 levels 0-2 days ago will increase suicide risk by 9%, PM2.5 levels 1 day ago will increase suicide risk by 10.1%. Increased fine particulate matter will increase suicide risk, especially for individuals with cardiovascular disease. Men and people aged 36 to 64 are at the highest risk of suicide after exposure to short-term air pollution (Kim et al., 2010). High concentrations of particulate matter increase the risk of suicide (Fleehart et al., 2014, Jee et al., 2011). The increase of optical oxidants will increase the suicide risk of patients with mental illness (Cavanagh et al., 2003).

Researchers have analyzed the quantitative relationship between more than 4000 suicide cases and PM10.Suicide increases when pollution peaks. When pollution peaks above the median, the suicide rate increases by 9%. For people with cardiovascular disease, the risk of suicide increased by 19% (Ha et al., 2015). The influence of air pollution on suicide varies from season to season. Studies have shown that cold seasons are more likely to trigger the effects of air pollution on suicide risk (Szyszkowicz et al., 2010). There are also studies that suggest that over-season (spring and autumn) air pollution is associated with an increased risk of suicide (Bakian et al., 2015).

European and American countries have a clear understanding of mental illness, and will make reasonable suggestions for this weather in weather forecasting. Air pollution is very serious in our country, and haze has been paid more and more attention. People pay attention to haze, and choose to wear masks when the air quality is not good, and use air purifiers indoors to reduce the harm of air pollution to health. However, the threat of haze to mental health and public mental health is still neglected in China. There is little attention paid to mental health in China. There is almost no research on the relationship between air pollution and mental health in China, and corresponding early warning measures have not been set up. But according to the online health survey of domestic portals, haze can make people fear, anxious and depressed. Some counseling rooms in China claim that patients in haze days will increase by about 10% than usual. Those with serious mental illness will feel unhappy and unhappy because of the haze.

In this paper, suicide data of different cities in China were observed, combined with air quality index (AQI). The results showed that people's suicide risk increased when air quality was poor, and there was a positive correlation between the two. Men are more sensitive to air pollution and are more likely to commit suicide in polluted weather. More attention should be paid to the psychological impact of air pollution and its harm should be actively prevented.

The main contents of the follow-up part of this paper are as follows: The second part mainly introduces the sources and processing methods of the data used in the research. The data are mainly divided into three parts: AQI daily value data of cities, meteorological daily value data, insurance policy data of a large domestic insurance company, and the time and city of suicide claims can be extracted from the data. The following descriptive statistics describe the processed data in detail and briefly analyze the data set. The third part introduces the regression model and results of empirical analysis, and explains the regression results in detail. The last part is the conclusion. It summarizes the main conclusions of this paper and puts forward reasonable suggestions accordingly. It calls on people to pay attention to the psychological impact of air pollution and actively defend against it.

3.2 Data and descriptive statistics

3.2.1 Data sources

AQI data is a daily air quality index issued by China Environmental Monitoring Station from 2014 to 2016 (a total of three years). China's air quality standards come from "Environmental Air Quality Standards" (GB3095-2012) issued by the Ministry of Environmental Protection, and "Technical Regulations for Environmental Air Quality Index (AQI)" (HJ633-2012). The main pollutants involved in the evaluation are fine particulate matter, inhalable particulate matter, sulfur dioxide, nitrogen dioxide, ozone and carbon monoxide. The classification standard of AQI index in China is shown in Table 1.

Table 1 Air Quality Index Levels and Grades

AQI value	Category	Warnings
0-50	1 excellent	Outdoor activity recommended
50-100	2 good	Outdoor activity as normal
101-150	3 lightly polluted	Sensitive individuals avoid outdoors
151-200	4 moderate	Sensitive individuals will be affected
201-300	5 heavy	Everyone should avoid being outdoors
>300	6 very heavy	Do not go outside

Suicide data come from more than 10 million insurance policies of a large insurance company in China. The observation period is from 2012 to 2016. Firstly, the insurance data are processed, and the insurance company code is used to identify the prefecture-level cities to which the insurance policy belongs, and the cities with AQI data are retained. Finally, 258 cities at or above the prefecture-level are covered. Next, the claim data are processed. According to the reasons for the insurance, the insured's claim is described in detail, supplemented by the conclusion of the insurance company's claim determination (audit opinions). By using keyword search and regular expression, the above-mentioned written records of each claim case are retrieved step by step, and the five-level risk reasons are obtained. After clearing up the claim data, all claim data for suicide are selected and matched with the claim data by the insurance policy number, so as to know the sex of the suicide. So we can know the city, the date of each suicide, and the sex of the suicide.

China Meteorological Station is divided into base station, basic station and general station. This paper takes meteorological data of base station and basic station. The daily observation data of ground meteorological stations downloaded from China Meteorological Data Network are used to determine the prefecture-level cities according to the longitude and latitude of meteorological stations. The data of each station include daily average temperature, daily maximum temperature, daily minimum temperature, relative humidity, rainfall, wind speed and sunshine hours, and can be used to judge whether there is snow or not. The daily data of each prefecture-level city can be obtained by averaging the observation data of each prefecture-level city. The data is matched with the above data set as a control variable. Station number, annual, monthly, daily and daily average temperature (0.1 degrees Celsius), relative humidity (0.01), rainfall (0.1 mm), wind speed (0.1 m/s), sunshine hours (0.1 hours)

Match the above data with cities and dates to get panel data (2014-2016). It contains daily AQI, meteorological data and suicide events in nearly 200 cities. If there are suicides in the city on the same day, it is recorded as 1, otherwise it is recorded as 0. Table 2 lists the sources of the data and the explanations of the variables.

Table 2 Variable definitions and data sources

variable	Variable definition	data sources
Suicide	Does anyone suicide? 1 for existence of suicide in the city and 0 for no suicide	Policy data
Suicide_woman	Whether there is female suicide or not, 1 for existence of woman suicide in the city and 0 for no woman suicide	
Suicide_man	Whether there is male suicide or not, 1 for existence of man suicide in the city and 0 for no man suicide	
Eve_temp	Daily average temperature	China Meteorological Data Network
Humidity	relative humidity	
Rain	rainfall	
Wind	wind speed	
Sun	Sunshine hours	
Snow	Whether it snows or not, 1 for existence of snow during the day and 0 for no snow.	
AQI	AQI Index Value of the Day	Environmental Monitoring of China
LX (AQI)	Value of AQI with a X-day lag	
MeanX_AQI	AQI index (X+1)-day mean	

3.2.2 Descriptive statistics

Descriptive statistical analysis of the variables used in this paper shows that all the data of AQI index are logarithmic. The results are shown in Table 3.

Table 3 Descriptive statistics

variable	N	Maximum value	mean value	minimum value	Median	standard deviation
Suicide	223668	1. 0000	0.0011	0.0000	0.0000	0.0337
Suicide_woman	223668	1. 0000	0.0005	0.0000	0.0000	0.0227
Suicide_man	223668	1. 0000	0.0006	0.0000	0.0000	0.0250
Eve_temp	175083	31. 1000	15. 7058	-13. 4000	17. 8800	10.3252
Htemp	175083	36. 6000	20.8770	-7. 2500	23. 0500	10.2583
LTEMP	175082	27. 6000	11. 6921	-19. 0500	13. 7000	10.8069
Humidity	175075	0.9800	0.7036	0.2450	0.7333	0.1726
Rain	170205	49. 0000	3. 1150	0.0000	0.0000	8. 1155
Wind	175047	6. 2000	2. 1366	0.6000	1. 9000	1. 0339
Sun	175069	12. 5500	5. 3372	0.0000	5. 9000	4. 0831
Snow	175087	1. 0000	0.1135	0.0000	0.0000	0.3172
AQI	221274	5. 5530	4. 3086	3. 2581	4. 2905	0.4754
L (AQI)	221016	5. 5530	4. 3089	3. 2581	4. 2905	0.4754
L2 (AQI)	220758	5. 5530	4. 3093	3. 2581	4. 2905	0.4754
L3_AQI	220500	5. 5530	4. 3099	3. 2581	4. 2905	0.4753
Mean_AQI	221016	5. 5013	4. 3232	3. 3142	4. 3108	0.4487
Mean2_AQI	220758	5. 4482	4. 3328	3. 3557	4. 3219	0.4295
Mean3_AQI	220500	5. 4128	4. 3395	3. 3759	4. 3307	0.4155

3.2.3 Matrix of correlation coefficient

The main explanatory variable is AQI index. Table 4 is the correlation coefficient matrix of AQI and AQI mean. The correlation coefficient of AQI and AQI index with two days lag is not high. The correlation coefficients of AQI mean and AQI index of the last two days were 0.93 and 0.94 respectively. In the following regression, the AQI mean of nearly two days and AQI value of one day lag were used to regression respectively, and the results were basically the same.

Table 4 AQI Exponential Coefficient Matrix

	AQI	L (AQI)	L2 (AQI)	Mean_AQI
AQI	1. 0000	0.7479	0.5688	0.9286
L (AQI)	0.7479	1. 0000	0.7478	0.9365
L2 (AQI)	0.5688	0.7478	1. 0000	0.7064
Mean_AQI	0.9286	0.9365	0.7064	1. 0000

The main control variables in this paper are the daily values of meteorological data. The correlation coefficient matrix of all control variables is calculated (see Table 5). The correlation coefficient between meteorological data and AQI is small, and the absolute value is less than 0.3. The correlation coefficient between meteorological data is not high, and the correlation coefficient is basically lower than 0.5.

Table 5 Explains the correlation coefficient matrix of variables

	Eve_temp	Humidity	Rain	Wind	Sun	Snow	AQI
Eve_temp	1. 0000	0.3248	0.1543	-0.0905	0.0895	-0.6872	-0.2299
Humidity	0.3248	1. 0000	0.3980	-0.2108	-0.5870	-0.2825	-0.2963
Rain	0.1543	0.3980	1. 0000	0.0144	-0.3600	-0.1188	-0.2972
Wind	-0.0905	-0.2108	0.0144	1. 0000	0.1095	0.0604	-0.1178
Sun	0.0895	-0.5870	-0.3600	0.1095	1. 0000	0.0506	0.1236
Snow	-0.6872	-0.2825	-0.1188	0.0604	0.0506	1. 0000	0.1325
AQI	-0.2299	-0.2963	-0.2972	-0.1178	0.1236	0.1325	1. 0000

3.3 Empirical Analysis

3.3.1 Logistic regression model

Considering the linear relationship between the occurrence of urban suicide and air quality index, the following is considered:

$$Suicide = AQI + city + year*month$$

Among them, suicide indicates whether someone committed suicide in the first city on a certain day, with a suicide score of 1 and no suicide score of 0. AQI is the daily air quality index of each city, which is the main explanatory variable in this regression. City is a city fixed effect. The data set in this paper contains 258 cities above Prefecture level. Year * month is a time-fixed effect that generates different virtual variables every month.

Existing studies have shown that pineal gland can perceive the degree of sunshine exposure and affect emotional state through hormone secretion. Changing weather, extreme weather and so on can affect people's mood and mood. The above regression model does not control the daily weather conditions, so this paper based on the above regression, and added meteorological data as the control variable for regression. The regression model is as follows:

$$Suicide = AQI + Weather + city + year*month$$

Meteorological data are added as control variables, and the fixed effect remains unchanged. Weather refers to temperature, relative humidity, rainfall, wind speed, sunshine hours, extreme weather (whether snow) and other variables.

In order to further study the effects of air pollution on different genders, the dependent variables were replaced by man_suicide and woman_suicide, respectively, and the last regression was repeated. If there is a male suicide on that day, the man_suicide score is 1, otherwise it is 0. Similarly, women_suicide indicated whether there had been a female suicide. The sensitivity of men and women to air pollution can be compared by gender-based regression results.

3.3.2 Regression results

The main explanatory variable of this paper is the AQI index, which focuses on suicide incidents in corresponding cities. The first four columns of regression in Table 6 are AQI index of the same day, AQI index of one day, AQI index of two days and AQI index of three days. The coefficients of AQI and AQI index with one day lag were significantly positive at the level of 10% and 5% respectively. Regression with AQI index of two and three days lag showed that the coefficient was negative and the result was not significant. The regression results showed that the regression coefficients of AQI and AQI were significant only on the same day, but there was no significant correlation between AQI and suicide events before a longer time interval. The latter three columns of regression, at the same time increase the day, lag one, two, three days AQI index. On that day, AQI did not show significant results in these three regressions. The coefficients of AQI index with one day lag were significantly positive, 0.39, 0.63 and 0.61, respectively. The regression coefficients of AQI index with three days lag are negative in the fourth and seventh columns, and are not significant.

In addition to using lag days data, the author also used AQI index mean regression from the past two days to the past 15 days. The AQI mean over the past two days showed a regression coefficient of 0.39 and significant at the level of 5%. The AQI mean values of the past three and four days were 0.29 and 0.15 respectively, but the regression coefficients were not significant. Regression coefficients were not significant with longer AQI mean regression.

Table 6 AQI Index and Suicide Events

Variable	Estimate	Estimate2	Estimate3	Estimate4	Estimate5	Estimate6	Estimate7
Intercept	-15. 544	-16. 122	-14. 3291	-13. 685	-16. 2131	-15. 4882	-15. 1525
	(115)	(109. 3)	(126. 2)	(132. 7)	(105. 4)	(109. 2)	(115)
AQI	0.2762*				0.0527	0.0191	0.025
	(0.1632)				(0.2005)	(0.2015)	(0.2018)
L (AQI)		0.4093**			0.3786*	0.6305***	0.6124**
		(0.1633)			(0.2015)	(0.2429)	(0.2441)
L2 (AQI)			-0.0072			-0.3887*	-0.2695
			(0.1638)			(0.2071)	(0.2462)
L3 (AQI)				-0.1583			-0.1864
				(0.1641)			(0.2054)
Municipal Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: "*",**"and"***"are significant at 10%, 5% and 1% levels respectively.

China's AQI classification, 101-150 for light pollution, 151-200 for moderate pollution, 201-300 for heavy pollution, and > 300 for serious pollution. The main explanatory variable was the number of days in which the AQI index exceeded 100, 150, 200, 250 and 300 in the past few days (survival variables in the past 1-15 days respectively). Regression results showed that when AQI threshold was 200, 250 and 300, the regression coefficient was not significant. When the number of days with AQI exceeding 150 thresholds in the past day was the main explanatory variable, the regression coefficient was 0.24 and significant at the level of 5%, while the regression coefficient for the rest of the time was not significant. When the threshold value is 100, the AQI is more than 100 days in two days and more than 100 days in three days. The regression coefficients are 0.22 and 0.14, respectively, and both of them are significant at the level of 5%. Regression over a longer period of time, regardless of the threshold value, the results are not significant.

In conclusion, suicide incidents were positively correlated with air quality in the last two days, but not earlier. The results are consistent with the existing conclusions of foreign studies. Mild pollution has been positively correlated with suicide. It can be seen that people are more sensitive to air pollution, and suicide can occur without serious pollution. The above regression only controls the fixed effect of time and city, but from the perspective of pineal gland sensitization, meteorological conditions can also affect people's emotions, which has a certain impact on the occurrence of suicide. In the regression, the daily average temperature, relative humidity, rainfall, wind speed, sunshine hours, whether there is snow and other control variables are added. The regression results are shown in Table 7.

Table 7 AQI Index, Meteorological Data and Suicide Events

Variable	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate
Intercept	-16. 1238	-16. 176	-16. 2046	-15. 9897	-16. 0955	-16. 3789
	(124. 3)	(114)	(118. 9)	(119. 3)	(106. 7)	(119. 8)
L (AQI)	0.5324***	0.5422***	0.5436***	0.545***	0.5452***	0.5471***
	(0.1803)	(0.1808)	(0.1839)	(0.1841)	(0.1839)	(0.1839)
Eve_temp	-0.0459**	-0.0464**	-0.0493**	-0.0714**	-0.072**	-0.0722**
	(0.0217)	(0.0218)	(0.0222)	(0.0292)	(0.0292)	(0.0292)
Eve_temp2	0.000937	0.000806	0.000875	0.00135*	0.00137*	0.00137*
	(0.00068)	(0.000708)	(0.000718)	(0.000821)	(0.000822)	(0.000822)
Sun		0.0129	0.01	0.0103	0.00919	0.0152

		(0.0189)	(0.0204)	(0.0204)	(0.0204)	(0.0244)
Rain			-0.00556	-0.00554	-0.0062	-0.00722
			(0.0107)	(0.0107)	(0.0107)	(0.011)
Snow				-0.4596	-0.4544	-0.4544
				(0.4081)	(0.4081)	(0.4086)
Wind					0.0585	0.0673
					(0.0817)	(0.084)
Humidity						0.3139
						(0.6942)
Week1	-0.4238**	-0.4222**	-0.4027**	-0.4034**	-0.4054**	-0.4055**
	(0.2022)	(0.2022)	(0.2025)	(0.2025)	(0.2025)	(0.2025)
Week2	0.461***	0.4604***	0.4158***	0.4157***	0.414***	0.4133***
	(0.1428)	(0.1428)	(0.1467)	(0.1467)	(0.1467)	(0.1467)
Week3	0.0367	0.0361	0.0213	0.0197	0.0191	0.0203
	(0.168)	(0.168)	(0.1706)	(0.1706)	(0.1706)	(0.1706)
Week4	0.2979*	0.2979*	0.3219**	0.3238**	0.3251**	0.325**
	(0.1534)	(0.1534)	(0.1537)	(0.1537)	(0.1538)	(0.1538)
Week5	0.0344	0.0332	0.0511	0.052	0.0536	0.0541
	(0.168)	(0.1681)	(0.1684)	(0.1683)	(0.1684)	(0.1684)
Week6	-0.2649	-0.2659	-0.248	-0.2481	-0.2473	-0.2479
	(0.1904)	(0.1904)	(0.1906)	(0.1906)	(0.1907)	(0.1907)
Municipal Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Year-month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes

Note: "*",**"and"***"are significant at 10%, 5% and 1% levels respectively.

On the basis of the fixed effect, the first column of regression added the daily average temperature as the explanatory variable, and the second term of the daily average temperature, the first term coefficient was significantly negative. The second column regression further increases the sunshine hours as explanatory variables, and the coefficient of sunshine hours is not significant. In the third and sixth columns, the meteorological variables such as precipitation, snow, wind speed and relative humidity are increased in turn, and the coefficients of these variables are not significant. The coefficient of AQI is significant at 1%. The coefficient of daily average temperature is significant, the coefficient of quadratic term is negative, the coefficient of primary term is positive, the influence of extreme temperature is greater, and the influence of suitable temperature is less. The daily mean temperature is replaced by the daily maximum temperature and the daily minimum temperature respectively. The regression results are consistent with the existing results. For every standard deviation of AQI, the suicide rate increased by 26%.

Next, a cohort of observations was used to calculate the proportion of suicides due to increased air pollution. Among them, if the air quality at the time of suicide or the day before suicide was slightly polluted or above (AQI > 100), the number of molecules would be increased by 1. From the data, we can calculate that the proportion of suicide is 43. 4%. According to a report released by the Beijing Center for Psychological Crisis Research and Intervention ("Suicide Situation and Countermeasure in China"), 287,000 people die of suicide every year in China (data source, NetEase News). By multiplying the above ratio with the number of suicides, it can be estimated that the annual air pollution-related suicide deaths in China are about 125,000.

In addition, the National Death Cause Surveillance System established by the China Center for Disease Control and Prevention and the Center for the Prevention and Control of Chronic Non-communicable Diseases has a total of 605 surveillance points. The total population under surveillance is over 300 million, accounting for about 24% of the national population. The data set of death cause monitoring in China published by the center is classified by ICD-10 coding. The number of suicides and sequelae deaths is 18 865, which estimates that the number of suicides in China is about 78 604. Multiplicating with the suicide rate of air pollution, the number of suicides caused by air pollution is about 34,000.

In order to study the sensitivity of different sex groups to air pollution, the regression results were divided into female suicide and male suicide. The first and second columns of Table 8 were the regression results of male suicide, and the third and fourth columns were the regression results of female suicide. The first column only adds AQI index and daily average temperature as explanatory variables, and the coefficient of AQI index is significantly positive at the level of 5%. In the second column, all meteorological data are added as explanatory variables, and the coefficient of AQI index is significantly positive at the 1% level. The third column only adds AQI index and daily average temperature as explanatory variables. The coefficient of AQI index is significantly positive at the level of 10%. The AQI coefficient of the fourth column is not significant.

Table 8. AQI index and suicide incidents in men and women

Variable	Estimate_man	Estimate_man	Estimate_woman	Estimate_woman
Intercept	-19. 4187	-18. 4277	-18. 5552	-19. 6177
	(193. 7)	(116. 5)	(108. 2)	(100.9)
L (AQI)	0.5895**	0.6394***	0.4627*	0.4332
	(0.2374)	(0.2412)	(0.2783)	(0.2859)
Eve_temp	-0.0362	-0.0715**	-0.0386	-0.0486
	(0.0265)	(0.0357)	(0.0403)	(0.0549)
Eve_temp2	0.002**	0.00222**	-0.00076	-0.00027
	(0.000844)	(0.000993)	(0.00123)	(0.00157)
Sun		0.0139		0.0212
		(0.0326)		(0.0369)
Rain		-0.0179		-0.00292
		(0.0188)		(0.0138)
Snow		-0.5575		-0.1942
		(0.5225)		(0.6707)
Wind		0.0569		0.0691

		(0.1104)		(0.1288)
Humidity		-0.2317		1. 172
		(0.906)		(1. 0961)
Week1	-0.2805	-0.2668	-0.5939*	-0.5794*
	(0.253)	(0.2533)	(0.339)	(0.3393)
Week2	0.2756	0.2402	0.7057***	0.6442***
	(0.2038)	(0.208)	(0.2037)	(0.2102)
Week3	-0.3504	-0.3391	0.4354*	0.4026*
	(0.2619)	(0.2622)	(0.2247)	(0.2299)
Week4	0.3805*	0.4029**	0.2188	0.2446
	(0.1972)	(0.1977)	(0.2462)	(0.2466)
Week5	0.1579	0.1746	-0.1184	-0.1019
	(0.2117)	(0.2121)	(0.2791)	(0.2795)
Week6	-0.1886	-0.1745	-0.3278	-0.3112
	(0.2452)	(0.2455)	(0.3038)	(0.3041)
Municipal Fixed Effect	Yes	Yes	Yes	Yes
Year-month fixed effect	Yes	Yes	Yes	Yes

Note: "*",**"and"***"are significant at 10%, 5% and 1% levels respectively.

The coefficients in columns 1 and 2 of Table 8 are larger than those in the latter two. It can be seen that men are more sensitive to air pollution, air quality index increases, and suicide incidents of men increase more, and this positive correlation is significant. Existing research also shows that men are at the highest risk of suicide after exposure to short-term air pollution. Our results also show that men are at a higher risk of suicide than women after exposure to air pollution, and the results are significant.

3.4 Conclusion

In this paper, we use the insurance data of a large insurance company in China to observe more than 10 million people. The correlation between air pollution and suicide was studied by logistic regression. The results showed that there was a significant positive correlation between the occurrence of suicide and AQI index, with a lag of 0-3 days. This conclusion is basically consistent with the existing research results abroad. It can be seen that short-term exposure to polluted air will affect the mood of residents, and even cause suicide. For every standard deviation of AQI, the suicide rate increased by 26%. The number of suicides caused by air pollution is about 34,000, which should be paid attention to.

European and American countries pay more attention to mental health. In weather forecast, reasonable suggestions will be put forward to the psychologically sensitive people according to the weather conditions on that day. China's air pollution is serious. The residents of our country will feel unhappy and unhappy because of the haze, but this harm has not been paid enough attention. Psychological intervention and people's mental health problems have not received sufficient attention. This paper quantitatively studied the correlation between suicide and AQI index. The results showed that air pollution was significantly correlated with suicide events. The conclusion shows that the effect of haze on mental health should be paid more attention. This conclusion provides a scientific basis for decision-making to reduce the occurrence of suicide. The society should take measures to help people solve their mental health problems so as to reduce the occurrence of suicides. In highly polluted weather, risk prevention tips can be increased and psychological counseling can be guaranteed. We should pay attention to the impact of air pollution on mental health, actively prevent and reduce pollution from the source, so as to ensure the physical and mental health of residents.

Chapter 4: Air Pollution and Insurance Purchase Decision

—— Risk Perception and Prospect Theory

Absrtact: What factors affect insurance purchase decision-making and whether information asymmetry exists in the insurance market are basic issues in Insurance Research field. This paper mainly uses COX proportional risk regression and quasi-unrelated regression to study the influencing factors of insurance purchase decision. The air pollution level in the city where the insurer stay will affect the insurance purchase decision. As air pollution becomes heavier, more people are likely to purchase insurance. High-risk individuals usually choose insurance products with high insurance coverage and low premium, which reflects the adverse selection of China's health insurance market.

4.1 Introduction

Information asymmetry is an important research field of insurance economics. Information asymmetry mainly discusses moral hazard and adverse selection (Dionne et al., 2000). Policy holders have more information on their risk probability than insurance companies do. That is to say, the insurance company can not observe what kind of risk group each person is when buying insurance, but the insured knows what kind of risk group he belongs to. There are some classic papers on adverse selection (Hoy, 1982; Akerlof, 1970; Rothschild and Stiglitz, 1976), which also point out that information asymmetry has an impact on welfare (Dahlby, 1980). The insured with different risks in the insurance market, if the cost of insurance is zero and the risk of insurance company is neutral. The high-risk group will buy full insurance under actuarial fair premium, while the low-risk group will buy partial insurance under actuarial fair premium (Wilson, 1977). The model has been further extended in subsequent studies (Miyazaki, 1977; Spence, 1978).

The insurer can not know the risk level of each person, and can only get the average risk for all the insured through empirical calculation. The insured knows his/her risk level, and the person with higher risk will buy insurance, because they are willing to pay more premiums; the person with lower risk will not buy insurance, because the premium they are willing to pay is also low. Whether the insurance market is based on private information to make purchasing decisions, as assumed by the information asymmetry related model, remains to be attested. Therefore, many researchers have made empirical studies on adverse selection in the insurance market. The earliest empirical study of adverse selection, using data from the Canadian automobile insurance market, found that there is adverse selection (Dahlby, 1983). Some scholars studied the French automobile insurance market with parametric and non-parametric methods and found that there was no information asymmetry (Chiappori and Salanie B, 1997). American scholars analyzed the new insurers who entered the automobile insurance market, such that insurance companies know least about their risk level. The lower the purchase of deductibles was, the higher the number of subsequent claims was (Cohen, 2005).

Empirical studies on adverse selection in foreign property and health insurance markets have not yet drawn a unified conclusion on the existence of information asymmetry (Pierre, 2000). Empirical studies of overseas insurance markets have yielded many results (Dionne G et al., 2013; Cohen and Siegelman, 2010; Puelz and Snow, 1994; Barseghyan L et al., 2013; Fang H et al., 2008). Scholars have used additional medical insurance data to confirm that those who spend more on health care are more willing to buy additional insurance (Spinnewijn, 2013). However, China's insurance market started late, and there are many differences from foreign countries, and the conclusions are not necessarily consistent. Therefore, it is very important to study whether there is adverse selection in the insurance market, which is of great significance to the development of the insurance industry and product innovation of insurance companies.

For policyholders, insurance products have great specialty and complexity. Rational consumers need to carry out a large number of information collection, comprehensive product evaluation, prudent purchase decision-making and serious post-purchase evaluation and other stages. Behavioral insurance integrates psychology into insurance science. It explains and predicts various decision-making behaviors in the insurance market from the micro-individual behavior and the motivation of individual behavior. Nowadays, the Internet is well-developed, the cost of information acquisition is very low, and the information people have access to is at explosive growth. From the perspective of dissemination, negative news disseminates much faster and more widely than positive news. Many negative reports spread explosively, while positive reports can not get widespread attention of the whole society. The dissemination of negative news will hinder the development of the industry, but not for the insurance industry. Widespread haze weather in China has aroused widespread concern in society. People have begun to realize the health hazards of the haze and worry about its impact on health. The dissemination of negative information will Increase fear, but increase the demand for insurance. Insurance consumption decision-making is directly related to the risk perceived by the insurer.

Psychological sociology has proved that the individual's coping behavior, especially risk decision-making, will change with the risk perception after suffering catastrophe. After 9-11, many people chose to drive rather than fly for long-distance travel. But in fact, airplanes are three times safer than driving. Air pollution has existed in China for many years, but according to the search volume of search engines, the search volume of air pollution has started to increase gradually since 2013. It reflects that the public's real understanding and concern about haze and air pollution began in December 2013. This information and cognition may increase the individual's risk perception, that is, the subjective probability of the individual increases, leading to an increase in insurance demand.

People mainly rely on intuition to estimate the risk of things, but limited by memory and information, they usually generalize in a partial way. For example, when catastrophes occur and afterwards, tension and fear will spread among the public. However, as time goes by, people tend to ignore the events with low probability and high loss. In prospect theory, people have a subjective evaluation of objective probability, and in decision-making, they will make the best choice based on the subjective risk. Some scholars analyzed insurance consumption behavior from the perspective of subjective risk perception (Wang Jun, Gao Feng, 2008). Psychological, emotional or seemingly unrelated events can change people's risk perception and affect their decision-making behavior.

When individuals are subjected to strong emotional and perceptual shocks, the probability of subjective risk usually increases, which leads to an increase in the insurance demand. Prospect theory studies show that insurance consumption decision-making is directly related to the risk perceived by insurers, but the uncertainty perceived by consumers is not consistent with the objective probability. Risk exists objectively, but risk perception is an individual's subjective understanding of external objective risks. When the level of individual risk perception rises, the subjective probability assessment of a certain risk rises, so the demand for insurance increases.

Empirical research on information asymmetry in the insurance market with insurance policy data is based on the idea of testing the positive correlation between risk categories and the insurance degree. Usually, in empirical research, whether an accident occurs or not is used as the basis for judging the degree of risk of the insured, and the degree of insurance is reflected in the selection of the insurance amount and the deductible amount. People with higher risk tend to buy higher premiums and choose lower deductibles. Combining with the subjective probability mentioned in behavioral insurance, adverse selection also reflects that individuals who consider themselves at higher risk are more willing to buy insurance and insurance products with higher insurance coverage.

Some scholars used the data of health insurance policies of a large domestic insurance company, and found that individuals who claimed after the accident would choose a lower premium in advance and usually bought additional insurance (Wang Jun, Gao Feng, 2007). Using the data of a domestic property insurance company's automobile insurance policy and the number of claims after the event to posteriorily identify the high-risk individuals, the research showed that the high-risk individuals purchase high insurance and low deduction (Peng Fei, 2004). This paper uses the policy data of a large domestic insurance company from 2014 to 2016, and uses COX proportional hazard model to study the influencing factors of insurance purchase decision, including personal information, claim situation, weather condition and air quality condition of the city. And the proportional risk model is used to study the influencing factors of insurance amount, premium rate and insurance interval. It is found that adverse selection exists in China's health insurance market. At the same time, the decision-making of insurance purchase is also related to the air pollution situation. As a result of the mass media coverage of the impact of air pollution on health, the risk perception of residents in heavily polluted cities will increase, and the purchase of insurance policies will also increase.

The main contents of this paper are as follows: The second part describes the data sources used in this study, and introduces the specific processing methods in detail. The data used in this paper mainly come from three sources: the AQI daily value data of cities published by the Ministry of Environmental Protection; the meteorological daily value data published by the State Meteorological Administration; and the insurance policy data of a large insurance company in China, which mainly provides individual information. And the descriptive statistics of these data is set after processing. The third part introduces COX proportional risk regression model, and analyses the regression results in detail. On this basis, without considering personal information, we only study the impact of policy purchase and AQI daily value from the urban level. The last part summarizes this paper and provides reference for the supervision and development strategy of China's commercial health insurance market.

4.2 Data and descriptive statistics

4.2.1 data sources

Major diseases generally refer to more common diseases with high severity and cost of treatment, such as malignant tumors, acute myocardial infarction, sequelae of stroke, etc. Major disease insurance originated in South Africa. Since the 1980s, major disease insurance has been introduced into Europe and Southeast Asia. In China, middle-income and high-income people and private owners are the main buyers of major disease products. Since it was introduced into the Chinese market in the mid-1990s, insurance companies have launched major disease insurance products with their own characteristics, and serious disease insurance has developed rapidly. By the end of 2005, there were more than 36 million serious illness insurance policies in mainland China. At the end of 2010, the number of effective insurance policies for major diseases reached 63.7 million, an increase of 78% over the end of 2005 and an average annual increase of 13%. By the end of 2010, the number of effective insurance policies for major disease products in the Chinese market was about 11 times that in 2000, with an average annual growth of 27%. Major diseases insurance covers more than 30 common major diseases and provides strong support for the multi-level development of China's medical security system. Nowadays, major disease insurance is one of the most important insurance types.

Based on the data of a large insurance company in China, the company's market share in 2013-2015 is 30.4%, 26.1% and 23%, which is the highest in China's life insurance market share. The insurance purchase data of the company are used for analysis, which is representative. In order to ensure that the insured know enough about the insured information, only the insured is selected as the insured's own policy data. Referring to the classification of the insurance company, the occupational categories of the insured are divided into 22 categories: catering and tourism, service industry, public utilities, family management, construction engineering, transportation, education, financial insurance, mining and quarrying, timber and forestry, agriculture and animal husbandry, commerce, health, culture and education, press and publication advertising, general occupation, fishery and entertainment. Manufacturing industry, public security personnel, and others. Because this paper only considers the insured who buy insurance for themselves, they are all capable persons, and have jobs or income sources to pay premiums. There is no case of neonatal insured.

Gender was the categorized variable, 1 was male and 0 was female. The insured's income unit is 10,000 yuan, and the abnormal value of income is deleted from the observed data. The basic insurance amount is the amount chosen by the insured. The premium rate is the ratio of the amount of insurance to the premium, indicating the amount of insurance purchased by the insured per unit of premium. As the study of serious illness insurance includes lifelong, fixed interval and other different insurance intervals, the following unified insurance intervals. Calculate the time interval between the effective date and the insurance period, and the time interval of life insurance calculated in the data is more than 1000 years. For these policies, the insurance period is regarded as 90 years. For adults, the insurance period of 90 years is sufficient to reflect the life-long guarantee. For other policies that are insured to a fixed age, the calculated insurance period is an integer within a reasonable range. If the insurance is up to 80 years old, the difference between the age at which the insurance is purchased and the age at which the insurance is purchased is calculated as the duration of the insurance.

Using the date of birth and purchase of the policy, the age of the policy holder at the time of purchase is calculated. The Age used here is full age. From birth, there is 1 year for every year of life, but not for less than one year. Because the premium of serious illness insurance is related to the purchase age, the COX proportional hazard model is used to study the cohort. When the insured enters the observation set on the day of the insured's birthday, the observation is completed until the purchase date, and the birth of the variable is obtained for the length of the observation. In order to ensure the robustness of the results, we also use another way to start observation. From January 1 of each year, we record the purchase time of each policy as the date of that year, and get the variable Jan 1st.

If there is no claim, claim will be counted as 0. If there is no claim, claim will be counted as 0. If there is a claim, claim will be counted as 0. If the claim date and the effective date are more than two years, claim will be counted as 0. If the time between two points is less than two years, claim will be counted as 1. The data set is up to 2016. To ensure that the observation period expires for two years, the last two years (2015 and 2016) data are deleted. All variables reflecting personal information and personal purchase decisions are derived from policy data. The following are urban data, including air quality and meteorological data.

This study uses AQI to reflect the air quality of cities. AQI data is a daily air quality index issued by China Environmental Monitoring Station from 2014 to 2016 (a total of three years). China's air quality standards come from "Environmental Air

Quality Standards" (GB3095-2012) issued by the Ministry of Environmental Protection, and "Technical Regulations for Environmental Air Quality Index (AQI)" (HJ633-2012). The main pollutants involved in the evaluation are fine particulate matter, inhalable particulate matter, sulfur dioxide, nitrogen dioxide, ozone and carbon monoxide. The grading standard of AQI index in China is shown in Table 1. The AQI value of more than 100 is polluted weather, and that of less than 100 is non-polluted weather.

Table 1 AQI Index Classification Criteria

AQI index	Grade	Matters needing attention
0-50	1Grade optimization	Take part in outdoor activities to breathe fresh air
50-100	2Grade Liang	Outdoor activities can be carried out normally.
101-150	3Mild grade	Sensitive people reduce physical exertion in outdoor activities
151-200	4Moderate grade	Greater impact on sensitive population
201-300	5Grade severe	Everyone should reduce outdoor activities appropriately
>300	6Serious grade	Try not to stay outdoors

People's decision-making may be affected by emotions. Existing studies have shown that weather conditions may affect people's emotions and decision-making. For example, light-sensitive pineal gland can make people feel depressed in less sunshine. Therefore, the meteorological conditions of each city are controlled in order to ensure the reliability of the results. The meteorological data in this paper are the monitoring data of meteorological stations. Chinese meteorological stations are divided into base stations, basic stations and general stations. In this paper, the meteorological data of base stations and basic stations are taken, and the daily observation data of ground meteorological stations downloaded from China Meteorological Data Network are used to determine the prefecture-level cities according to the longitude and latitude of meteorological stations. The data of each station include daily average temperature, relative humidity, rainfall and sunshine hours, and can be used to judge whether there is snow or not. The daily data of each prefecture-level city can be obtained by averaging the observation data of each prefecture-level city. The data is matched with the above data set as a control variable. Station number, annual, monthly, daily and daily average temperature (0.1 degrees Celsius), relative humidity (0.01), rainfall (0.1 mm), sunshine hours (0.1 hours)

Match the above data with cities and dates to get panel data (2014-2016). It contains daily AQI and meteorological data of more than 200 cities. Then the city data and personal data are matched to match 1495 237 insurance policies successfully. Only the insurance policies which are purchased for themselves are retained. Finally, 742709 observation data, including 224 cities, are obtained. The source and meaning of all variables are shown in Table 5. 2.

Table 2 Sources and Implications of Variables

Individual characteristic variable of insured person	Variable definition	data sources
Claim	Whether there is a claim within two years of insurance purchase is 1 or 0.	Serious Disease Insurance Policy Data
Birth	The time point at which the policy was purchased (from the beginning of the observation to the end of the observation)	
Priem_time	Insurance period	
Priem	premium rate	
Policy_count	Basic insurance amount	
Occup	Policy holder occupation	
Buy_age	Age at the time of insurance purchase	
Income	Insured's Income	
Gender	The insured sex is 1 male and 0 female.	
City variable		
AQI	Daily air quality index	Environmental Monitoring of China
Mean_temp	Daily average temperature	China Meteorological Data Network
Mean_humidity	Daily average humidity	
Rain	rainfall	
Snow	Does it snow?	
Sunlight	Sunshine duration	

Next, descriptive statistics are made on the above variables.

4.2.2 descriptive statistics

This part gives the descriptive statistics of the variables used in the following part. Table 3 shows that the standard deviation of the insured's income is large. Therefore, in the later regression, the abnormal value of excessive income is deleted for regression.

Table 3 Descriptive statistics of variables used in the model

Variable	Obs	Mean	Std. Dev.	Min	Max
birth	719,392	197.069	107.328	1	364
Jan 1st	719,808	193.242	96.505	1	364
AQI100-10	688,314	3.215	2.958	0	10
AQI100-20	672,717	6.394	5.334	0	20
AQI100-30	650,893	9.473	7.587	0	30
AQI150-10	688,314	1.101	1.923	0	10
AQI150-20	672,717	2.123	3.359	0	19
AQI150-30	650,893	3.084	4.667	0	27
AQImean-10	719,808	94.855	38.205	27.9	296
AQImean-20	719,808	95.167	35.864	28.55	296
AQImean-30	719,808	95.697	35.449	28.7	296
policy_count	719,808	1.500	0.941	-5.591	6.908
priem_time	719,808	62.471	24.785	5	90
log_priem	719,808	3.868	0.931	-0.896	9.208
claimyear2	719,808	0.004	0.064	0	1
insure_income	719,808	2.146	0.701	0.693	4.605
insure_gender	719,808	0.374	0.484	0	1
insure_age	719,808	37.669	8.378	15	66
mean_temp	719,808	17.690	9.761	-13.3	30.8
mean_humid	719,778	0.708	0.167	0.26	0.97
mean_snow	719,808	0.084	0.277	0	1
mean_rain	697,760	3.370	8.740	0	53.233
mean_sun	719,683	5.297	4.026	0	12.5

In this paper, based on AQI, a series of variables, mainly threshold variables, reflect the level of urban air pollution, focusing on the number of days of continuous pollution. The pollution mentioned in this paper is that the daily value of AQI is higher than the threshold value. The air pollution variable is obtained by recording the days of pollution. This variable can reflect the air pollution situation, and the days of air pollution will be longer if the air pollution is serious. The correlation matrix of regression variables is calculated below.

4.2.3 Correlation matrix

The urban air quality and meteorological data are put into the regression as control variables. It can be seen that the correlation coefficient between meteorological data and AQI is very small, and there is no collinearity in the regression.

Table 4 Coefficient Matrix

	birth	Jan 1st	AQI100-30	AQI150-30	AQImean-30	temp	snow
birth	1						
Jan 1st	0.0175*	1					
AQI100-30	-0.0236*	-0.1849*	1				
AQI150-30	-0.0191*	-0.1835*	0.8214*	1			
AQImean-30	-0.0236*	-0.3320*	0.9299*	0.9141*	1		
temp	0.0170*	0.1019*	-0.3483*	-0.4547*	-0.5099*	1	
snow	-0.0018	-0.1083*	0.2104*	0.2836*	0.2919*	-0.6521*	1

Using the average of AQI over the past 30 days, the number of AQI over 100 days over the past 30 days and the number of AQI over 150 days over the past 30 days, these variables are respectively put into the regression equation as indicators of air pollution status. From the correlation coefficient matrix, it can be seen that the correlation coefficients of these variables are greater than 0.8.

The above data sets are described in detail, from the data sources, processing methods. The descriptive statistics of the processed data are also given. In addition, the correlation coefficient matrix of the used control variables is given. Next, we will use the data set to carry out empirical analysis. First, we introduce the COX model used in regression. From the perspective of individual insurance policies, we will study the factors affecting policy purchase decisions. Then from the urban level, the relationship between the purchase of group insurance policies and air quality and weather conditions is studied. The analysis process and main conclusions are introduced in detail. Based on the empirical results, we have a further understanding of China's health insurance market and provide a reference for the development of the insurance market.

4.3 Empirical Analysis

4.3.1 COX proportional hazard model

COX proportional hazard model is often used to analyze data sets containing censored data, mainly used to study factors affecting survival time. In this paper, COX proportional hazard model is used to analyze individual data, and to study what factors affect policy-holders' insurance purchase decisions. These factors are called covariates. Because covariates are related to decision-making time, the time-dependent survival analysis model with covariates (Peng Fei, 2004).

P-dimensional column vectors are column vectors composed of covariates, which depend on time t or constant variables. The time-dependent proportional risk model can be expressed as follows:

$$\lambda\left(t|Z\right)=\phi\left(Z\left(t\right)\right)\lambda_0\left(t\right)$$

The conditional risk rate $\lambda_0\left(t\right)$, which represents Z = 0, is called the benchmark risk rate function. $\lambda\left(t|Z\right)$ is a risk rate function. The risk function is a mono-incremental function of risk factors, and the non-risk factor is a constant. It can reflect the individual's risk level.

The value of a function $\phi\left(\cdot\right)$ is usually positive and is a function of covariate Z. Therefore, the form of function is usually taken as the linear combination of the covariate vector Z and the index, and the form of function is usually taken as the index $\phi\left(Z;\beta\right)=\exp\left(\beta^T Z\right)$. By substituting the above formula and simplifying it, the following models are obtained:

$$\lambda\left(t|Z\right)=\exp\left(\beta^T Z\left(t\right)\right)\lambda_0\left(t\right)$$

Among them, the P-dimension unknown regression coefficient vector β can be estimated by maximizing the following likelihood functions:

$$L\left(\beta\right)=\prod_{i=1}^{n}\left\{\frac{\exp\left(\beta^T Z_i\left(t\right)\right)}{\sum_{j\in R\left(X_i\right)}\exp\left(\beta^T Z_{\mathrm{j}}\left(t\right)\right)}\right\}^{\delta_i}$$

Among them $\delta = \min(T \leq C)$, $X = \min(T, C)$

This paper mainly uses the proportional risk model to carry out empirical research on individual insurance policy data. Next, the covariates used in regression will be described in detail, and the empirical results will be analyzed.

4.3.2 regression model

The COX proportional hazard model is used to conduct a cohort study. The time when the insured buys insurance is the time when the observation ends. There is no literature reference on the timing of entry into the queue, because there is no literature on insurance purchase decision-making at the individual level. Considering that the price of health insurance is age-related, and the price of products is different at different ages, the insured person's birthday is chosen as the time to enter the queue for observation. The regression equation is as follows:

Buy_time (cansor) = AQI + claim + personal + time + city + cross + Weather

In the regression equation, the survival time is the time interval from the beginning to the insured's purchase of insurance, that is, the purchase time. The data used in this paper is Chinese health insurance policy data. The data set contains only insurance purchasers, so there is no deleted data. The cansor variable in the regression model always takes 1. In other words, all observations are data of end-point events.

All personal information in the regression comes from the health insurance policy data. Personal represents the personal information data in the insurance policy data, including the insured's gender, occupation, the amount of insurance purchased, and the insured's income. In addition, the time of purchase is added as a time-fixed effect, and the purchase behavior occurs on the day of the week as a control variable.

Claim is a 0-1 variable. If the insured has a claim within two years of purchasing the insurance policy, claim = 1; if the insured has not made a claim or the claim has been more than two years away from the purchase time, claim = 0. Since there is usually 180 days' waiting period after purchase, two years are chosen as the study interval. Because people may be in poor health before major diseases and malignant tumors occur. If it's thyroid cancer (or other non-fatal cancers), you may have checked for cancer before buying insurance. Therefore, if the claim is made within two years, we believe that the insured has some obvious or potential information when they purchase the insurance, which conveys the signal that their health is not very healthy. A posteriori conclusion is that the claimant is at high risk within two years. This variable is used to examine the effect of the insured's risk level on the purchase time.

In COX survival analysis model, stratified variables can be selected, and the implicit assumption is that the basic risk levels of different levels are different. In this study, because of the different level of insurance development in each city, the depth of insurance purchase varies greatly in each city, and the total amount of insurance purchase varies greatly. Therefore, cities are chosen as stratified variables, and the gap of insurance purchase level among cities is included in the model.

Next, the weather conditions at the time of purchase are taken as control variables, including daily average temperature, humidity, rainfall, whether there is snow, sunshine duration, and the influence of weather conditions on purchase decisions is controlled. Finally, AQI is cross-multiplied with individual characteristic variables to get cross-variable CROSS, and all cross-variables are put into regression. From the regression results, we can see what characteristics of individuals are more sensitive to air pollution. The above regression will study the impact of air pollution level on the purchase time point of people's serious illness insurance, and whether the purchase time point of insurance is affected by other factors.

4.3.3 Empirical results

Since the end of 2013, China has published daily air quality index. The air quality below AQI index 50 is excellent, below 100 is good, over 100 is mild pollution, over 150 is moderate pollution, and over 200 is serious pollution. Tom et al. 's paper on the relationship between air pollution and insurance purchase decision (Chang et al., 2018) uses the AQI index converted from PM2.5 concentration and uses 150 as the threshold of pollution. Therefore, 150 is also used as AQI threshold in this paper. The number of days with AQI exceeding 150 in the 30 days before the purchase date can reflect the air quality situation in the past period, so the number of days with pollution is used as the main explanatory variable. In this paper, the cohort study observers are insurance purchasers. The period before insurance purchases is the starting observation point, the day of insurance purchases is the closing observation point, and the observation period is the time interval between two points. The COX model was used to analyze the effect of pollution days before the purchase date on the purchase decision.

Table 5 shows the number of AQI days exceeding 150 in the 30 days before the purchase of the policy, reflecting the air quality situation in the period before the purchase decision of the policy. Among them, model (1) adds AQI variable, policy purchase decision is the day of the week, time fixed effect, city stratification. The regression coefficient of AQI variable is significantly positive, which indicates that if the air condition in the past 30 days is not good, it will promote insurance purchase behavior. Widespread haze weather in China has aroused widespread concern in society. People begin to realize the haze's health hazards and worry about the haze's impact on health. People mainly rely on intuition to judge and estimate the risk of things, but limited by memory and information, they usually generalize in a partial way. Insurance consumption decision-making is directly related to the risk perceived by the insurer. Over a period of time, air pollution (AQI values are higher than 150 days), subjective risk probability usually increases, which leads to an increase in insurance demand.

In the model (1), only time and city fixed effects were added, but individual characteristics were not controlled. Further, the regression results were obtained by adding personal characteristics information as shown in the model (2). The AQI coefficient in model (2) was still significantly positive, and the gender regression coefficient was significantly positive, indicating that men were more willing to buy insurance. The results of behavioral insurance research show that when faced with uncertainty, female decision makers are more cautious than men (Wan Yan Ruiyun, Suo Lingyan, 2016). The regression results of model (1) showed that men preferred to buy health insurance, possibly because adult males were the backbone of the family and might be the main source of income. Men are more likely to buy health insurance if they lose their ability to work due to major illnesses and have a greater impact on their families.

Table 5 Pollution Days with AQI Value over 150 and Policy Purchase Decision

	(1)	(2)	(3)	(4)	(5)
VARIABLES	birth	birth	birth	birth	birth
AQI150-30	0.001**	0.001**	0.001**	0.003***	0.007*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.004)
insure_income		-0.008***	-0.008***	-0.009***	-0.009***
		(0.002)	(0.002)	(0.002)	(0.002)
insure_gender		0.015***	0.015***	0.016***	0.018***
		(0.003)	(0.003)	(0.003)	(0.003)
insure_age		0.001***	0.001***	0.001***	0.002***
		(0.000)	(0.000)	(0.000)	(0.000)
claimyear2			0.015	0.011	0.022
			(0.019)	(0.019)	(0.024)
mean_eve_temp				0.002***	0.002***
				(0.000)	(0.000)
mean_humidity				-0.025**	-0.026**

				(0.013)	(0.013)
mean_snow				-0.009	-0.009
				(0.007)	(0.007)
mean_rain				0.000*	0.000*
				(0.000)	(0.000)
mean_sun				-0.001	-0.001
				(0.000)	(0.000)
cros_income					0.000
					(0.000)
cros_gender					-0.001
					(0.001)
cros_age					-0.000***
					(0.000)
cros_claim					-0.003
					(0.004)
Observations	650,517	650,517	650,517	629,216	629,216

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

In the result of model (2), the regression coefficient of income is significantly negative, and the high-income people are more reluctant to buy insurance. This study uses the policy data of commercial health insurance. In China, the main customer group of commercial health insurance is the middle and high income group, which is the general consensus of the industry. High-income people may have other ways of protection, or have been more comprehensive protection, rather than choose to buy insurance. The regression coefficient of the insured's occupation is basically not significant. Almost all health insurance is based on the age of the insured. The higher the age, the higher the premium, because the older the age, the higher the risk of major diseases. The regression coefficient of the policy holder's age is significantly positive, and the older policy holder is more willing to buy insurance. Maybe it's because older people tend to pay more attention to health, because they feel that health is important and they are more willing to buy insurance. The regression results of all personal information variables are consistent with the existing studies, and the regression results are reasonable.

Next, the regression result (3) is obtained by adding the variable of whether to claim or not in two years after insurance. This paper only studies the insurance policy for myself. Usually, I will have a better understanding of my health. If the insured person is out of danger two years after the insurance, the insured person's physical condition may have been out of danger at the time of the insurance, indicating that his or her health has been in sub-health. The regression coefficient of the claim is not significant in two years, and the individual risk level has nothing to do with the time point of insurance purchase.

In addition to personal characteristics and air pollution, insurance purchase decisions may also be influenced by weather factors, weather conditions affect people's emotions, and further affect people's behavior. On the basis of model (3), the model (4) controls the meteorological variables, including five meteorological data: daily average temperature, humidity, rainfall, whether snow falls or not, and sunshine duration. After adding meteorological data as control variable, the regression results are still significant and robust.

Furthermore, the AQI values and individual variables are crossed to study which characteristics of people are more sensitive to air quality. Model (5) is a regression result with crossing terms. The regression coefficient of AQI* income cross-term is not significant, and the sensitivity of different income to air pollution is not significantly different. The regression coefficients of AQI* gender cross-terms were not significant, and the sensitivity of male and female insurance to air pollution was the same. AQI* age coefficient was significantly negative, and younger people were more sensitive to air pollution. The health risks of air pollution have not been recognized until 13 years ago. Young people are more likely to pay attention to the information because they understand that air pollution is harmful to health faster and earlier. The information of older people is relatively poor, because our data is in 2014, young people get information faster in a shorter period of time after reporting, and the results are reasonable. The AQI* claim coefficient is also negative, but not significant. There is no significant difference in sensitivity to air quality among the people who claim for insurance within two years, that is, the high-risk group. In order to simplify the typesetting, table 5 omits the regression results of the insured's occupation, AQI* occupation intersection and purchase decisions that take place on the weekdays.

Many variables can be used to describe air pollution. Referring to the AQI classification standard, different pollution levels (100, 150, 200) can be selected as thresholds. For the selected AQI threshold, exposure duration can choose different days. The author uses the pollution days in different time intervals (1-60 days) as explanatory variables. The regression results show that when the AQI threshold is not less than 100 (100, 150, 200 respectively), the purchase of insurance policies is related to air quality, and the regression results are significantly positive. In other words, in the past, the number of days of pollution was usually a factor affecting policy purchase decisions. If pollution is serious, people are more willing to buy health insurance.

Table 6　Different AQI variables and policy purchase decisions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	birth	birth	birth	birth	birth	birth	birth	birth	birth
AQI100-10	0.009*								
	(0.005)								
AQI100-20		0.005							
		(0.003)							
AQI100-30			0.002						
			(0.002)						
AQI150-10				0.015*					
				(0.009)					
AQI150-20					0.009*				
					(0.005)				
AQI150-30						0.007*			
						(0.004)			
AQImean-10							0.001***		
							(0.000)		
AQImean-20								0.001***	
								(0.000)	
AQImean-30									0.001***
									(0.000)
claimyear2	0.036	0.050*	0.053*	0.010	0.023	0.022	0.075	0.110**	0.102*
	(0.028)	(0.030)	(0.031)	(0.022)	(0.023)	(0.024)	(0.049)	(0.051)	(0.053)
insure_income	-0.009***	-0.012***	-0.015***	-0.008***	-0.009***	-0.009***	-0.004	-0.009	-0.013**
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.005)	(0.005)	(0.006)
insure_gender	0.015***	0.016***	0.017***	0.017***	0.018***	0.018***	0.011	0.011	0.013*
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.007)	(0.007)	(0.007)
insure_age	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.003***	0.003***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
cros_income	-0.000	0.000	0.001**	-0.002	-0.001	0.000	-0.000	0.000	0.000
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
cros_gender	0.001	0.000	-0.000	0.000	-0.000	-0.001	0.000	0.000	0.000
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
cros_age	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
cros_claim	-0.008	-0.006*	-0.004	-0.001	-0.006	-0.003	-0.001	-0.001**	-0.001**
	(0.006)	(0.004)	(0.003)	(0.009)	(0.006)	(0.004)	(0.000)	(0.000)	(0.001)
mean_temp	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***

	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
mean_humid	-0.031**	-0.024*	-0.014	-0.040***	-0.033***	-0.026**	-0.033***	-0.030**	-0.025**
	(0.012)	(0.012)	(0.013)	(0.012)	(0.012)	(0.013)	(0.012)	(0.012)	(0.012)
mean_snow	-0.001	-0.003	-0.010	-0.001	-0.002	-0.009	0.005	0.005	0.005
	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
mean_rain	0.000**	0.000**	0.000*	0.000**	0.000**	0.000*	0.000**	0.000**	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
mean_sun	-0.000	-0.000	-0.000	-0.000	-0.001	-0.001	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
timefix	yes	yes	yes	yes	yes	yes	yes	yes	yes
stratified by city	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	665,938	650,685	629,216	665,938	650,685	629,216	697,203	697,203	697,203

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6 uses the number of days in which AQI values exceed a threshold in the past 10, 20 and 30 days as the main explanatory variables. Selecting light pollution (AQI greater than 100) and moderate pollution (AQI greater than 150) as thresholds, the results are positive. Using the AQI mean over the past few decades as the main explanatory variable, the regression results were also significantly positive. In short, insurance purchase decisions are positively correlated with persistent air pollution, and the results are significant.

In Table 6, the threshold value of model (1) (2) (3) air pollution variable AQI is 100. Model (1) is the number of days with AQI exceeding 100 in the past 10 days. Model (2) and model (3) are the past 20 days and the past 30 days, respectively. The main explanatory variables of model (4) (5) (6) are the number of days with AQI exceeding 150 in the past 10, 20 and 30 days respectively. The regression results are more significant and the coefficients are larger. Because AQI over 100 is slightly polluted and over 150 is moderately polluted, moderately polluted people will feel a higher risk of disease, so the possibility of buying insurance is higher. Thus, the higher the threshold is, the greater the regression coefficient is, and the higher the pollution level is, the more willing people are to buy health insurance.

The main explanatory variable of model (1) - (6) is the days of continuous pollution. Model (7) - (9) directly uses AQI mean as the main explanatory variable, and chooses the time interval of 10, 20 and 30 days respectively. The regression coefficients of AQI mean are smaller and more significant. The two air quality variables are mean AQI and threshold AQI. The correlation coefficient matrix of mean

AQI and threshold AQI is calculated, and the correlation coefficient is higher than 0. 8. If the air pollution is serious, the AQI mean and AQI threshold days are both large, and vice versa, both laugh.

This is related to the individual's sensitivity to air quality. Some individuals can fully perceive and make purchasing decisions within a week or two of continuous pollution. Others need a longer period of continuous pollution to purchase health insurance. These individuals are less sensitive. In short, the longer the pollution lasts, the more obvious the individual perceives the air pollution, and the more likely he is to buy insurance.

In Table 6, the regression coefficients of model (7) (8) (9) are smaller than those of model (1) - (6), which are significant. From the regression results, the threshold has a greater impact on insurance purchase decision-making than the mean value. People's perception of air pollution risk is not necessarily due to the short-term exposure effects of short-term persistent polluted weather, but may also be due to the impact of information dissemination on people's risk perception. Subjectively, it is believed that the long duration of pollution has a greater potential risk to one's health. This subjective risk judgment directly affects the decision-making of insurance purchase, so the regression coefficient of the threshold is larger than the mean value.

Cox proportional hazard model will be used to study the cohort of the above regressions. The model studies the relationship between the time point of insurance purchase and air pollution. The policyholder enters the observation set from the birthday before the insurance purchase. The observation duration variable is the time interval from the last birthday to the insurance purchase. Because of the different purchase time points, we can see the influence of different pollution conditions before the purchase time points on purchase decisions. In order to ensure the reliability of the results, different survival time measurement methods were used to regression. Calculating the time interval from the beginning of the year (January 1) to the insurance purchase time, the regression results are shown in Table 7.

The results of the regression in Table 7 are consistent with those mentioned above. It can be seen that the results are not affected by the algorithm of survival time, and the regression results are robust. Next, we study the influence of each variable's standard deviation on insurance purchase decision-making.

Table 7 Different Survival Time Measurement Methods

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Jan1st	Jan1st	Jan1st	Jan1st	Jan1st	Jan1st	Jan1st	Jan1st	Jan1st
AQI100-10	0.031***								
	(0.005)								
AQI100-20		0.049***							
		(0.003)							
AQI100-30			0.061***						
			(0.002)						
AQI150-10				0.076***					
				(0.011)					
AQI150-20					0.106***				
					(0.007)				
AQI150-30						0.119***			
						(0.005)			
AQImean-10							0.006***		
							(0.001)		
AQImean-20								0.017***	
								(0.001)	
AQImean-30									0.030***
									(0.001)
claimyear2	0.020	0.005	0.023	0.051***	0.048***	0.050**	-0.059	-0.069	-0.054
	(0.021)	(0.023)	(0.025)	(0.018)	(0.019)	(0.020)	(0.064)	(0.066)	(0.070)
insure_income	0.009***	0.004*	0.006**	-0.001	0.004*	0.004**	0.018***	-0.018**	-0.040***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.006)	(0.007)	(0.008)
insure_gender	-0.009***	-0.013***	-0.017***	0.004*	-0.004*	-0.007**	-0.106***	-0.123***	-0.120***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.008)	(0.009)	(0.009)
insure_age	0.001***	0.001***	0.001***	0.002***	0.002***	0.002***	-0.004***	-0.003***	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
cros_income	-0.007***	-0.002***	-0.001***	-0.012***	-0.007***	-0.004***	-0.000***	0.000	0.000***
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
cros_gender	0.011***	0.005***	0.003***	0.020***	0.012***	0.006***	0.001***	0.002***	0.001***
	(0.001)	(0.000)	(0.000)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
cros_age	0.001***	0.000***	0.000***	0.001***	0.001***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
cros_claim	0.018***	0.012***	0.006**	0.027*	0.014*	0.009	0.001**	0.002**	0.001
	(0.006)	(0.004)	(0.003)	(0.014)	(0.008)	(0.006)	(0.001)	(0.001)	(0.001)
mean_temp	0.079***	0.094***	0.113***	0.083***	0.101***	0.123***	0.072***	0.086***	0.100***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
mean_humid	-0.477***	-0.502***	-0.282***	-0.618***	-0.662***	-0.440***	-0.520***	-0.360***	-0.077***

	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)	(0.013)
mean_snow	0.250***	0.082***	-0.043***	0.286***	0.116***	-0.003	0.344***	0.337***	0.275***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.009)	(0.010)	(0.010)
mean_rain	0.013***	0.014***	0.014***	0.014***	0.014***	0.014***	0.013***	0.013***	0.013***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
mean_sun	-0.008***	-0.009***	-0.008***	-0.010***	-0.011***	-0.011***	-0.007***	-0.003***	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
timefix	yes	yes	yes	yes	yes	yes	yes	yes	yes
stratified by city	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	666,326	651,062	629,582	666,326	651,062	629,582	697,607	697,607	697,607

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 8 The effect of a standard deviation of variables on policy purchase decisions

Variable	回归系数	Mean	Std. Dev.	影响的值
AQI100-10	0.003	3.215	2.958	1.009
AQI100-20	0.002	6.394	5.334	1.011
AQI100-30	0.001	9.473	7.587	1.008
AQI150-10	0.003	1.101	1.923	1.006
AQImean-10	0.000	94.855	38.205	1.000
AQI150-20	0.000	2.123	3.359	1.000
AQImean-20	0.000	95.167	35.864	1.000
AQI150-30	-0.002	3.084	4.667	0.991
AQImean-30	0.001	95.697	35.449	1.036
claimyear2	0.010	0.004	0.064	1.001
insure_income	-0.008	2.146	0.701	0.994
insure_gender	0.016	0.374	0.484	1.008
insure_age	0.001	37.669	8.378	1.008
mean_temp	0.002	17.690	9.761	1.020
mean_humid	-0.010	0.708	0.167	0.998
mean_snow	-0.008	0.084	0.277	0.998
mean_rain	0	3.370	8.740	1.000
mean_sun	-0.000	5.297	4.026	1.000

Table 8 shows that the first half mainly compares the effects of threshold AQI variables and mean AQI variables, and the second half gives the effects of explanatory variables in regression on insurance purchase decisions. Meteorological data are controlled, and AQI mean and AQI threshold are put into the regression equation as explanatory variables to compare the two. The regression coefficients are used to calculate the impact of various variables on insurance purchase decisions.

Calculate the product of the standard deviation of the variable and the regression coefficient, and then take the index of the value to get the last column value in Table 8. What we get is the degree of the influence of each standard deviation of the corresponding variables on the decision-making of insurance purchase. Similarly, we can also get the degree of influence on insurance purchase decision when the variable change is an average value.

Air pollution has existed in China in the early years, but according to the search volume of search engines, the search volume of air pollution peaked at the end of 2013. Later, the search volume was higher than 13 years ago. It can be seen that the public's real understanding and concern about haze and air pollution began in December 2013. Risk exists objectively, and the risk perception is an individual's subjective understanding of external objective risks. Insurance consumption decision-making is directly related to the risk perceived by insurers, but the uncertainty perceived by consumers is not consistent with the objective probability. The subjective probability of the individual increases, which leads to the increase of insurance demand.

Before 2013, the air pollution assessment system used in China was called the air pollution index (API). After 2013, it was changed to air quality index(AQI) and three new pollutant concentration monitoring systems were added, including the widely concerned PM2.5. Next, we use the purchase data of insurance policies before 2013, and use similar methods to process API to get threshold variables that reflect air quality. A total of 700,000 observation data covering 103 cities at or above the prefecture level are obtained. The author took threshold 100, threshold 150 and average AQI respectively and observed different time lengths (1-60 days). The regression results were not significant. Due to layout limitations, only the regression results in Table 9 are shown here. In Table 9, model (1) - (3) takes API 100 as the threshold, and takes 10, 20 and 30 days’ interval respectively. The main explanatory variables of model (4) - (6) are API over 150 days in the past 10, 20 and 30 days respectively. Model (7) - (9) is the API mean over the past few decades, which reflects air pollution in the past even days.

Table 9 API Policy Purchase Decisions

VARIABLES	(1) birth	(2) birth	(3) birth	(4) birth	(5) birth	(6) birth	(7) birth	(8) birth	(9) birth
API100-10	-0.000								
	(0.005)								
API100-20		0.002							
		(0.003)							
API100-30			0.002						
			(0.002)						
API150-10				0.003					
				(0.008)					
API150-20					0.004				
					(0.005)				
API150-30						0.003			
						(0.003)			
APImean-10							0.000		
							(0.001)		
APImean-20								0.001	
								(0.001)	
APImean-30									0.001
									(0.001)
claimyear2	-0.036	-0.019	-0.021	-0.036**	-0.037**	-0.040**	-0.070	-0.062	-0.056
	(0.051)	(0.055)	(0.058)	(0.017)	(0.018)	(0.018)	(0.056)	(0.060)	(0.064)
insure_income	-0.008	-0.011*	-0.011*	-0.003	-0.004*	-0.005**	-0.003	-0.010	-0.016**
	(0.005)	(0.006)	(0.006)	(0.002)	(0.002)	(0.002)	(0.006)	(0.007)	(0.007)
insure_gender	0.029***	0.035***	0.033***	0.026***	0.026***	0.026***	0.037***	0.040***	0.041***
	(0.008)	(0.008)	(0.009)	(0.003)	(0.003)	(0.003)	(0.009)	(0.009)	(0.010)
insure_age	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
cros_income	0.001	0.000	0.000	-0.001	-0.000	0.000	-0.000	0.000	0.000
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
cros_gender	-0.001	-0.001	-0.000	-0.002	-0.001	-0.001*	-0.000	-0.000*	-0.000*
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
cros_age	-0.000	-0.000	-0.000	-0.000***	-0.000***	-0.000***	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
cros_claim	0.001	-0.001	-0.000	0.004	0.003	0.003	0.001	0.000	0.000
	(0.006)	(0.003)	(0.002)	(0.008)	(0.005)	(0.003)	(0.001)	(0.001)	(0.001)
mean_temp	0.002***	0.002***	0.002***	0.001***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
mean_humid	0.000	0.008	0.011	-0.003	-0.002	-0.002	-0.001	0.003	0.005

	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
mean_snow	0.009*	0.011**	0.012**	0.008	0.008	0.008	0.008	0.009	0.009*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
mean_rain	0.000*	0.000**	0.000**	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
mean_sun	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
timefix	yes	yes	yes	yes	yes	yes	yes	yes	yes
stratified by city	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	708,236	708,236	708,236	708,236	708,236	708,236	708,236	708,236	708,236

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

According to the previous analysis, air pollution has existed in the period before 2013, but it has not attracted people's attention, and people do not consciously pay attention to air quality index, so the API regression results are not significant. Comparing the regression results of Table 9, Table 6 and Table 7, the policy-holder's insurance purchase decision is positively correlated with risk perception. Information about the status of air pollution, as well as the impact of air pollution on health, makes people feel that the level of risk rises after decades of continuous pollution and stimulates them to buy health insurance.

The above regression studies the relationship between air pollution and insurance purchase decision. The main explanatory variable is the level of air pollution in the city where the insured is located. The policy-holder's insurance purchase decision can not affect the air quality of his city, so there is no reverse causality in this paper. Referring to the existing research, this paper controls the fixed effect of time and city, and also controls the meteorological data of city. Because this paper is from the individual level, it controls some personal characteristics variables of the insured, and reduces the possibility of missing variables. There is no endogenous problem in this paper.

4.3.4 Seemingly Unrelated Regression and Empirical Results

The above regression results show that the policy-holder's insurance purchase decision will be affected by air pollution, and long-term continuous air pollution will prompt people to buy health insurance. The longer the polluted weather lasts, the more obvious the effect will be. In addition to deciding whether to buy insurance, the insured will choose different insurance products. These different products correspond to different insurance periods, premiums and basic insurance amounts.

This part mainly studies what factors influence the choice of insurance period, premium rate and insurance amount of the insured. Since the decision-making of these three factors is not independent of each other, the three equations are estimated jointly to improve the estimation efficiency. The seemingly unrelated regression equation used in this part is as follows:

$$\begin{cases} priem\_time = AQI + personal + week + timefix + cityfix + cross + weather \\ priem = AQI + personal + week + timefix + cityfix + cross + weather \\ policy\_count = AQI + personal + week + timefix + cityfix + cross + weather \end{cases}$$

The variables used for regression are consistent with the previous ones. Table 10 (1) - (3) is a regression model. The model includes personal information, time-fixed effect, city-fixed effect, urban meteorological data, the number of days in which AQI variables exceed the threshold of 150 in the past 30 days, and the cross-term between personal information and AQI. The higher the pollution level is in the past 30 days, the more people are willing to pay for it. The regression coefficient is 0. 04 and significant. The corresponding purchase amount is lower and the insurance period is shorter. When exposed to polluted air, people are more worried about the risk of illness, and their subjective risk perception increases, so they are willing to pay higher premiums for the risk.

The regression coefficients of whether or not the insurance purchase claims within two years are significantly positive for both the insurance amount and the insurance period, indicating that the high-risk individuals purchase more insurance and tend to have longer insurance coverage. The regression coefficient of the premium is - 0. 01 and significant, indicating that high-risk individuals will choose products with lower premiums. From the specific product decision-making, it can be seen that the policy-holder uses his own advantages to make a favorable choice for himself when he purchases insurance. In other words, there is information asymmetry in China's commercial health insurance market. High-income people will buy higher premiums, choose longer insurance periods and choose products with lower premiums in a continuously polluted environment. In order to make the layout neat, the regression results of some variables are omitted in Table 10.

Table 10 Seemingly Unrelated Regression of Amount, Premium and Period of Insurance

	(1)	(2)	(3)
VARIABLES	policy_count	priem_time	log_priem

AQI150-30	-0.015***	-0.747***	0.041***
	(0.003)	(0.098)	(0.003)
cros_income	0.005***	0.174***	-0.007***
	(0.000)	(0.011)	(0.000)
cros_gender	-0.002***	-0.074***	0.000
	(0.000)	(0.013)	(0.000)
cros_age	-0.000	0.026***	-0.001***
	(0.000)	(0.001)	(0.000)
cros_claim	0.006**	0.429***	-0.012***
	(0.003)	(0.095)	(0.003)
claimyear2	0.140***	4.348***	-0.137***
	(0.019)	(0.569)	(0.019)
insure_income	0.314***	-2.260***	0.045***
	(0.002)	(0.059)	(0.002)
insure_gender	-0.078***	1.438***	-0.126***
	(0.003)	(0.077)	(0.003)
insure_age	-0.022***	-0.467***	-0.045***
	(0.000)	(0.004)	(0.000)
mean_ temp	0.001***	-0.098***	-0.010***
	(0.000)	(0.005)	(0.000)
mean_humidi	0.062***	-2.057***	0.030***
	(0.010)	(0.302)	(0.010)
mean_snow	0.073***	-1.272***	-0.139***
	(0.005)	(0.162)	(0.006)
mean_rain	0.001***	-0.013***	-0.000
	(0.000)	(0.004)	(0.000)
mean_sun	0.003***	-0.057***	-0.000
	(0.000)	(0.011)	(0.000)
Constant	1.141***	87.960***	5.818***
	(0.028)	(0.832)	(0.028)
timefix	yes	yes	yes
stratified by city	yes	Yes	yes
Observations	629,582	629,582	629,582
R-squared	0.290	0.103	0.263

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Through the regression results of the cross-term, we can discuss the impact of air pollution on the policy-holders with different characteristics, such as insurance amount, premium rate and decision-making during insurance period. The risk level of the insured is posteriori measured by whether the insured claims two years after the purchase of insurance. High-risk groups are also more sensitive to air pollution, choosing higher premiums, longer insurance periods and lower premiums. Continuous high-pollution weather will increase the breadth and depth of insurance coverage. In the selection of specific terms, the high-risk group has obvious adverse selection, and the high-risk group is more sensitive to air pollution. It can be seen that in China's commercial health insurance market, there are adverse choices in terms of specific terms.

Table 10 shows the AQI variable of air pollution, which is "the number of days with AQI higher than 150 in the past 30 days". The author uses different thresholds and different time lengths to regression. Because of the limited layout, only the above regression results are given in this paper.

All the regressions above are from the individual point of view, to study the individual insurance purchase decision-making, specific insurance product purchase decision-making factors. The impact of AQI and personal risk on these decisions was discussed, and the sensitivity of various populations to air pollution was studied. Next, we study the relationship between the purchase of urban insurance policies and the air pollution index AQI from the urban level, instead of focusing on the impact of individual characteristics.

4.3.5 The Regression of Urban Policy Purchase

This part summarizes the policy data according to the city, and obtains the total amount of policy purchases in the city every day. Because the development level of insurance industry in each city is different, the baseline level of insurance policy purchase is different. Taking the total amount of insurance purchase in a city as a reference standard, the ratio of the total amount of insurance purchase per day to the total amount of insurance purchase in that year is calculated. The ratio is used as the explanatory variable to eliminate the differences in the development level of insurance industry in different cities. The specific regression formulas are as follows:

$$totalbuy_{jt} = AQI_{jt} + city_{jt} + year_{jt} + week_{jt} + \varepsilon$$

$$totalbuy_{jt} = AQI_{jt} + city_{jt} + year_{jt} + week_{jt}$$

totalbuy responds to the proportion of new policies purchased every day. *City* is the fixed effect of the city, and *year* is the fixed effect of time, also controls the variables of the *week*. *AQI* is the main explanatory variable. The specific form is consistent with the above. The number of days in which AQI exceeded the threshold in the past X days. Subscript J denotes the jth city and t denotes the date. The regression results are shown in Table 11.

Table 11 Urban Policy Purchase and AQI

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	totalbuy	totalbuy	totalbuy	totalbuy	totalbuy	totalbuy
AQI100_SUM10	-0.001					
	(0.001)					
AQI100_SUM20		-0.001				
		(0.001)				
AQI100_SUM30			-0.000			
			(0.000)			
AQI150_SUM10				0.003**		
				(0.001)		
AQI150_SUM20					0.001*	
					(0.001)	
AQI150_SUM30						0.001
						(0.001)
mean_temp	0.008***	0.007***	0.007***	0.008***	0.008***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
mean_humidi	-0.013	-0.007	-0.013	-0.010	-0.003	-0.010
	(0.017)	(0.018)	(0.018)	(0.017)	(0.017)	(0.018)
mean_snow	0.083***	0.080***	0.076***	0.083***	0.079***	0.075***
	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)
mean_rain	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
mean_sun	-0.001	-0.001	-0.001	-0.001	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.117***	0.118***	0.109***	0.103***	0.105***	0.098***
	(0.028)	(0.030)	(0.031)	(0.028)	(0.029)	(0.030)
timefix	yes	yes	yes	yes	yes	yes
stratified by city	yes	yes	yes	yes	yes	yes
Observations	39,663	38,327	36,999	39,663	38,327	36,999

R-squared	0.074	0.072	0.071	0.074	0.072	0.071

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 11 discusses the impact of air pollution on insurance purchases at the urban level. Model (1) - Model (3) uses the past 10, 20 and 30 days of AQI over 100 days as the main explanatory variable. The regression results are negative. Continuous air pollution does not lead to an increase in insurance purchases. The threshold value of model (4) - model (6) is 150, and the regression coefficients of new insurance policies to air pollution level are positive. When the threshold is 100, the regression coefficient is negative.

As a whole, air pollution will increase the growth of new insurance policies. When the air pollution situation reaches moderate pollution or above, the result is stable. Similar conclusions have been drawn from existing studies on health insurance demand and pollution (Chang et al., 2018). However, the regression results of city-level data are related to the threshold value of AQI variable, which will increase the demand for insurance purchase when AQI exceeds 150. So Chang et al. also had significant positive regression results on the same day. When the threshold value is 100, the regression result is negative. Chang et al. did not mention the case of different thresholds. According to the results of city-level regression in this paper, the cause of negative significance at the threshold of 100 is not clear yet.

In summary, from a personal level, persistent air pollution will lead to an increase in policy purchases. The increase in demand for insurance purchase can be observed after sustained mild pollution, and similar results can be observed for moderate and severe pollution. From the urban level, the increase of insurance purchase demand is more related to the pollution level of the city, but not to the days of continuous pollution. The regression results of cohort study are more in line with common sense, and the model takes into account the relevant variables of personal characteristics. The results obtained by using this model to study insurance demand are more credible.

4.4 Conclusion

This paper uses the health insurance data of a large life insurance company to study the influencing factors of insurance purchase decision. The conclusion of this paper is at least valid for the policy holder of the company. As the company's market share is about 30%, it has the highest market share in China's life insurance industry. Using the company's insurance purchase data for analysis, it is more representative, and the possibility of sample selection bias is small. From the insurance policy data, the information of the insured's income, age, gender, occupation and claim status can be obtained, and the AQI and meteorological data of the insured's city can be matched. The COX proportional hazard model is used to study the influence of the factors on insurance purchase decision-making from the individual insurance policy level. The increment of days with air pollution in the past period of time, will rise the willing of buying health insurance. Empirical test of prospect theory shows that subjective risk perception is not necessarily identical with objective risk. Subjective risk perception lead to people's anxiety about illness. So people may choose to buy health insurance to provide security and alleviate anxiety.

Furthermore, the choice of insurance interval, insurance amount and premium rate is studied. The regression results of the two-year claim and the two-year claim * AQI cross-item are significantly positive for both the insurance amount and the insurance period, and negative for the rate. It can be seen that high-risk individuals tend to buy insurance with higher amount and longer period, in order to get more sound protection. The results confirm the existence of adverse selection in China's health insurance market. High-risk individuals are more sensitive to AQI pollution. They buy commercial health insurance because of persistent air pollution and choose higher insurance to ensure adequate insurance coverage. Due to the limitation of data availability, this paper only controls several personal characteristic variables, and the observation time of insurance purchase is relatively short. There may be some personal habits and other characteristics that affect the regression results. Next, natural experiments can be used to obtain more credible results.

Cox proportional risk is used to conduct a cohort study. It is found that long-term air pollution will increase insurance purchase. Air pollution has an impact on people's health. The insurance industry takes this opportunity to broaden its coverage by meeting people's health security needs. On the basis of providing insurance consumers with protection, we should also pay attention to the fact that the incidence is increasing year by year. Insurer should be more cautious in pricing and take appropriate measures to prevent adverse selection.

Chapter 5: Buy, Hold or Sell? Air Quality and Financial Analyst Reports

Abstract:

Recent literature has suggested a negative impact of poor air quality on several cognitively related tasks. We test the influence of air pollution in the domain of financial reports, in which analysts make forecasts and recommendations on equity investments. Financial analysts are significantly affected by poor air quality in the city of their employment along several dimensions. Poor air quality leads to significantly higher forecasting errors, meaning less precise forecasts by the analysts. Although forecasts are not significantly inflated in a positive direction, except on the very margin of optimism. In addition, poor air quality leads analysts to make mistakes, including incorrect references, statistics and typos in their reports. In summary, we find evidence that analysts are less precise in their forecasts and in their work more generally.

Keywords: financial analysts, air quality, pollution, forecasting, mistakes

JEL codes: G20, G02, J24, D01

5.1 Introduction

As the world increasingly relies on developing economies to drive economic growth, one of the key questions is how environmental conditions affect decisions and the functioning of markets. Carrying the burden of heavy manufacturing activity as well as using older technologies for modern necessities such as heating systems, many of the worst air quality cities in the world are located in developing economies. What are the consequences of air pollution for advisors and decision-makers in these emerging financial markets?

An increasing discussion has circulated around the potential negative effects of air pollution on cognitive ability, risk attitudes, worker productivity, and other influential facets of marketplaces. While the underlying effects may originate in the psychological and even biological influence on decision-making and limits on human abilities under environmental constraints, the potential influence of these effects are far-reaching, to the level of markets, the macroeconomy and beyond. Financial markets provide one fruitful channel for understanding the effects of environmental pollution on agents in the marketplace.

In terms of the relationship between air pollution and cognitive ability, Zhang, Chen and Zhang (2018) find that cumulative exposure to air pollution reduced test scores in both verbal and mathematical domains. In a laboratory setting, Chew, Huang and Li (2017) find that pollution increases risk aversion in the gain domain, while reducing risk aversion in the loss domain. X, Lien and Yuan (2019) find that concurrent pollution levels increase lottery gambling sales, with greater effects in poorer provinces. These studies provide evidence that air pollution levels alter the decision-making abilities and patterns of individuals. More generally, Zhang, Zhang and Chen (2017) found a negative relationship between air pollution and mental health statuses, including subjective well-being and depression symptoms.

A related line of literature studies the effects of pollution on workers in labor markets. In the physical labor domain, studies have found productivity reductions associated with increased particulate air pollution (Chang, Graff Zivin, Gross and Neidell, 2016; Lichter, Pestel and Sommer, 2017; He, Liu and Salvo, 2019) and increased ozone (Graff Zivin and Neidell, 2012), in professions based on manual or physical labor. Evidence for the effect of pollution on office work is sparser. In an office experiment, Lagercrantz et al. (2000) find that air pollution from an old carpet increased office workers' reporting of cognitive difficulty and mathematical error, and decreased typing productivity. In addition, Chang, Graff Zivin and Neidell (forthcoming) find significant negative productivity effects of pollution among call center workers.

Our study is positioned to contribute to the knowledge in the domains of belief formation and biases, willingness to recommend to others, and job performance. Firstly, in terms of the effect of pollution on decision-making fundamentals, we observe how analysts' forecasts for stocks are influenced by the cumulative pollution exposure. There are two main potential effects on the forecasts, the precision of the forecast, and whether there is an optimistic or pessimistic bias. The precision reflects how accurate an analyst is in performing his or her forecasting work, while the bias reveals whether pollution exposure has influenced the analyst's attitude towards the future. Secondly, using text analysis and error detecting software, we test how carefully the analyst has performed in terms of putting together their recommendation report. Mistakes in the report itself are indicative of either reduced cognitive ability or carefulness in the job performance. Our study is one of few to our knowledge, to study the negative effect of pollution on the actual quality of workers' output beyond more general productivity measures.

We find that moderately poor local air quality (in the AQI 100 to 150 and above range) within the most recent 10 to 60 days has a negative effect on analysts' forecasting performance. Absolute forecasting errors are higher when the report is written by an analyst in a recently polluted city, even when accounting for forecasting errors for the market as a whole. While forecasting errors are higher under polluted conditions, pollution does not appear to substantially change the level of overall pessimism or optimism level of an analyst. However, polluted conditions are more likely to lead to an optimistic assessment than a pessimistic one by the analyst.

In other words, pollution makes analysts simultaneously noisier in their forecasting, while being more positive or confident in their recommendation. These two seemingly contradictory facts are also consistent with a separate finding in our results, which is that analysts also make more mistakes under recent accumulated exposure to pollution. We analyze each financial report using a software designed to detect inconsistencies and errors in the report, independently of the forecast and recommendation made. When an analyst was subjected to more days of moderately poor air quality, there are more mistakes in their written report. This serves as a cross-check to confirm that analysts' job performances are indeed negatively affected by recent local pollution conditions.

The remainder of the paper is organized as follows: Section 2 provides a detailed description of the various data sources and variables constructed; Section 3 provides the main results of the paper; Section 4 discusses and concludes.

5.2 Description of Data

To study the effect of China's air pollution on the work of financial analysts, we collect the financial reports on A-share stocks, and the corresponding analyst's firm and A-share company information. The data is combined with the air quality information in the city where the analyst works. This enables us to discern how the recent history of local air quality affects the analyst's forecast on returns on equity, their optimism or pessimism about a stock, their level of recommendation for a stock, and other detailed features of the financial report.

Analyst report data as well as data on the analyst and their employer are collected from the GTA China Stock Market & Accounting Research (CSMAR) Database.[1] Historical daily air pollution data is collected at the city level and matched to the location that the analyst is primarily based in. Daily trading data and performance measures for the stock market are publicly available.

The data on mistakes in the analyst reports is from GoGoal, a company which focuses on analyst reports.[2] Based on private communication, the company applied a textual analysis method first and then checked manually to find the mistakes in the reports. Since only errors which could be detected on the basis of specific rules are found, there may exist non-detected errors.

Table 1: Air Quality Index Levels and Grades

AQI value	Category	Warnings
0-50	1 excellent	Outdoor activity recommended
50-100	2 good	Outdoor activity as normal
101-150	3 lightly polluted	Sensitive individuals avoid outdoors
151-200	4 moderate	Sensitive individuals will be affected
201-300	5 heavy	Everyone should avoid being outdoors
>300	6 very heavy	Do not go outside

Table 1 shows the official Air Quality Index (AQI), categorizations and descriptions. Table 2 provides summary statistics for the variables related to analyst performance. The

1 CSMAR (http://www.gtarsc.com/)

2 (http://www.go-goal.com/)

database consists of over 1000 analysts (906 analyst groups and 481 first author analysts). Table 3 shows the summary statistics for variables related to city pollution.

Table 2: Summary Statistics, Analyst Variables

Variable	Obs	Mean	Std. Dev.	Min	Max
error	51,596	0.038	0.192	0	1
Forecast_error	47,830	0.759	1.630	0.002	11.849
reForecast_error	47,601	0.003	0.016	-0.048	0.077
Forecast_opti	47,830	0.006	0.019	-0.058	0.083
reForecast_opti	47,601	-0.003	0.019	-0.081	0.061
authnumber	51,596	2.190	0.950	1	6
freq	51,596	4.825	3.802	1	19
gender	43,447	0.773	0.419	0	1
exper	44,657	4.179	0.519	1.946	5.017
cover2	51,596	3.587	0.751	1.609	5.620
spc2	51,596	0.374	0.286	0.013	1
previous	51,596	3.766	1.181	0.693	5.130
degree	51,596	0.739	0.439	0	1
rank	51,596	0.035	0.185	0	1
growth2	47,135	1.242	0.498	0.592	4.596
roa	49,209	0.057	0.048	-0.076	0.214
sd_roa	40,637	0.074	1.188	0.001	48.666
state	49,151	0.099	0.298	0	1
lev	49,209	0.442	0.212	0.069	0.935
top1	49,151	34.827	14.801	8.540	73.820
size	49,151	13.457	1.687	10.884	19.723
cage	49,251	2.148	0.692	0	3.178
big4com	49,151	0.122	0.327	0	1
follow	48,901	2.508	0.774	0	3.738
horizon	49,251	4.837	0.766	1.609	5.684
uw	49,251	0.018	0.133	0	1

Table 4 displays the variables of the analysis. The main dependent variable we consider (Forecast Error) is the forecast error for the earnings per share, which is a measure of the analyst's precision of forecasting. To control for the overall difficulty level in making predictions based on market conditions, we also consider the relative forecast precision as compared to the market forecast error (RForecast Error).

Table 3: Summary Statistics, Pollution

Variable	Obs	Mean	Std. Dev.	Min	Max
aqimean30	51,596	87.891	23.264	41.6	155.233
aqimean20	51,596	87.695	25.159	38.8	167.5
aqimean10	51,596	87.478	28.452	36.7	186.4
aqimean5	51,596	87.585	32.595	33.6	198.8
aqi100_sum30	49,491	8.370	5.698	0	27
aqi100_sum20	50,034	5.588	4.153	0	19
aqi100_sum10	50,731	2.774	2.453	0	10
aqi100_sum5	51,229	1.417	1.472	0	5
aqi150_sum30	49,491	2.883	3.617	0	19
aqi150_sum20	50,034	1.902	2.588	0	14
aqi150_sum10	50,731	0.937	1.527	0	9
aqi150_sum5	51,229	0.477	0.930	0	5
aqi200_sum30	49,491	0.714	1.954	0	16
aqi200_sum20	50,034	0.502	1.477	0	11
aqi200_sum10	50,731	0.250	0.846	0	8
aqi200_sum5	51,229	0.125	0.508	0	5
humidimean_5	51,596	0.709	0.129	0.307	0.906
tempmean_5	51,596	21.020	7.081	-2.807	32.320
snowmean_5	51,596	0.022	0.143	0	1.000
sunmean_5	51,596	5.195	2.792	0	10.833
rainmean_5	51,596	4.454	6.220	0	29.750

In addition to the forecast itself, we also examine analyst’s official recommendation levels for the stock. Raw recommendations (recomm) have 5 categories, ranging from ‘buy’ to ‘sell’. We also consider a binary form of the recommendation variable (recomm 2) which is 1 for buy or outperform, and 0 for hold, underperform or sell. We account for the overall recommendation level (Relative Recomme) in the market as well.

The next set of dependent variables considers the potential direction of the forecasting bias (Forecast Opti) and relative to the overall market condition (Relative Optimism), by measuring the price normalized direction of forecasting error rather than simply the absolute value.

Finally, we examine the analyst's report for mistakes in the write-up, including typos, incorrectly quoted data, and so on, using the data from GoGoal, where a binary variable (error) indicates the presence of at least one such mistake. This variable is a measure of non-forecasting accuracy in the analyst's work and performance.

Our main explanatory variable of interest is the count of days within a given number of days in which the AQI exceeded a particular threshold in the analyst's place of employment. In the baseline estimate, an AQI of over 150 during the past 30 days is considered. Other thresholds and accumulation periods are also considered.

Table 4: Variable Descriptions

Variable Content	*Variable Name*	*Variable Description*
Dependent Variables		
Forecast Error for EPS	Forecast_error	|Forecasted eps – actual eps|/ |actual eps| Absolute value of the difference between the analyst's forecasted EPS (earnings per share) and the actual EPS, divided by the absolute value of the actual EPS
Relative Forecast Precision	RForecast_error	The difference between the absolute value of the analyst's forecast error (for EPS) and the average absolute value of market's forecast error, divided by the closing price of the stock for the trading day before the report is released.
Optimism Bias	Forecast_opti	(Forecasted eps – actual eps)/prior day's price The difference between the analyst's forecasted EPS (earnings per share) and the actual EPS, divided by the closing price of the stock for the trading day before the report is released
Relative Optimism Bias	relativeoptimism	(Forecast opti – Forecast opti market)/prior day's price The difference between the analyst's optimism bias and the average optimism bias of the market, divided by the closing price of the stock for the trading day before the report is released
Report Error	error	1 for existence of errors in the report and 0 for no errors
Explanatory		

Variables		
Air Quality Index (AQI)	aqiX_sumY	Number of days for which AQI is greater than X for the city where the analyst is located among the previous Y days before the report is released
Control Variables		
Forecast Horizon	horizon	the logarithm of (1 plus the difference between the last day the forecast covers and the day the report is released)
About the analyst		
Forecast Frequency	freq	Number of reports on a given company (stock) by the analyst, during the year where reports are released
Number of Companies Analyzed	cover	Number of companies the analyst has analyzed
gender	gender	Gender of the analyst
Number of author	authnumber	Number of the author
degree	degree	1 for master and PhD, 0 for undergraduate and high school
Time gap	previous	the logarithm of (1 plus the number of days passed between the day the report is released and the day the Annual Report or Semi-annual Report of this company)
Experience	exper	the logarithm of (1 plus the number of seasons passed between the day the report is released and the day the first report by the analyst was released)
Star Analyst Dummy	rank	1 for star analyst, 0 for not a star analyst
Specialty	spc	Number of companies analyzed by the analyst in a given industry, divided by the number of companies analyzed by the analyst in total
About the brokerage the analyst works for		
Lead Underwriter Dummy	uw	Whether or not the brokerage which the analyst works for is the lead underwriter of the company analyzed by the analyst
About the city the analyst works in		
temperature	tempmean	

humidity	humidimean	
rain	rainmean	
snow	snowmean	
Sun	sunmean	

About the company

Rating on Information Disclosure	rating	4 for A(highest), 3 for B, 2 for C, and 1 for D (lowest)
Growth Rate	growth	This year's revenue divided by previous year's revenue
Rate of Return on Common Stockholders' Equity (ROA)	roa	Income divided by Common Stockholders' Equity
SOE Dummy	state	1 if the largest shareholder is SOE, 0 if not
Debt ratio	lev	Total Debts/Total Assets
Percentage of Shares	top1	Percentage of shares held by the largest shareholder
Size of the Company	size	the logarithm of the total asset of the company
Age of the Company	cage	the logarithm of 1 plus the difference between the day the report is released and the day the company goes for IPO
Big 4 Dummy	big4com	1 if the auditing report is released by the big 4 accounting firms, 0 if not
Analyst's Attention received	follow	the logarithm of the number of analysts that release reports
Standard Deviation of the ROA (Return of Assets)	sd_roa	Adjusted ROA=current year's ROA of the company – current year's ROA of the industry that the company belongs to. The Standard Deviation is calculated using the ROA data in 3 years before the year when the report is realeased.

Control variables include forecast horizon, as well as relevant descriptive variables about the analyst that wrote the report, the firm the analyst works for, and the company the analyst is writing about. The details of these and all variables in the analysis are provided in Table 4.

According to the AQI index classification standard, the air pollution level is divided into six levels, and the average values of net income prediction error, optimism bias and investment rating under each level of pollution level are calculated respectively. The data are used to plot, as shown in Figure 1. By analyzing the cities where the analysts are located, it is found that the securities analysts are concentrated in the first-tier cities of China: Beijing, Shanghai, Guangzhou and Shenzhen. The number of research reports issued by these four cities accounts for 95.67% of the total number of reports. The AQI monthly mean values of these four cities are calculated, and the AQI monthly trend maps of the four cities are drawn, respectively, as shown in Figure 2.

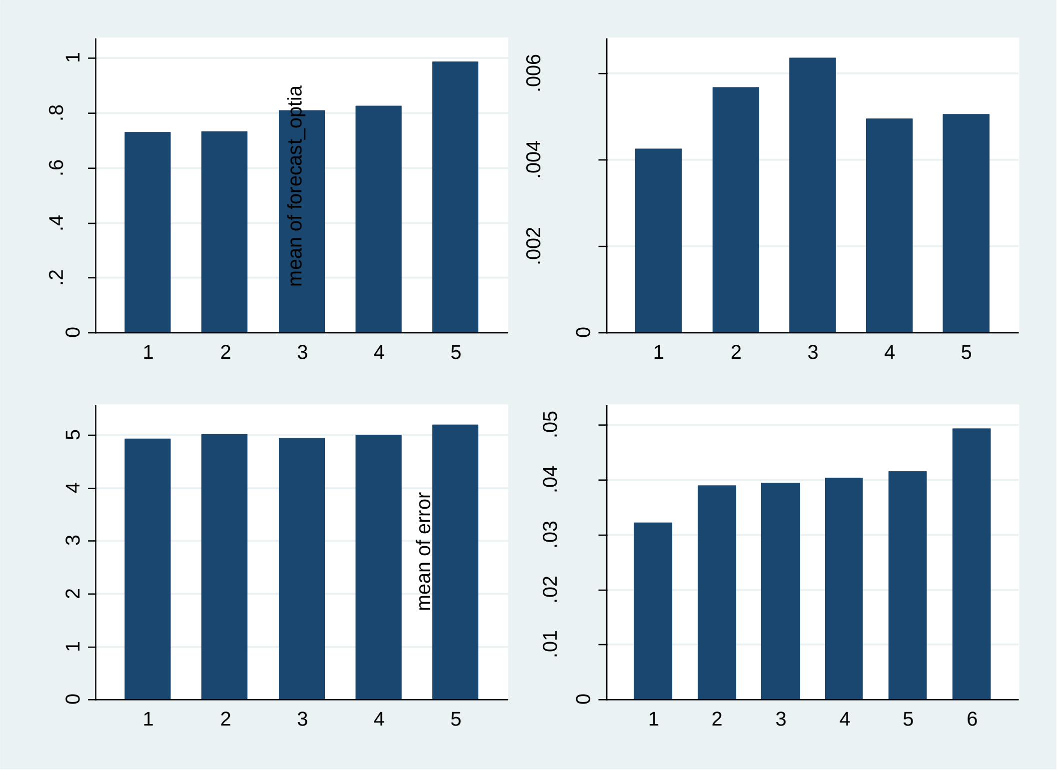

Fig. 1 Prediction accuracy, optimism, recommendation and error rate at different air pollution levels

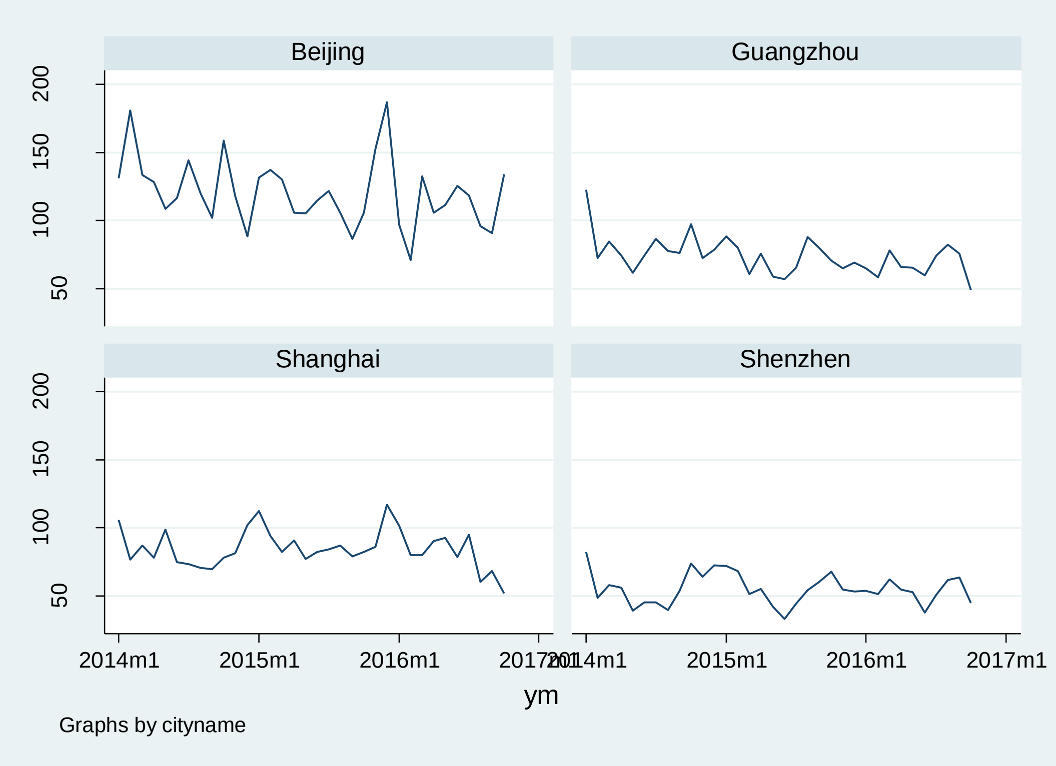

Figure 2 Air quality trend charts of Beijing, Shanghai, Guangzhou and Shenzhen

Table 5 provides correlation coefficients of some of the variables in the analysis. Return on equity and the standard deviation of return on assets, are positively correlated with analysts' recommendations, while these two returns variables are negatively correlated with one another. The number of analysts releasing reports that day is positively correlated with an analyst's recommendation as well as the return on equity.

Table 5: Correlations

	recomm	roe	sd_roa	follow
recomm	1			
roe	0.1826*	1		
sd_roa	0.0097*	-0.0159*	1	
follow	0.1437*	0.5053*	-0.0270*	1

**** p<0.01, ** p<0.05, * p<0.1*

5.3 Results

5.3.1 Prediction accuracy regression results

Table 6 shows the baseline results, with forecasting error as the dependent variable. The number of days with AQI mean within the last 30 days is a significant predictor of a higher forecast error in the analyst's report. The significance of the result is robust to control variables for the company in question, the analyst, and the reporting company, while the control variables hold intuitive direction of effects.

Column (1) contains the basic result with only mean of AQI as the control variable, along with year and industry fixed effects. In addition to pollution increasing the forecast error. Column (2) additionally controls for the city's meteorological data (temperature, snow, sunshine, rainfall, humidity), in which the regression coefficient of temperature is significantly positive, and the higher the temperature, the greater the prediction error. The regression coefficient of sunshine duration is significantly negative. The longer the sunshine, the more photosensitive the amygdala in the brain, the happier the mood and the better the work.

After controlling the meteorological data, column (3) additionally includes the analyst's personal characteristic variables. The more frequently an analyst predicts, the more reports he publishes about a company, the more time he spends on the company, the more information he collects and the more sensitive he is to the company's relevant information. So the analysis is clearer to the company's situation, more accurate to forecast, and the regression coefficient is significantly negative. The regression coefficient of cover is significantly negative when analysts study the variables of the number of companies, which indicates that the more companies the analysts follow, the more accurate the forecast is. Firstly, the stronger the analyst's ability, the more companies will be allocated, and the more accurate the analyst's forecast, which is consistent with the previous analysis.

Table 6: Main Results, Dependent Variable: Forecast Error

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	forecast_error	forecast_error	forecast_error	forecast_error	forecast_error	forecast_error
aqimean30	0.020***	0.016***	0.019***	0.020***	0.021***	0.021***
	(0.003)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)

tempmean_30	0.054*	0.069**	0.099**	0.109***	0.109***
	(0.032)	(0.035)	(0.039)	(0.040)	(0.040)
snowmean_30	-1.234	-1.067	-1.018	-1.105	-1.118
	(0.891)	(0.976)	(1.040)	(1.038)	(1.036)
sunmean_30	-0.292***	-0.245***	-0.214**	-0.215**	-0.216**
	(0.076)	(0.082)	(0.093)	(0.093)	(0.093)
rainmean_30	0.025	0.032	0.038	0.039	0.039
	(0.031)	(0.033)	(0.038)	(0.038)	(0.038)
humidimean_30	-6.905***	-5.862***	-4.658**	-4.627**	-4.635**
	(1.542)	(1.637)	(1.816)	(1.817)	(1.816)
authnumber		0.007	-0.244	-0.241	-0.249
		(0.151)	(0.170)	(0.170)	(0.170)
freq		-0.105***	0.077***	0.078***	0.078***
		(0.021)	(0.023)	(0.023)	(0.023)
gender		0.922***	0.622***	0.616***	0.615***
		(0.191)	(0.217)	(0.217)	(0.217)
exper		0.151	0.240	0.246	0.243
		(0.155)	(0.174)	(0.174)	(0.174)
cover2		-0.057	-0.434***	-0.436***	-0.433***
		(0.129)	(0.147)	(0.147)	(0.147)
spc2		-2.411***	-1.131**	-1.134**	-1.148**
		(0.426)	(0.484)	(0.484)	(0.484)
previous		-0.064	-0.225**	-0.233**	-0.234**
		(0.086)	(0.093)	(0.094)	(0.094)
degree		0.232	0.124	0.128	0.125
		(0.266)	(0.279)	(0.279)	(0.279)
rank		1.472***	1.183**	1.183**	1.174**
		(0.546)	(0.525)	(0.525)	(0.525)
growth2			9.155***	9.157***	9.157***
			(0.440)	(0.440)	(0.440)
roa			-0.633***	-0.633***	-0.633***
			(0.036)	(0.036)	(0.036)
sd_roa			-0.210**	-0.212**	-0.211**
			(0.100)	(0.100)	(0.100)
state			-1.332***	-1.331***	-1.327***
			(0.308)	(0.308)	(0.309)
lev			5.324***	5.339***	5.351***
			(0.901)	(0.901)	(0.901)
top1			-0.052***	-0.052***	-0.052***
			(0.007)	(0.007)	(0.007)
size			-1.490***	-1.492***	-1.490***

				(0.128)	(0.128)	(0.128)
cage				1.346***	1.350***	1.355***
				(0.238)	(0.238)	(0.238)
big4com				0.590**	0.596**	0.591**
				(0.265)	(0.265)	(0.265)
follow				-1.841***	-1.842***	-1.842***
				(0.187)	(0.187)	(0.187)
horizon					-0.928*	-0.929*
					(0.523)	(0.523)
uw						0.699
						(0.680)
Constant	-1.284	4.759*	3.862	16.620***	21.784***	21.786***
	(1.911)	(2.502)	(2.825)	(3.185)	(4.238)	(4.237)
timefix	yes	yes	yes	yes	yes	yes
indusfix	yes	yes	yes	yes	yes	yes
Observations	47,830	47,830	40,221	31,713	31,713	31,713
R-squared	0.061	0.061	0.063	0.176	0.176	0.176

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Secondly, the more companies that analysts track, the more time they may spend thinking about relevant research, thus more exercise in their predictive ability. Finally, tracking different companies may have an enlightening effect on each other, but increase the accuracy of prediction. The empirical regression coefficient of analysts is positive, but not significant. The regression coefficient of SPC is - 2.41, which is significant. The regression coefficients of star analysts are significantly positive, while the regression coefficients of analysts' educational background and the number of people writing reports are not significant. The gender of the analyst was significantly positive, and the predictive accuracy of men was lower than that of women.

Column (4) adds the control variables of listed companies, and the regression coefficient of growth of companies is significantly positive. The fast-growing enterprises are more uncertain, more difficult to predict and less accurate. The regression coefficient of ROA is negative and significant. Lev, regression coefficient is 5.32 and significant. The regression coefficient of state, the property of ownership, is also significantly negative. The higher the analyst's attention, the more information about the company in the market, and the lower the uncertainty, the more accurate the forecast is. The regression coefficient is significantly negative, which indicates that the high attention of analysts leads to high prediction accuracy. The regression coefficient of enterprise size is significantly negative, and the above regression results are consistent with the prediction results. The regression coefficient of Top1 is negative and significant, which is also consistent with the forecast. The listed year of the company is significantly positive, the listed time is long, and the prediction accuracy is low.

Column (5) and column (6) respectively add the relevant variables of horizon interval and the securities firm where the analyst is located. Whether the analyst's securities firm is listed underwriter UW of listed companies or not, the regression coefficient of this variable is positive, but not significant. If the analyst's securities firm is a listed underwriter of the listed company, there is an interest correlation between the two, which leads to distorted forecasts, so the report of the underwriter's analyst is less accurate.

Because analysts' time to complete reports is not necessarily the time spent on more invested reports may take a long time to collect and collate information, process analysis and draw conclusions to generate research reports. An irresponsible analyst may finish a research report in a week. The time windows of 5, 10, 20 and 30 days before the release of the report were taken to obtain the different variables of air pollution mean. In addition, with different thresholds of 100, 150 and 200, different air pollution variables are also obtained. The regression results are shown in Table 7.

Table 7: Air Quality Variations, Dependent Variable: Forecast Error

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	forecast_error	forecast_error	forecast_error	forecast_error	forecast_error	forecast_error	forecast_error
aqimean30	0.021***						
	(0.006)						
aqi100_sum30		0.069***					
		(0.024)					
aqi200_sum30			0.142*				
			(0.078)				
aqi150_sum30				0.138***			
				(0.040)			
aqi150_sum20					0.097*		
					(0.050)		
aqi150_sum10						0.094	
						(0.073)	

aqi150_sum5							0.151
							(0.115)
tempmean_30	0.109***	0.073*	0.065	0.071*	0.037	0.021	0.028
	(0.040)	(0.038)	(0.040)	(0.037)	(0.036)	(0.032)	(0.029)
snowmean_30	-1.118	-1.283	-2.625**	-2.250*	-1.928*	-1.731**	-1.446*
	(1.036)	(1.165)	(1.205)	(1.153)	(0.992)	(0.826)	(0.782)
sunmean_30	-0.216**	-0.216**	-0.269***	-0.196**	-0.238***	-0.160**	-0.170***
	(0.093)	(0.094)	(0.093)	(0.095)	(0.082)	(0.069)	(0.055)
rainmean_30	0.039	0.019	0.002	0.017	0.033	0.005	0.010
	(0.038)	(0.038)	(0.037)	(0.038)	(0.034)	(0.026)	(0.019)
humidimean_30	-4.635**	-4.039**	-5.922***	-3.044	-5.177***	-4.674***	-4.694***
	(1.816)	(1.957)	(1.825)	(2.014)	(1.797)	(1.467)	(1.241)
authnumber	-0.249	-0.243	-0.250	-0.249	-0.261	-0.256	-0.263
	(0.170)	(0.175)	(0.175)	(0.175)	(0.174)	(0.172)	(0.170)
freq	0.078***	0.072***	0.071***	0.074***	0.074***	0.077***	0.076***
	(0.023)	(0.024)	(0.024)	(0.024)	(0.024)	(0.023)	(0.023)
gender	0.615***	0.685***	0.716***	0.712***	0.701***	0.665***	0.629***
	(0.217)	(0.219)	(0.220)	(0.219)	(0.219)	(0.219)	(0.219)
exper	0.243	0.274	0.282	0.280	0.269	0.217	0.239
	(0.174)	(0.177)	(0.177)	(0.177)	(0.176)	(0.176)	(0.175)
cover2	-0.433***	-0.426***	-0.417***	-0.430***	-0.390***	-0.401***	-0.427***
	(0.147)	(0.151)	(0.151)	(0.151)	(0.150)	(0.148)	(0.148)
spc2	-1.148**	-1.177**	-1.208**	-1.173**	-1.205**	-1.282***	-1.310***
	(0.484)	(0.492)	(0.491)	(0.492)	(0.488)	(0.488)	(0.485)
previous	-0.234**	-0.209**	-0.199**	-0.212**	-0.236**	-0.201**	-0.201**
	(0.094)	(0.094)	(0.094)	(0.094)	(0.094)	(0.094)	(0.093)
degree	0.125	0.053	0.048	0.073	0.075	0.074	0.086
	(0.279)	(0.288)	(0.288)	(0.288)	(0.286)	(0.282)	(0.280)
rank	1.174**	1.219**	1.218**	1.220**	1.195**	1.125**	1.170**
	(0.525)	(0.537)	(0.537)	(0.537)	(0.534)	(0.523)	(0.522)
growth2	9.157***	9.183***	9.187***	9.185***	9.210***	9.177***	9.187***
	(0.440)	(0.452)	(0.452)	(0.452)	(0.450)	(0.444)	(0.443)
roa	-0.633***	-0.613***	-0.614***	-0.613***	-0.614***	-0.626***	-0.630***
	(0.036)	(0.037)	(0.037)	(0.037)	(0.037)	(0.036)	(0.036)
sd_roa	-0.211**	-0.219**	-0.221**	-0.225**	-0.219**	-0.212**	-0.211**
	(0.100)	(0.100)	(0.099)	(0.100)	(0.099)	(0.099)	(0.099)
state	-1.327***	-1.140***	-1.142***	-1.140***	-1.171***	-1.229***	-1.275***
	(0.309)	(0.316)	(0.317)	(0.316)	(0.316)	(0.312)	(0.310)
lev	5.351***	5.528***	5.523***	5.501***	5.704***	5.369***	5.376***
	(0.901)	(0.912)	(0.912)	(0.912)	(0.915)	(0.911)	(0.904)
top1	-0.052***	-0.053***	-0.053***	-0.053***	-0.052***	-0.051***	-0.051***

	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
size	-1.490***	-1.480***	-1.483***	-1.477***	-1.477***	-1.496***	-1.497***
	(0.128)	(0.129)	(0.129)	(0.129)	(0.130)	(0.129)	(0.128)
cage	1.355***	1.266***	1.269***	1.273***	1.282***	1.343***	1.341***
	(0.238)	(0.242)	(0.242)	(0.242)	(0.241)	(0.240)	(0.239)
big4com	0.591**	0.635**	0.641**	0.635**	0.606**	0.644**	0.624**
	(0.265)	(0.268)	(0.268)	(0.268)	(0.269)	(0.268)	(0.265)
follow	-1.842***	-1.840***	-1.831***	-1.839***	-1.858***	-1.873***	-1.845***
	(0.187)	(0.190)	(0.190)	(0.190)	(0.189)	(0.188)	(0.187)
horizon	-0.929*	-0.924*	-0.908	-0.992*	-0.835	-0.799	-0.872*
	(0.523)	(0.559)	(0.559)	(0.561)	(0.566)	(0.523)	(0.514)
uw	0.699	0.678	0.668	0.679	0.706	0.649	0.716
	(0.680)	(0.693)	(0.694)	(0.693)	(0.692)	(0.684)	(0.682)
Constant	21.786***	21.687***	24.019***	21.939***	23.507***	23.147***	23.808***
	(4.237)	(4.771)	(4.676)	(4.754)	(4.193)	(4.049)	(3.950)
timefix	yes	yes	yes	yes	yes	yes	yes
indusfix	yes	yes	yes	yes	yes	yes	yes
Observations	31,713	30,416	30,416	30,416	30,728	31,212	31,515
R-squared	0.176	0.177	0.177	0.177	0.177	0.176	0.176

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Column (1) - Column (4) is the air pollution situation in the past 30 days. Threshold and mean variables are used to describe the air pollution level of each city. Four regressions showed that the more serious the air pollution situation in the cities where the major analysts were located 30 days before the report was released, the lower the accuracy of the forecast report. The results show that the regression coefficients are significantly positive when different variables are used to reflect the air pollution situation, which indicates that the conclusion is robust.

Column (4) - Column (7) uses moderately polluted AQI levels as thresholds and takes different time lengths. The regression coefficient of air pollution in the past 20 and 30 days is significantly positive, indicating that air pollution in the past 20 and 30 days will reduce the accuracy of analysts' prediction. However, the air pollution level in the past 5 days and 10 days has no significant impact on the accuracy of the analyst's prediction, and the two regressions are not significant. The analyst first conducts research, interviews, collects information, then analyses information, calculates data, and finally combs and summarizes the conclusions and organizes them into research reports. From the time point of view, analysts are more likely to analyze information and calculate data in 20-30 days, and 1-10 days are more likely to summarize and write reports. Therefore, the accuracy of the analyst's prediction is only related to the air pollution situation in 20 or 30 days, but not to the air pollution situation in 5 or 10 days.

Next, the explanatory variables such as relative accuracy and optimism are used for regression, and table 8is obtained. In order to facilitate the comparison of the regression results of different interpreted variables, the Column (1) in Table 8 is consistent with Table 6, which is the regression of prediction accuracy. The explanatory variable of Column (3) is relative accuracy, and the regression coefficient of air pollution is still significant. Air pollution will reduce the analyst's prediction accuracy, and no matter which variable is used to measure the prediction accuracy results are significant, the conclusion is robust.

The explanatory variable of Column (2) is the optimistic bias of prediction, and Column (3) is the relative optimistic degree. Both of them respond to the optimistic degree of prediction. The regression results of Column (2) are significant. Air pollution will increase the degree of blindness and optimism of prediction. However, the regression results of Column (4) are not significant. Air pollution does not lead to excessive optimistic prediction. Although air pollution will increase the prediction error, the prediction error does not show a specific bias.

Table 8: Dependent Variables: Forecast Error, Optimism

	(1)	(2)	(3)	(4)	(5)
VARIABLES	forecast_error	forecast_opti	rforecast_error60	relativeoptimism60	forecast_error
aqimean30	0.021***	0.187***	0.103*	0.062	0.021***
	(0.006)	(0.064)	(0.060)	(0.073)	(0.007)
tempmean_30	0.109***	0.063	0.756*	-0.469	0.119**
	(0.040)	(0.419)	(0.392)	(0.476)	(0.049)
snowmean_30	-1.118	8.607	-7.387	-4.530	-0.993
	(1.036)	(11.401)	(10.012)	(12.272)	(1.241)
sunmean_30	-0.216**	0.470	-6.263***	-2.705**	-0.169
	(0.093)	(1.056)	(0.961)	(1.182)	(0.112)
rainmean_30	0.039	0.864**	-0.832**	0.298	0.007
	(0.038)	(0.394)	(0.373)	(0.460)	(0.046)
humidimean_30	-4.635**	-12.873	-68.449***	-57.127***	-0.948
	(1.816)	(17.611)	(16.601)	(19.806)	(2.077)

authnumber	-0.249	-1.499	-0.521	0.733	-0.340*
	(0.170)	(1.880)	(1.745)	(2.083)	(0.187)
freq	0.078***	0.973***	2.572***	2.142***	0.014
	(0.023)	(0.273)	(0.231)	(0.291)	(0.028)
gender	0.615***	3.272	5.207**	7.185***	0.752***
	(0.217)	(2.415)	(2.271)	(2.781)	(0.274)
exper	0.243	2.765	0.610	-3.291	-0.077
	(0.174)	(2.011)	(1.781)	(2.258)	(0.208)
cover2	-0.433***	-2.828*	-6.721***	-7.501***	-0.220
	(0.147)	(1.500)	(1.409)	(1.696)	(0.176)
spc2	-1.148**	-13.954***	-20.017***	-25.007***	-2.253***
	(0.484)	(5.109)	(4.660)	(5.651)	(0.598)
previous	-0.234**	-4.044***	-2.006**	-5.218***	-0.292**
	(0.094)	(1.098)	(1.000)	(1.226)	(0.115)
degree	0.125	-0.890	-0.931	6.403*	0.279
	(0.279)	(3.368)	(3.076)	(3.737)	(0.340)
rank	1.174**	22.531***	10.274	12.651*	1.256**
	(0.525)	(6.808)	(6.337)	(7.414)	(0.611)
growth2	9.157***	56.278***	30.285***	54.484***	9.526***
	(0.440)	(2.505)	(2.508)	(2.591)	(0.565)
roa	-0.633***	2.722***	3.335***	4.282***	-0.491***
	(0.036)	(0.306)	(0.282)	(0.334)	(0.043)
sd_roa	-0.211**	-8.985***	6.488***	-10.465***	-0.267**
	(0.100)	(2.490)	(2.410)	(3.356)	(0.114)
state	-1.327***	-21.777***	-16.328***	-19.667***	0.052
	(0.309)	(3.753)	(3.493)	(4.216)	(0.523)
lev	5.351***	180.505***	108.273***	175.856***	4.483***
	(0.901)	(8.548)	(8.065)	(9.224)	(1.185)
top1	-0.052***	-0.593***	-0.082	0.007	-0.080***
	(0.007)	(0.079)	(0.074)	(0.090)	(0.009)
size	-1.490***	1.902	5.743***	-15.930***	-1.783***
	(0.128)	(1.335)	(1.295)	(1.574)	(0.193)
cage	1.355***	15.522***	-3.869*	15.853***	1.934***
	(0.238)	(2.263)	(2.048)	(2.533)	(0.366)
big4com	0.591**	17.097***	-12.337***	24.583***	2.707***
	(0.265)	(4.349)	(3.973)	(4.798)	(0.476)
follow	-1.842***	-20.207***	0.905	0.082	-1.802***
	(0.187)	(1.725)	(1.666)	(1.940)	(0.240)
horizon	-0.929*	18.469***	-15.430***	-2.923	-0.188
	(0.523)	(5.958)	(5.170)	(6.371)	(0.620)
uw	0.699	0.397	11.713	9.295	2.180**

	(0.680)	(7.974)	(7.691)	(8.638)	(0.874)
rating					-2.364***
					(0.225)
Constant	21.786***	-59.129	-306.109***	27.749	29.027***
	(4.237)	(82.370)	(66.627)	(162.906)	(5.167)
timefix	yes	yes	yes	yes	yes
indusfix	yes	yes	yes	yes	yes
Observations	31,713	31,713	31,554	31,554	20,847
R-squared	0.176	0.166	0.097	0.074	0.195

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The relative accuracy in Table 8 is the prediction error subtracted from the industry average prediction error in the past 60 days; the relative optimism is the difference between the analyst's prediction error and the other analysts' prediction error in the past 60 days, divided by the other analysts' prediction error. The author used the industry average of the past 30 days to calculate the relative accuracy and optimism, and the regression results remained unchanged. Because of the limited layout, the results of robust regression are not listed here.

5.3.2 Analysis Report Error Regression Results

The author collects all the analyst forecast reports in recent years as far as possible, and classifies the possible errors in the report roughly, and checks whether there are corresponding errors in the report with code. Next, we will study the air pollution situation in the analyst's city, and the relationship between the various input and calculation errors in the analysis report. The model in Table 9 (1) uses the AQI mean five days before the release of the report as the main explanatory variable to logistic regression for reporting errors. The regression result was 0.003, which was significantly positive. The more serious the air pollution level in the city where the analyst is located, the higher the probability of errors in the analyst's forecast report, such as errors in writing, calculation and expression, etc. The Column (2) adds meteorological data, and the regression results are still significant.

Table 9: Dependent Variables: Mistakes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	error	error	error	error	error	error	error	error	error

aqimean5	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
humidimean_5		-0.438	-0.395	-0.420	-0.529	-0.516	-0.514	-0.500	-0.510
		(0.308)	(0.309)	(0.310)	(0.329)	(0.329)	(0.329)	(0.329)	(0.329)
tempmean_5		-0.018**	-0.016**	-0.017**	-0.021***	-0.022***	-0.022***	-0.021***	-0.022***
		(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
snowmean_5		-0.468**	-0.471**	-0.472**	-0.369	-0.391	-0.401	-0.395	-0.397
		(0.235)	(0.236)	(0.236)	(0.245)	(0.245)	(0.245)	(0.246)	(0.245)
sunmean_5		0.007	0.006	0.006	0.001	0.002	0.001	0.001	0.001
		(0.014)	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
rainmean_5		0.013***	0.012**	0.012**	0.011**	0.011**	0.011**	0.011**	0.011**
		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
authnumber			-0.106***	-0.139***	-0.125***	-0.123***	-0.103***	-0.100***	-0.101***
			(0.024)	(0.024)	(0.026)	(0.026)	(0.027)	(0.027)	(0.028)
freq				0.050***	0.052***	0.053***	0.056***	0.057***	0.057***
				(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
gender					0.294***	0.283***	0.292***	0.291***	0.298***
					(0.067)	(0.067)	(0.067)	(0.067)	(0.067)
exper						-0.127***	-0.098**	-0.099**	-0.096**
						(0.045)	(0.046)	(0.046)	(0.046)
cover2							-0.113***	-0.139***	-0.136***
							(0.038)	(0.040)	(0.040)
spc2								-0.254**	-0.255**
								(0.124)	(0.124)
previous									0.031
									(0.026)
degree									0.083
									(0.081)
rank									-0.036
									(0.152)
Constant	-3.994***	-3.513***	-3.299***	-3.394***	-3.457***	-2.919***	-2.734***	-2.617***	-2.877***
	(0.368)	(0.451)	(0.455)	(0.456)	(0.473)	(0.507)	(0.511)	(0.514)	(0.536)
timefix	yes	yes	yes	yes	yes	yes	yes	yes	yes
indusfix	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	51,441	51,441	51,441	51,441	43,204	43,203	43,203	43,203	43,203

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Column (3) increased the number of report writers, and the regression coefficient was significantly negative. The more people involved in writing reports, the more they will help each other check, thus reducing such errors as clerical errors and inappropriate statements in analytical reports. Column (4) Increased the frequency of forecasting. The higher the frequency of forecasting, the less time and energy spent on each article. Therefore, the probability of errors is higher, and the regression results are significant. In the Column (5), the regression coefficient of analysts' gender is significantly positive, and the probability of male analysts' errors is higher.

Column (6) Increases the analyst's experience. The more experienced the analyst is, the more skilled he is in writing the analysis report. His experience reduced the probability of reporting errors, so the regression coefficient was significantly negative. Similar models (7) and (8) increase the number of companies and analysts' expertise predicted by analysts respectively, and the regression results are significantly negative. The Column (9) adds three variables: analyst degree, star analyst or not, release gap, and the coefficient of air pollution variable is still significant.

Column (1) - Column (4) in Table 10. The main explanatory variables are the mean air pollution of 5, 10, 20 and 30 days before the release of the report. It can be seen that the air pollution level in the past 5 or 10 days will increase the error rate of the analyst's forecast report, while the air pollution level in the past 20 or 30 days will not increase the error rate of the analyst's forecast report. As mentioned in the previous section, analyst reports are mainly written in 1-10 days, and analytical calculations are before that. Therefore, the errors in the analysis report are only affected by air pollution in recent days, because analysts usually write reports in the days prior to the onset of the disease.

Table 10: Dependent Variables: Mistakes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	error	error	error	error	error	error	error
aqimean30	0.002						
	(0.002)						
aqimean20		0.001					
		(0.001)					
aqimean10			0.003***				
			(0.001)				
aqimean5				0.003***			
				(0.001)			
aqi100_sum5					0.040**		
					(0.018)		
aqi150_sum5						0.064**	

						(0.025)	
aqi200_sum5							0.115**
							(0.046)
previous	0.030	0.030	0.031	0.031	0.032	0.030	0.031
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
freq	0.057***	0.057***	0.057***	0.057***	0.056***	0.056***	0.056***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
degree	0.078	0.080	0.084	0.083	0.066	0.071	0.069
	(0.081)	(0.081)	(0.081)	(0.081)	(0.081)	(0.081)	(0.081)
rank	-0.044	-0.040	-0.038	-0.036	-0.022	-0.027	-0.030
	(0.153)	(0.153)	(0.153)	(0.152)	(0.152)	(0.153)	(0.153)
gender	0.299***	0.299***	0.298***	0.298***	0.306***	0.309***	0.312***
	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)
exper	-0.102**	-0.102**	-0.101**	-0.096**	-0.102**	-0.101**	-0.102**
	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
cover2	-0.135***	-0.134***	-0.135***	-0.136***	-0.134***	-0.133***	-0.133***
	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
spc2	-0.239*	-0.238*	-0.240*	-0.255**	-0.270**	-0.272**	-0.276**
	(0.125)	(0.125)	(0.125)	(0.124)	(0.124)	(0.124)	(0.124)
authnumber	-0.101***	-0.100***	-0.101***	-0.101***	-0.103***	-0.102***	-0.103***
	(0.027)	(0.027)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
humidimean_5	-0.555	-1.162**	-0.598	-0.510	-0.570*	-0.593*	-0.697**
	(0.507)	(0.453)	(0.388)	(0.329)	(0.330)	(0.326)	(0.324)
tempmean_5	-0.032***	-0.032***	-0.028***	-0.022***	-0.025***	-0.025***	-0.022***
	(0.011)	(0.010)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
snowmean_5	-0.650**	-0.655**	-0.557**	-0.397	-0.314	-0.390	-0.432*
	(0.329)	(0.282)	(0.255)	(0.245)	(0.245)	(0.248)	(0.253)
sunmean_5	-0.005	-0.039	-0.006	0.001	-0.003	-0.001	0.001
	(0.028)	(0.024)	(0.020)	(0.015)	(0.015)	(0.015)	(0.015)
rainmean_5	-0.008	-0.009	0.001	0.011**	0.008	0.008	0.009*
	(0.011)	(0.010)	(0.008)	(0.005)	(0.005)	(0.005)	(0.005)
Constant	-2.498***	-1.911***	-2.656***	-2.877***	-2.509***	-2.478***	-2.445***
	(0.735)	(0.645)	(0.580)	(0.536)	(0.526)	(0.525)	(0.525)
timefix	yes	yes	yes	yes	yes	yes	yes
indusfix	yes	yes	yes	yes	yes	yes	yes
Observations	43,203	43,203	43,203	43,203	42,934	42,934	42,934

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Since the air pollution level in the past five days will significantly affect the reported error rate, the main explanatory variables of Column (4) - Column (7) are the air pollution level in the last five days, and Column (4) is the mean value. The other three models take different thresholds (light pollution, moderate pollution, severe pollution) to obtain air pollution variables. The regression results are significantly positive, and the results are robust, and are not affected by the calculation method of air pollution variables. Previous regression results (Tables 8 and 9) show that the air pollution in the analyst’s city is serious, which will reduce the accuracy of the analyst’s forecast and increase the error rate of the analyst’s report, but air pollution will not make the forecast optimistic. Next, we evaluate the loss caused by the increase of error rate from the perspective of operational risk.

The explanatory variables of the above regression describe the analyst's report from different perspectives. The main explanatory variable is the air pollution level of the analyst's city. The forecast accuracy, recommendation degree and error rate of the analyst report can not affect the air quality of the city where the analyst is located, so there is no reverse causality in this study. The size of listed companies, the nature of equity, the ratio of assets to liabilities and other corporate characteristics, the experience, ability of analysts, and personal characteristics of the securities firms they belong to may affect the accuracy of the forecast report. Referring to the relevant research on the forecast accuracy of analyst reports, this paper controls the variables that may affect the forecast accuracy, the time-fixed effect and the industry-fixed effect of listed companies. It also controls the meteorological data of the cities where the analysts are located, and there is basically no problem of missing variables.

Because this paper takes the lead in studying the impact of air pollution on analyst error rate, this paper tries to control the analyst's personal characteristic variables, the meteorological data of the analyst's city, as well as the fixed effect of time and industry. In short, because of the comprehensive control variables, this paper can not consider the endogenous problem.

5.3.3 Operational risk and loss

Operational risk is the fluctuation of income or cash flow caused by customers, insufficient internal control, system or control failure, uncontrollable events in financial institutions. Man-made factor is an important cause of operational risk. As indicated in the previous study, air pollution increases the likelihood of human error, so we use operational risk to assess the loss caused by air pollution.

A study collected 307 loss incidents from 2000 to 2005, with a total loss of 11.874 billion yuan, with an average loss of 38.68 million yuan (Yuan Delei and Zhao Dingtao, 2007). This study only collects loss events, which can be used to calculate the loss frequencies of different amounts caused by operational risks. Because there is no total observation value, it is impossible to calculate the loss-free probability. The author uses Guotaian's bank financial research database to select the financial data (profit table) of listed and unlisted banks in China from 2000 to 2005 to calculate the total business income of the banking industry. Using banking business income as denominator and total loss as molecule, the probability of loss is 1.76% and the probability of lossless loss is 98.24%. According to the proportion of 307 loss incidents, the specific loss distribution can be obtained (see Table 11).

Table 11 Loss distribution of operational risk

amount of damages	Loss distribution	Loss distribution after pollution
0	0.98241	0.98235
[0-10]	0.00395	0.00397
[10-100]	0.00516	0.00517
[100-1000]	0.00361	0.00362
[1000-10000]	0.00315	0.00316
100000	0.00172	0.00172

Taking the median of each stage as the loss value, the expected loss value can be calculated as 64.905 million yuan according to the loss distribution. Because air pollution increases the error rate, i.e. changes the distribution of existing operational risks. Assuming that air pollution increases operational risk and leads to loss, the risk-free probability decreases and the loss probability increases at all stages, we can obtain the loss distribution after air pollution (see Table 11). Because the loss probability caused by air pollution increases, if AQI rises from good to light pollution (50 to 100), the loss will increase by 15%.The expected loss value also increased to 74.635 million yuan. Finally, it was estimated that the expected loss caused by air pollution ranged from mild to moderate pollution, which was about 9.735 million yuan.

5.4 Conclusions

The growing evidence on the effect of air pollution on behaviors beyond physical health indicates that pollution reduces workers' productivity. In this study, we find an additional and more specific effect, which is that pollution reduces workers' ability to do a precise job in their work. In other words, even if workers appear to remain as productive as without pollution exposure, the quality of their work may have deteriorated undetectably.

In economies with increasing reliance on cognitively-based work responsibilities with perhaps limited means for measuring quality-adjusted productivity very precisely, the influence of pollution could be widespread and detrimental. Such is the case for the financial analysts we study here. Without a systematic and comprehensive study, it would be difficult to detect that exposure to air pollution reduces forecasting ability and the quality of the financial reports, since each analyst could not be expected to be completely precise in their report.

BIBLIOGRAPHY

Kampa, Marilena, and E. Castanas. "Human health effects of air pollution." Environmental Pollution 151.2(2008):362-7.

Franklin, Barry A., Robert Brook, and C. Arden Pope. "Air pollution and cardiovascular disease." Current problems in cardiology 40.5 (2015): 207-238.

Hoek G; Krishnan RM; Beelen R; Peters A; Ostro B; Brunekreef B; Kaufman JD. Long-term air pollution exposure and cardio- respiratory mortality: a review[J]. Environ Health. 2013, 12(1):43.

Brook RD; Rajagopalan S; Pope CA 3rd; Brook JR; Bhatnagar A; Diez-Roux AV; Holguin F; Hong Y; Luepker RV; Mittleman MA; Peters A; Siscovick D; Smith SC Jr; Whitsel L; Kaufman JD; American Heart Association Council on Epidemiology and Prevention, A. C. O. Nutrition, and P. Activitymetabolism. "Particulate Matter Air Pollution and Cardiovascular Disease An Update to the Scientific Statement From the American Heart Association." Circulation 121.21(2010):2331-78.

Friedman, Michael S., et al. "Impact of Changes in Transportation and Commuting Behaviors During the 1996 Summer Olympic Games in Atlanta on Air Quality and Childhood Asthma." Jama the Journal of the American Medical Association 285.7(2001):897-905.

Kipen, Howard, et al. "Measurement of inflammation and oxidative stress following drastic changes in air pollution during the Beijing Olympics: a panel study approach." Annals of the New York Academy of Sciences 1203.1 (2010): 160-167.

Su, C., et al. "Short-term effects of fine particulate air pollution on cardiovascular hospital emergency room visits: a time-series study in Beijing, China. " International Archives of Occupational and Environmental Health 89.4(2016):1-17.

Su, C., et al. "Assessing responses of cardiovascular mortality to particulate matter air pollution for pre-, during- and post-2008 Olympics periods. " Environmental Research 142(2015):112-122.

Rich, David Q., et al. "Association between changes in air pollution levels during the Beijing Olympics and biomarkers of inflammation and thrombosis in healthy young adults." Jama 307.19 (2012): 2068-2078.

Rich, D. Q., et al. "Differences in Birth Weight Associated with the 2008 Beijing Olympics Air Pollution Reduction: Results from a Natural Experiment." Environmental Health Perspectives 123.9(2015).

Schwartz, J, D. W. Dockery, and L. M. Neas. "Is daily mortality associated specifically with fine particles?." Journal of the Air & Waste Management Association 46.10(1996):927-39.

Schwartz, J, and L. M. Neas. "Fine particles are more strongly associated than coarse particles with acute respiratory health effects in schoolchildren. " Epidemiology 11.1(2000):6-10.

Petersen, Floyd F. "Relationships of mortality with the fine and coarse fractions of long-term ambient PM10 concentrations in nonsmokers. "Journal of Exposure Analysis & Environmental Epidemiology10.5(2000):427-436.

Anderson, Hr, et al. "Particulate matter and daily mortality and hospital admissions in the west midlands conurbation of the United Kingdom: associations with fine and coarse particles, black smoke and sulphate."Occupational & Environmental Medicine 58.8(2001):504-10.

Brook R D, Shin H H, Bard R L, et al. Exploration of the Rapid Effects of Personal Fine Particulate Matter Exposure on Arterial Hemodynamics and Vascular Function during the Same Day[J]. Environmental Health Perspectives, 2011, 119(5):688-694.

Ai, Milojevic, et al. "Short-term effects of air pollution on a range of cardiovascular events in England and Wales: Case-crossover analysis of the MINAP database, hospital admissions and mortality." Heart100.14(2014):1093-1098.

Zanobetti, A, and J. Schwartz. "The effect of fine and coarse particulate air pollution on mortality: a national analysis. " Environmental Health Perspectives 117.6(2009):898-903.

Ostro, B, et al. "Fine particulate air pollution and mortality in nine California counties: results from CALFINE. " Environmental Health Perspectives 114.1(2006):29-33.

Zhang, Minsi, et al. "Economic assessment of the health effects related to particulate matter pollution in 111 Chinese cities by using economic burden of disease analysis." Journal of Environmental Management 88.4(2008):947–954.

Chen, Renjie, et al. "Coarse particles and mortality in three Chinese cities: The China Air Pollution and Health Effects Study (CAPES)." Science of the Total Environment 409.23(2011):4934-4938.

Bank, World. "Cost of Pollution in China: Economic Estimates of Physical Damages." (2007):1-151.

Ma, Yanjun, et al. "Fine particulate air pollution and daily mortality in Shenyang, China." Science of the Total Environment 409.13(2011):2473–2477.

Xie, W., et al. "Relationship between fine particulate air pollution and ischaemic heart disease morbidity and mortality. " Heart101.4(2015):257-63.

Schwartz, J, F. Laden, and A. Zanobetti. "The concentration-response relation between PM(2.5) and daily deaths. " Environmental Health Perspectives 110.10(2002):1025-9.

Yang, Chunxue, et al. "Alternative ozone metrics and daily mortality in Suzhou: The China Air Pollution and Health Effects Study (CAPES)." Science of the Total Environment 426.2(2012):83-89.

Lu, Feng, et al. "Systematic review and meta-analysis of the adverse health effects of ambient PM 2.5, and PM 10, pollution in the Chinese population." Environmental Research 136(2015):196-204.

Dockery, D. W., et al. "An association between air pollution and mortality in six U.S. cities. " New England Journal of Medicine 329.24(1993):1753-9.

Kasiscovick, Miller. "Long-term exposure to air pollution and incidence of cardiovascular events in women." Digest of the World Core Medical Journals 356.5(2007):447-58.

Laden, F, et al. "Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard Six Cities study. " American Journal of Respiratory & Critical Care Medicine 173.6(2006):667-72.

Naess, et al. "Relation between concentration of air pollution and cause-specific mortality: four-year exposures to nitrogen dioxide and particulate matter pollutants in 470 neighborhoods in Oslo, Norway. "American Journal of Epidemiology 165.4(2007):50392-50420.

Iii, C. Arden Pope, et al. "Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. " Jama the Journal of the American Medical Association 287.9(2002):1132-41.

Pope, C. Arden, et al. "Cardiovascular Mortality and Long-Term Exposure to Particulate Air Pollution Epidemiological Evidence of General Pathophysiological Pathways of Disease." Circulation 109.1(2004):71-7.

Raaschou-Nielsen, Ole, et al. "Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE)." Lancet Oncology14.9(2013):813-22.

Turner, Michelle C., et al. "Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers." American journal of respiratory and critical care medicine 184.12 (2011): 1374-1381.

Rd, C A Pope,. "Respiratory disease associated with community air pollution and a steel mill, Utah Valley. " American Journal of Public Health 79.5(1989):623-8.

3Rd, Pope Ca. "Particulate pollution and health: a review of the Utah valley experience. " Journal of Exposure Analysis & Environmental Epidemiology 6.1(1996):23-34.

Anderson, Michael L. "As the Wind Blows: The Effects of Long-Term Exposure to Air Pollution on Mortality." Michael Anderson (2015).

Chen, Yuyu, et al. "Evidence on the impact of sustained exposure to air pollution on life expectancy from China’s Huai River policy." Proceedings of the National Academy of Sciences 110.32 (2013): 12936-12941.

Ebenstein, Avraham , et al. "New evidence on the impact of sustained exposure to air pollution on life expectancy from China’s Huai River Policy." Pnas 110.32(2013):12936-12941.

Lee Liu. "Made in China: Cancer Villages." Environment Science & Policy for Sustainable Development 52.2(2010):8-21.

Janssen, N. A., et al. "Personal exposure to fine particulate matter in elderly subjects: relation between personal, indoor, and outdoor concentrations. " Journal of the Air & Waste Management Association50.7(2000):1133-43.

Sarnat, Jeremy A., and H. H. Suh. "Gaseous pollutants in particulate matter epidemiology: confounders or surrogates?" Environmental Health Perspectives 109.10(2001):1053-61

Rijnders, Esther, and B. Brunekreef. "Personal and outdoor nitrogen dioxide concentrations in relation to degree of urbanization and traffic density. " Environmental Health Perspectives 109.Suppl 3(2001):411-7.

Oglesby, L, et al. "Validity of ambient levels of fine particles as surrogate for personal exposure to outdoor air pollution--results of the European EXPOLIS-EAS Study (Swiss Center Basel). " Journal of the Air & Waste Management Association 50.7(2000):1251-61.

Chen W, Zheng R, Zhang S, et al. Cancer incidence and mortality in China, 2013[J]. Cancer Letters, 2017, 24(1):1-8.

Dunford A, Schumacher S E, Savova V, et al. Escape from X-Inactivation Tumor Suppressor (EXITS) Genes Are Associated with Excess Incidence of Cancer in Men[J]. 2015.

Green M S, Symons M J. A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies[J]. Journal of chronic diseases, 1983, 36(10): 715-723.

Breslow N E, Day N E, Davis W. Statistical methods in cancer research[M]. Lyon: International Agency for Research on Cancer, 1987.

Dionne G, Doherty N, Fombaron N. Adverse Selection in Insurance Markets[C]// Universite de Montreal, Departement de sciences economiques, 2000:249-62.

Hoy M. Categorizing risks in the insurance industry[J]. The Quarterly Journal of Economics, 1982, 97(2): 321-336.

Akerlof G. VThe Market for Lemons: Qualitative Uncertainty and the Market MechanismV[J]. Quarterly Journal of Economics, 1970, 84(3).

Rothschild M, Stiglitz J. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information[J]. The quarterly journal of economics, 1976: 629-649.

Dahlby B G. The welfare effects of prohibiting statistical discrimination in insurance markets[J]. Research paper, 1980: 80-5.

Wilson C. A model of insurance markets with incomplete information[J]. Journal of Economic theory, 1977, 16(2): 167-207.

Miyazaki H. The rat race and internal labor markets[J]. The Bell Journal of Economics, 1977: 394-418.

Spence M. Product differentiation and performance in insurance markets[J]. Journal of Public Economics, 1978, 10(3): 427-447.

Dahlby B G. Adverse Selection and Statistical Discrimination[J]. Journal of Public Economics, 1983, 20(1):121-130.

Chiappori P A, Salanie B. Testing for Asymmetric Information in Insurance Markets[J]. Working Papers, 1997, 108(1):56-78.

Cohen A. Asymmetric Information and Learning: Evidence from the Automobile Insurance Market[J]. Social Science Electronic Publishing, 2005, 87(2):197-207.

Chang T Y, Huang W, Wang Y. Something in the air: Pollution and the demand for health insurance[J].

The Review of Economic Studies, 2018, 85(3): 1609-1634.

Pierre-André Chiappori. Econometric Models of Insurance under Asymmetric Information[M]// Handbook of Insurance. Springer Netherlands, 2000:365-393.

Dionne G, Michaud P C, Dahchour M. SEPARATING MORAL HAZARD FROM ADVERSE SELECTION AND LEARNING IN AUTOMOBILE INSURANCE: LONGITUDINAL EVIDENCE FROM FRANCE[J]. Social Science Electronic Publishing, 2013, 11(4):897–917.

Cohen A, Siegelman P. Testing for Adverse Selection in Insurance Markets[J]. Journal of Risk & Insurance, 2010, 77(1):39–84.

Puelz R, Snow A. Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market[J]. Journal of Political Economy, 1994, 102(2):236-257.

Barseghyan L, Molinari F, O'Donoghue T, et al. The Nature of Risk Preferences: Evidence from Insurance Choices[J]. American Economic Review, 2013, 103(6):2499-2529.

Fang H, Keane M P, Dan S. Sources of Advantageous Selection: Evidence from the Medigap Insurance Market[J]. Journal of Political Economy, 2008, 116(2):303-350.

Spinnewijn J. Insurance and Perceptions: How to Screen Optimists and Pessimists[J]. Social Science Electronic Publishing, 2013, 123(569):606–633.

Agrawal A, Chen M A. Do analyst conflicts matter? Evidence from stock recommendations[J]. The journal of Law and Economics, 2008, 51(3): 503-537.

Brown L D, Hagerman R L, Griffin P A, et al. Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings[J]. Journal of Accounting and Economics, 1987, 9(1): 61-87.

Baginski S P, Hassell J M. The market interpretation of management earnings forecasts as a predictor of subsequent financial analyst forecast revision[J]. Accounting Review, 1990: 175-190.

Brown L D. Analyst forecasting errors: Additional evidence[J]. Financial Analysts Journal, 1997, 53(6): 81-88.

Botosan C A, Harris M S. Motivations for a change in disclosure frequency and its consequences: An examination of voluntary quarterly segment disclosures[J]. Journal of Accounting Research, 2000, 38(2): 329-353.

Barber B, Lehavy R, McNichols M, et al. Can investors profit from the prophets? Security analyst recommendations and stock returns[J]. The Journal of Finance, 2001, 56(2): 531-563.

Byard D, Shaw K W. Corporate disclosure quality and properties of analysts' information environment[J]. Journal of Accounting, Auditing & Finance, 2003, 18(3): 355-378.

Bradshaw M T, Drake M S, Myers J N, et al. A re-examination of analysts’ superiority over time-series forecasts of annual earnings[J]. Review of Accounting Studies, 2012, 17(4): 944-968.

Clement M B. Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter?[J].

Journal of Accounting and Economics, 1999, 27(3): 285-303.

Chen Q, Jiang W. Analysts' weighting of private and public information[J]. The Review of financial studies, 2005, 19(1): 319-355.

Das S, Levine C B, Sivaramakrishnan K. Earnings predictability and bias in analysts' earnings forecasts[J]. Accounting Review, 1998: 277-294.

Dechow P M, Hutton A P, Sloan R G. The relation between analysts' forecasts of long-term earnings growth and stock price performance following equity offerings[J]. Contemporary Accounting Research, 2000, 17(1): 1-32.

Dugar A, Nathan S. The effect of investment banking relationships on financial analysts' earnings forecasts and investment recommendations[J]. Contemporary Accounting Research, 1995, 12(1): 131-160.

Easley D, O'Hara M, Paperman J. Financial analysts and information-based trade[J]. Journal of Financial Markets, 1998, 1(2): 175-201.

Francis J, Philbrick D. Analysts' decisions as products of a multi-task environment[J]. Journal of Accounting Research, 1993: 216-230.

Fang L, Yasuda A. The effectiveness of reputation as a disciplinary mechanism in sell-side research[J]. The Review of Financial Studies, 2009, 22(9): 3735-3777.

Gleason C A, Lee C M C. Analyst forecast revisions and market price discovery[J]. The Accounting Review, 2003, 78(1): 193-225.

Gu Z, Li Z, Yang Y G. Institutional investors, commission fees, and analyst bias[J]. 2009.

Groysberg B, Healy P M, Maber D A. What Drives Sell-Side Analyst Compensation at High-Status Investment Banks?[J]. Journal of Accounting Research, 2011, 49(4): 969-1000.

Hunton J E, McEwen R A. An assessment of the relation between analysts' earnings forecast accuracy, motivational incentives and cognitive information search strategy[J]. Accounting Review, 1997: 497-515.

Healy P M, Hutton A P, Palepu K G. Stock performance and intermediation changes surrounding sustained increases in disclosure[J]. Contemporary accounting research, 1999, 16(3): 485-520.

Healy P M, Palepu K G. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature[J]. Journal of accounting and economics, 2001, 31(1-3): 405-440.

Hilary G, Shen R. The role of analysts in intra-industry information transfer[J]. The Accounting Review, 2013, 88(4): 1265-1287.

Hope O K. Accounting policy disclosures and analysts' forecasts[J]. Contemporary Accounting Research, 2003, 20(2): 295-321.

Jennings R. Unsystematic security price movements, management earnings forecasts, and revisions in consensus analyst earnings forecasts[J]. Journal of Accounting Research, 1987: 90-110.

Jacob J, Lys T Z, Neale M A. Expertise in forecasting performance of security analysts[J]. Journal of Accounting and Economics, 1999, 28(1): 51-82.

Jackson A R. Trade generation, reputation, and sell-side analysts[J]. The Journal of Finance, 2005, 60(2): 673-717.

Jegadeesh N, Kim W. Value of analyst recommendations: International evidence[J]. Journal of Financial Markets, 2006, 9(3): 274-309.

Jacob J, Rock S, Weber D P. Do non-investment bank analysts make better earnings forecasts?[J]. Journal of Accounting, Auditing & Finance, 2008, 23(1): 23-61.

Kross W, Ro B, Schroeder D. Earnings expectations: The analysts' information advantage[J]. Accounting Review, 1990: 461-476.

Kadan O, Madureira L, Wang R, et al. Analysts' industry expertise[J]. Journal of Accounting and Economics, 2012, 54(2-3): 95-120.

Lang M H, Lundholm R J. Corporate disclosure policy and analyst behavior[J]. Accounting review, 1996: 467-492.

Lees F A. Public disclosure of corporate earnings forecasts[C]. Conference Board, 1981.

Lin H, McNichols M F. Underwriting relationships, analysts' earnings forecasts and investment recommendations[J]. Journal of Accounting and Economics, 1998, 25(1): 101-127.

Lim T. Rationality and analysts' forecast bias[J]. The Journal of Finance, 2001, 56(1): 369-385.

Lehavy R, Li F, Merkley K. The effect of annual report readability on analyst following and the properties of their earnings forecasts[J]. The Accounting Review, 2011, 86(3): 1087-1115.

Libby R, Tan H T, Hunton J E. Does the Form of Management's Earnings Guidance Affect Analysts' Earnings Forecasts?(Retracted)[J]. The Accounting Review, 2006, 81(1): 207-225.

Malloy C J. The geography of equity analysis[J]. The Journal of Finance, 2005, 60(2): 719-755.

Michaely R, Womack K L. Conflict of interest and the credibility of underwriter analyst recommendations[J]. The Review of Financial Studies, 1999, 12(4): 653-686.

Mola S, Guidolin M. Affiliated mutual funds and analyst optimism[J]. Journal of Financial Economics, 2009, 93(1): 108-137.

Meng X. Analyst reputation, communication, and information acquisition[J]. Journal of Accounting Research, 2015, 53(1): 119-173.

Malmendier U, Shanthikumar D. Do security analysts speak in two tongues?[J]. The Review of Financial Studies, 2014, 27(5): 1287-1322.

O'BRIEN P C, McNichols M F, Hsiou-Wei L. Analyst impartiality and investment banking relationships[J]. Journal of Accounting Research, 2005, 43(4): 623-650.

Previts G J, Bricker R J, Robinson T R, et al. A content analysis of sell-side financial analyst company reports[J]. Accounting Horizons, 1994, 8(2): 55.

Park C W, Stice E K. Analyst forecasting ability and the stock price reaction to forecast revisions[J]. Review of Accounting Studies, 2000, 5(3): 259-272.

Schipper K. Analysts' forecasts[J]. Accounting horizons, 1991, 5(4): 105.

Stickel S E. Reputation and performance among security analysts[J]. The Journal of Finance, 1992, 47(5): 1811-1836.

Waymire G. Additional evidence on the accuracy of analyst forecasts before and after voluntary management earnings forecasts[J]. Accounting Review, 1986: 129-142.

Zhang X F. Information uncertainty and analyst forecast behavior[J]. Contemporary Accounting Research, 2006, 23(2): 565-590.

Howarth E, Hoffman M S. A multidimensional approach to the relationship between mood and weather[J]. British Journal of Psychology, 1984, 75(1):15–23.

Brook R D, Shin H H, Bard R L, et al. Exploration of the Rapid Effects of Personal Fine Particulate Matter Exposure on Arterial Hemodynamics and Vascular Function during the Same Day[J]. Environmental Health Perspectives, 2011, 119(5):688-694.

VAN DONKELAAR, Aaron, et al. Global Fine Particulate Matter Concentrations from Satellite for Long-Term Exposure 2 Assessment 3. Assessment, 2015, 3: 1-43.

VAN DONKELAAR, Aaron, et al. Regional Estimates of Chemical Composition of Fine Particulate Matter using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors. Environmental science & technology, 2019.

World Health Organization. Ambient (outdoor) air quality and health http://www.who.int/mediacentre/factsheets/fs313/zh/

Wang Jun, Gao Feng. Study on the Coexistence of Adverse and Positive Choices in China's Health Insurance Market [J]. Financial Research, 2008 (11): 160-170.

Wang Jun, Gao Feng. Empirical Study on Adverse Selection in China's Automobile Insurance Market [J]. Financial Research, 2007 (12): 223-230.

Peng Fei. Survival Analysis [M]. Renmin University Press, 2004.

Wan Yan Ruiyun, Suo Lingyan. Analysis of Insurance Consumption Decision-making Behavior: A Research Framework of Behavioral Insurance [J]. Insurance Research, 2016 (1): 15-29.

Wagner J G, Allen K, Yang H, et al. Cardiovascular Depression in Rats Exposed to Inhaled Particulate Matter and Ozone: Effects of Diet-Induced Metabolic Syndrome[J]. Environmental Health Perspectives, 2014, 122(1):27.

Arvin B M, Lew B. Life satisfaction and environmental conditions: Issues for policy[J]. International Journal of Global Environmental Issues, 2012, 12(1):76-90.

Marques S, Lima M L. Living in grey areas: Industrial activity and psychological health[J]. Journal of Environmental Psychology, 2011, 31(4):314-322.

Fonken L K, Xu X, Weil Z M, et al. Air pollution impairs cognition, provokes depressive-like behaviors and alters hippocampal cytokine expression and morphology[J]. Molecular Psychiatry, 2011, 16(10):987.

Wagner J G, Allen K, Yang H, et al. Cardiovascular Depression in Rats Exposed to Inhaled Particulate

Matter and Ozone: Effects of Diet-Induced Metabolic Syndrome[J]. Environmental Health Perspectives, 2014, 122(1):27.

Lim Y H, Kim H, Kim J H, et al. Air pollution and symptoms of depression in elderly adults[J]. Environmental Health Perspectives, 2012, 120(7):1023.

Gary W. Evans, Joan M. Campbell. Psychological Perspectives on Air Pollution and Health[J]. Basic & Applied Social Psychology, 1983, 4(2):137-169.

Cavanagh J T O, Carson A J, Sharpe M, et al. Psychological autopsy studies of suicide: A systematic review.[J]. Psychological Medicine, 2003, 33(3):395.

Rotton J, Frey J. Psychological costs of air pollution: Atmospheric conditions, seasonal trends, and psychiatric emergencies[J]. Population & Environment, 1984, 7(1):3-16.

Rotton J, Frey J. Air pollution, weather, and violent crimes: concomitant time-series analysis of archival data.[J]. Journal of Personality & Social Psychology, 1985, 49(5):1207.

Bakian A V, Huber R S, Coon H, et al. Acute Air Pollution Exposure and Risk of Suicide Completion[J]. American Journal of Epidemiology, 2015, 181(5):295-303.

Kim C, Jung S H, Kang D R, et al. Ambient particulate matter as a risk factor for suicide.[J]. American Journal of Psychiatry, 2010, 167(9):1100.

Fleehart S, Fan V S, Nguyen H Q, et al. Prevalence and correlates of suicide ideation in patients with COPD: a mixed methods study[C]// Western Institute of Nursing Communicating Nursing Research Conference. 2014:1321-1329.

Jee S H, Kivimaki M, Kang H C, et al. Cardiovascular disease risk factors in relation to suicide mortality in Asia: prospective cohort study of over one million Korean men and women.[J]. European Heart Journal, 2011, 32(22):2773-80.
Cavanagh J T O, Carson A J, Sharpe M, et al. Psychological autopsy studies of suicide: A systematic review.[J]. Psychological Medicine, 2003, 33(3):395.
Ha K H, Cho J, Cho S K, et al. Air pollution and unintentional injury deaths in South Korea[J]. Environmental Science & Pollution Research, 2015, 22(10):1-9.

Szyszkowicz M, Willey J B, Grafstein E, et al. Air Pollution and Emergency Department Visits for Suicide Attempts in Vancouver, Canada[J]. Environmental Health Insights,2010,4(2010-10-15), 2010, 4(4):79-86.

Bakian A V, Huber R S, Coon H, et al. Acute Air Pollution Exposure and Risk of Suicide Completion[J]. American Journal of Epidemiology, 2015, 181(5):295.

Chang, T., Graff Zivin, J., Gross, T., and Neidell, M., Particulate Pollution and the Productivity of Pear Packers, American Economic Journal: Economic Policy, Vol. 8 (2016), No. 3, p. 141 – 169.

Chang, T., Graff Zivin, J., and Neidell, M., The effect of pollution on office workers: Evidence form call centers in China, American Economic Journal: Applied Economics, forthcoming.

Chew, S., Huang, W., and Li, X., Haze and decision making: A natural laboratory experiment, Working paper, 2017.

Graff Zivin, J., and Neidell, M., The Impact of Pollution on Worker Productivity, American Economic Review, Vol. 102 (2012), No. 7, p. 3652 – 3673.

He, J., Liu, H., and Salvo, A., Severe air pollution and labor productivity: Evidence from industrial towns in China, American Economic Journal: Applied Economics, Vol. 11 (2019), No. 1, p. 173 – 201.

Lagercrantz, L., Wistrand, M., Willen, U., Wargocki, P., Witterseh, T., and Sundell, J., Negative impact of air pollution on productivity: Previous Danish findings repeated in new Swedish test room., Proc. Of Healthy Buildings, Vol. 1 (2000), p. 653 – 658.

Li, S., Lien, J.W., and Yuan, J., Air Pollution Increases Lottery Gambling, Working Paper, February 2019.

Lichter, A., Pestel, N., and Sommer, E., Productivity effects of air pollution: Evidence from professional soccer, Labour Economics, Vol. 48 (2017), p. 54 – 66.

Zhang, X., Chen, X., and Zhang, X., The impact of exposure to air pollution on cognitive performance, Proceedings of the National Academy of Sciences, Vol. 115 (2018), 37, p. 9193 – 9197.

Zhang X., Zhang, X., and Chen, X., Happiness in the air: How does a dirty sky affect mental health and subjective well-being?, Journal of Environmental Economics and Management, Vol, 85 (2017), p. 81 – 94.

AFDELING
Adres
3000 LEUVEN, België
tel. + 32 16 00 00 00
fax + 32 16 00 00 00
@kuleuven.be
www.kuleuven.be