The Health Impact of Air Pollution in China: Evidence from Emergency Admissions

A Thesis

Presented to the Faculty of the Graduate School

Of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science in Applied Economics and Management

By

Kaihang Shi

August, 2016

ABSTRACT

This thesis aims to investigate the health impact of air pollution in China using data from four urban districts in Shanghai. Outdoor air pollution has been shown to be a leading environmental cause of respiratory and heart related diseases. In recent years, $PM_{2.5}$ (air particles with a diameter of 2.5 micrometers or less) has become an increasing health concern in China, as its concentration in some cities often reach over 10 times of the risk level defined by WHO. According to recent studies, over 96% of the Chinese population lives in areas where $PM_{2.5}$ concentration exceeds 75 $\mu g/m^3$ (CNAAQS Level 2 standard). This thesis utilizes three different data set: daily $PM_{2.5}$ concentration and weather conditions data; daily emergency admissions due to respiratory diseases in hospitals in four urban districts in Shanghai; and traffic volume data in these district. This thesis tries to estimate the impact of air pollution on human health in China through its impact on emergency admissions. To address the endogeneity issue in $PM_{2.5}$ level due to omitted variables, I use traffic volume as the instrumental variables (IV) for $PM_{2.5}$. The results show that an increase of PM2.5 by 10 $\mu g/m^3$ in average, it would increase emergency admissions due to respiratory and cardiovascular diseases by 8.3% , and cause averagely 2.7% more daily medicine expenses (￥19，386) in respiratory and cardiovascular related diseases.

BIOGRAPHICAL SKETCH

Kaihang Shi studied mechanical engineering in Shanghai Jiaotong University before obtaining a dual degree in civil and environmental engineering from University of Michigan. At Cornell University, his study focused on the empirical analysis of transportation and environmental policies.

ACKNOWLEDGMENTS

First and Foremost, I would like to express my gratitude to my thesis advisor Professor Shanjun Li who gave me valuable guidance throughout my career as a Master's student in applied economics and management. He taught me empirical skills that lay a strong foundation for my research.

I am also thankful for the advice from Professor Calum G. Turvey. He provided the valuable advices for this thesis.

Last but not least, I would like to express my gratitude to my friends and family for the support and willingness to spend time to read my thesis and give me feedbacks.

Thank you.

Kaihang Shi

TABLE OF CONTENTS

1. Introduction and Background

Studies indicate that air pollution in China has increased since 2001 and China is now the largest source of SO_2 emissions in the world. Outdoor air pollution has been a leading environmental cause of respiratory and heart related diseases. If a premature death is valued using a cost of 1 million yuan, which is the marginal willingness to pay to avoid death risk, the welfare loss associated with air pollution will be 3.8 percent of Chinese GDP in 2007[1]. In recent years, PM2.5 particles which are air pollutants with a diameter of 2.5 micrometers or less have raised people's concern in China, as the PM2.5 concentration in some cities is reported over 10 times the risk level defined by WHO. According to recent studies, over 96% of the Chinese population lives in areas where PM2.5 concentration exceeds 75 μg/m3 (CNAAQS Level 2 standard)[2]. Research has estimated that 2.1 million premature respiratory deaths are associated globally and annually with anthropogenic PM2.5-related cardiopulmonary diseases (93%) and lung cancer (7%)[3]. Recognizing the negative impacts of air pollution (PM2.5 concentration level) on Chinese people's health and daily life, we investigate that this particulate matter is emitted during the combustion of solid and liquid fuels such as power generation, domestic heating and vehicle engines. Currently most papers focus on air pollution like Sulphur dioxide and nitrogen dioxide in China. There is no detailed $PM_{2.5}$ data before 2013 in most cities in China. There still lacks the analysis on the impacts of $PM_{2.5}$ level on people' respiratory health problem in city

[1]The world bank, China Cost of pollution, Green edition, 2007
[2] Zongwei Ma et al., Estimating Ground-Level $PM_{2.5}$ in China Using Satellite Remote Sensing, American Chemical Society, 2014
[3] Raquel A Silva et al., Global premature mortality due to anthropogenic outdoor air pollution and the contribution of past climate change, Environmental Research Letters, 2013

like Shanghai. This thesis aims to find the relation between $PM_{2.5}$ level data and people's respiratory health problem using econometric methods. This result will be useful to policy designer in China who wants to relive the negative impact of $PM_{2.5}$ pollution on people's health. According to the result, some welfare analysis in case of people's health will be presented in this paper.

The broad objective of my thesis research is to investigate the impact of $PM_{2.5}$ level on people's respiratory health problem in 4 urban districts of Shanghai. In the 2SLS model, I use the traffic volume data as the instrumental variable because it is correlated with the local $PM_{2.5}$ and not related with the local weather condition. In those 4 urban districts of Shanghai, I collected the emergency respiratory patients count in each AAA hospital which is the highest level of hospital in China of each districts. Using the 2SLS model, I am seeking to address the following questions: (1) How does $PM_{2.5}$ level and other weather condition affect local resident's respiratory health? (2) Does $PM_{2.5}$ have time-lag effect on people's respiratory health outcomes? What is the time-lag effect of $PM_{2.5}$ level on people's respiratory health outcomes? (3)What is the impact of month(season) and year (time trend) on patient count and medicine expenses in respiratory and cardiovascular related diseases? (4) What is the daily welfare loss (medicine expenses) for people in Hongkou district be if 10 $\mu g/m^3$ more $PM_{2.5}$ level is generated? (5) What is the daily welfare loss (medicine expenses) for people in Hongkou district if a sudden burst of PM2.5 pollution is generated on each day?

These are important questions. First, Shanghai, one of the largest city in China, is the leader in setting local and national policies. Many other local authorities are looking for guidance in solving local resident's health problem caused by air pollution. It is important to examine them carefully and understand how air pollution affect people's respiratory health. Second, our analysis aims to quantify the economy loss of air pollution in terms of people's medical expenses and provide guidance on how to improve the policy design to prevent more respiratory health problem caused by air pollution.

2. Literature Review

Although China has made tremendous economic success in recent 40 years, air pollution problem begins to be a significant health and economic loss. In 1997, the World Bank observed that ambient concentrations of particulates and sulfur dioxide in many Chinese cities are among the highest in the world and are significantly above WHO (World Health Organization) guidelines and Chinese air quality standards (World Bank, 1997, p.9). China as the largest developing countries has some of the worst air quality in the world (Kan et al., 2012) [4]. Since January 2013, when a hazardous dense haze covered 1.4 million km^2 of China and affected more than 800 million people in that area (Xu et al., 2013)[5], China raised a national concern about air pollution especially $PM_{2.5}$ level.

Particulate matter (PM) pollution is closely related to haze and strong related to adverse health effects (Pope and Dockery, 2006)[6]. Among particulate matter, $PM_{2.5}$ is fine particulate matter with a diameter of 2.5 micrometers or less, which have adverse impacts on cardiovascular disease (CVD) and respiratory disease(RD) (Samoli et al.[7],2005; Lepeule et al.[8], 2012; Krewski et al., 2009[9]; Kloog et al., 2014[10]). The

[4] Kan, H., Chen, R., Tong, S., 2012. Ambient air pollution, climate change, and population health in China. Environ. Int. 42, 10-19

[5] Xu. P., Chen, Y., Haze, Ye X., 2013. Air pollution, and health in China. Lancet 382(9910), 2067

[6] Pope, C.A., Dockey, D.W.,2006. Health effects of the fine particulate air pollution: lines that connect. J. Air Waste Manag. Assoc. 56, 709-742

[7] Samoli, E.,Analitis,A.,Touloumi,G.,etal.,2005.Estimatingtheexposure–response relationships between particulate matter and mortality within the APHEA multicity project. Environ Health Perspect113(1),88–95.

[8] Lepeule, J.,Laden,F.,Dockery,D.,etal.,2012.Chronicexposureto fine particles and mortality: an extended follow-up of the Harvard Six Cities study from 1974 to 2009. Environ. Health Perspect.120(7), 965–970.

[9] Krewski,D.,Jerrett,M.,Burnett,R.T.,etal.,2009.Extended follow-up and spatial analysis of the American cancer society study linking particulate air pollution and mortality. Res. Rep. Health Eff. Inst. 140, 5-114(discussion 115-36)

[10] Kloog, I., Nordio, F., Zanobetti, A., et al., 2014. Short term effects of particle exposure on hospital admissions in the mid-Atlantic States: a population estimate. PLoS One 9 (2), e88578.

mankind's sources of $PM_{2.5}$ mainly comes from coal combustion emission and secondary particle deriving from oxidation of primary vehicle emission gas (Sulfur and nitrogen oxides). From the report about air pollution level in China from Deng (2015)[11], about 45% of the $PM_{2.5}$ pollution sources comes from transportation related emission, and 23.5% comes from coal combustion emission, 13.2% comes from dust/soil, 10.6% comes from secondary organic particles, 7.4% comes from industrial emission, and 0.4% from metallurgical.

Currently China has very severe $PM_{2.5}$ pollution in the whole country. Over 96% of the Chinese population lives in area where $PM_{2.5}$ exceeds healthy standard (CNAAQS Level 2 standard 75μg/m^3) according to the research by Zongwei Mat et. al.(2014). Research by Robert A. Rohde et al.(2015)[12] shows that the observed $PM_{2.5}$ pollution is calculated to contribute to 1.6 million deaths per year in China, roughly 17% of all deaths in China. Figure 1 from website Berkeley Earth shows the $PM_{2.5}$ pollution level distribution all over the country. We can find most part of Chinese territory has very high $PM_{2.5}$ level which is far above the 75μg/m^3 CNAAQS Level 2 standard in China.

We can find the hourly $PM_{2.5}$ concentration measured and reported by the US embassy in Beijing since 2008. Since 2013 Shanghai Environmental Protection Bureau (EPB) started to monitor $PM_{2.5}$ for city coverage. Although the severe air

[11] Furong Deng, Chinese Urban Transportation-related Air Pollution and Its Health Impact. 2015
[12] Robert A. Rohde et al, Air pollution in China: Mapping of Concentrations and Sources, PLoS ONE 10(8): e0135749. doi:10.1371/journal. pone.0135749

pollution in China has been warmly discussed in the last few years, we still lack some quantitative analysis literature about the impact of $PM_{2.5}$ on people's respiratory health using $PM_{2.5}$ data in large cities like Shanghai. A rich literature in the epidemiological field has been found according to the study by C. Arden Pope III et al. (2012)[13], that persuasive evidence that exposure to fine particulate air pollution has adverse effects on cardiopulmonary health after reviewing research pursued since 1997. Raquel A Silva et al. (2013) found that 2.1 million premature respiratory deaths are associated globally and annually with anthropogenic $PM_{2.5}$ -related cardiopulmonary diseases (93%) and lung cancer (7%). Many evidences have shown that $PM_{2.5}$ concentration level is the most fetal reason causing respiratory and heart related diseases. Another reason I use $PM_{2.5}$ concentration as the main pollution factor is that $PM_{2.5}$ may cause residents' transportation mode change. Residents in Shanghai concern $PM_{2.5}$ much more than other pollutants like NO_x and CO. People may prefer stay at home when $PM_{2.5}$ is too high. This point is very interesting for further research.

The research of the impact of fine particulate in China is categorized by types of pollutant (PM_{10}, $PM_{2.5}$, and heavy metal in the particulate matter) as well as types of mortality data (non-accidental, cardiovascular, and respiratory) according to a study by Lu et al. (2015)[14]. In Lu's study, they restrict the literature on PM_{10} only, and conclude a significant wide range of 23%-67% increase in mortality for a 10μg/m^3 in

[13] C. Arden Pope III et al., Health effects of Fine Particulate Air Pollution: Lines that Connect, Journal of the Air and Waste Management Association, Volume 56, Issue 6, 2006

[14] Lu, Feng, Dongqun Xu, Yibin Cheng, Shaoxia Dong, Chao Guo, Xue Jiang, and Xiaoying Zheng. 2015. 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: 196-204.

ambient PM_{10} for a long-term exposure. There still lacks analysis about the negative impact of $PM_{2.5}$ for a short-term exposure which will be my topic in this thesis. Other researches focuses on using mortality data in certain age groups, such as infants (Currie and Neidell, 2004[15]; Beatty and Shimshack, 2014[16]), children (Janke, 2014[17]), or seniors (Schwartz, 2000[18]). All those age groups are believed to be more vulnerable to air pollution exposure. Not like other research in mortality increase caused by $PM_{2.5}$ pollution, this thesis focus on the impact of $PM_{2.5}$ in terms of patient count of emergency admission and medicine expenses in cardiovascular and respiratory related diseases about 4 urban districts in Shanghai.

[15] Currie, Janet, and Matthew Neidell. 2004. "Air pollution and infant health: What can we learn from California's recent experience." National Bureau of Economic Research

[16] Beatty, Timothy KM, and Jay P Shimshack. 2014. \Air pollution and children's respiratory health: A cohort analysis." Journal of Environmental Economics and Management, 67(1): 39-57.

[17] Janke, Katharina. 2014. \Air pollution, avoidance behaviour and children's respiratory health: Evidence from England." Journal of health economics, 38: 23-42.

[18] Schwartz, Joel. 2000. \The distributed lag between air pollution and daily deaths." Epidemiology, 11(3): 320-326.

3. Data

In this research we have four datasets. The first dataset is traffic volume data in 4 urban districts of Shanghai. The data is based on the real-time vehicle speed and GPS location collected from Qiangsheng Taxi Company which is the largest taxi company in Shanghai. I filter the daily traffic volume according to the taxi GPS location data in the range of 500 meter within GPS location of each $PM_{2.5}$ monitor. The traffic volume data has sample for 4 urban districts in Shanghai which is Jingan, Huangpu, Hongkou and Pudong district, and the date ranges from April 1st to April 30th in 2015. The traffic volume data is shown in figure 2.

In figure 2, we can find the average traffic volume in the range of 500m of the monitor station is around 25,000 per day. The traffic volume of each district is quite similar and the fluctuation is much depending on the date. There is not much difference between days in April which means the impact of traffic volume on local $PM_{2.5}$ is quite steady in that month.

The second dataset is weather condition data in Shanghai collected from National Meteorological Information Center[19]. The daily dataset includes temperature (0.1celsius), humidity (%), rain volume (0.1mm), wind velocity (0.1m/s), sunshine hours (0.1 hour). All the data is collected from Baoshan monitor station which is the only weather condition monitor in Shanghai. Time ranges from January 2013 to June 2015.

[19] http://www.cma.gov.cn/en2014/

The third dataset is $PM_{2.5}$ data in Shanghai collected from Ministry of Environmental Protection[20] of China. I choose the 4 urban districts data in the time period from 2013-2015. The data is plotted in figure 3.

From the figure 3 we find all those 4 urban districts have very similar $PM_{2.5}$ level, $PM_{2.5}$ pollution tends to be severe in months like January, December and be light in months in June, July. Especially in month December and January, some days have extraordinarily high $PM_{2.5}$ level in the year 2013. As those 4 urban districts are very close to each other, it is reasonable to have similar $PM_{2.5}$ level. The reason that $PM_{2.5}$ is very high in the winter of 2013 is that a hazardous dense haze covered 1.4 million KM^2 of China and affected more than 800 million people including most area of Shanhai. After 2013, Chinese government begins to monitor $PM_{2.5}$ level in the national wide. Year 2013 is a signal that Chinese people come to be aware of air pollution problem which is a severe obstacle in the way of developing China.

In figure 4, I plot the relation between traffic volume data and $PM_{2.5}$ data in the range of 2015.04.01 to 2015.04.30. We can find the relation between $PM_{2.5}$ level and the traffic volume data in the range of 500m near the monitor station. The correlation between daily traffic volume data and $PM_{2.5}$ level in Hongkou District is 0.0603 which is very high.

[20] http://english.mep.gov.cn/

From figure 4, we know that there is a very close relation between traffic volume and $PM_{2.5}$ volume data which will be a very useful information for further discussion. The correlation number is 0.603 which means the traffic volume within 500 m of the monitor is highly correlated with $PM_{2.5}$ level. That result is quite impressive and it is a good evidence that how transportation affected the urban air quality.

The fourth dataset is the detailed data of emergency registration in respiratory diseases. The dataset includes the emergency medical center record for respiratory diseases for 4 urban districts in Shanghai since 2010. I prefer the emergency registration data to the daily admission data of the hospital, because the admission data may be enlarged because of unequal distribution of health resources in China, which means large portion of patients in Shanghai are not local. The emergency record can reflect the local residents' health condition. The sample covers the closest AAA hospital which is the highest level hospital in the local district. Each hospital's location which is shown in figure 5 is within the range of 1-3 miles of the PM2.5 monitor to reflect the most sensitive influence of PM2.5 level on people's respiratory and cardiovascular related health. The count of respiratory diseases category includes asthma, rhinitis, pneumonia, emphysema etc.

Table 1 and 2 provides the summary statistics of the dataset of weather condition, PM2.5 level and patient count. In table 1, we calculate the mean, median, standard deviation, range, minimum, maximum value of temperature(0.1 celsius), Humidity(%), Rainvolume(mm), Windvelocity(0.1m/s) and sunshine hour(0.1 hour). In table 2, we calculate the mean, median, standard deviation, range, minimum, maximum value of PM2.5 (μg/m3) and Respiratory Disease Patient Count (people) in 4 different urban district hospitals.

In Figure 6 and 7, I plot the data of patient count in four different districts and medicine expenses in respiratory and cardiovascular related diseases in Hongkou district. In Figure 6, I found the trend of patient count is very close to the $PM_{2.5}$ level trend in figure 3. Patient count is high in month December, January. We can find there are some days which have extraordinary high patient count in the winter of 2013. This is caused by a huge hazardous dense haze covered 80% area of China. In figure 7, the medicine expenses data is also high in month December and January, but the point is a little bit looser between each other. In the year 2014, the difference in each month is not as obvious as 2013. I think that is caused by the pollution in 2013 which may raise more chronic respiratory and cardiovascular disease. Those diseases will add to medicine expenses in 2014.

4 Model Estimation

4.1 OLS model with fixed effect feature

In OLS model, we measure the impact of pollution and weather condition on the respiratory and cardiovascular disease count for the duration from 2013.1.18 to 2015.06.30. The respiratory disease count in each urban district is estimated by the following equation (1) and equation (2):

$$\mathrm{Health}_{dt} = \gamma_0 + \gamma_k \mathrm{Pollution}_{dt} + \theta Z_{dt} + \varphi_{dt} + \epsilon_{dt} \qquad (1)$$

$$Z_{dt}: \mathrm{temperature}_{dt};\ \mathrm{humidity}_{dt}; \mathrm{windvelocity}_{dt}; \mathrm{sunshinehours}_{dt} \qquad (2)$$

In the equation, Health_{dt} is the respiratory and cardiovascular related emergency records from the hospital in district d and time t; $\mathrm{Pollution}_{dt}$ is the $PM_{2.5}$ level in district d at time t; γ_0 is the constant term; ψ_{dt} is fixed effect term which includes fixed day of week, fixed month, fixed year and fixed district; ε_{dt} is the error term; Z_{dt} is the weather condition variable in district d and time t which is listed in equation (2).

The reason I add fixed effect of year, month, day of week, and district, is that all those effect may affect the patient count variable. For example, Shanghai as one of the largest cities in China is developing rapidly each year, so the pollution level may be different as the population, traffic congestion condition and industrialization level is increasing each year. I set year dummy variable as the indicator of year effect on air pollution in Shanghai. The month is the indicator of temperature which may affect people's respiratory diseases, so I set month fixed effect. Day of week fixed effect will eliminate the weekday avoidance in patient's behavior. District fixed effect will eliminate the difference of patient count number in different districts.

The result of OLS models with 4 different fixed effect features are shown in the table 3. In Table 3, I gradually add district fixed effect, month fixed effect and day of week fixed effect. The R-square result shows that OLS 1-day model 4 with fixed district, month and day of week effect has the best estimation result. This result makes sense that estimation with all fixed effect considered will be more significant. From the estimated coefficient, I found negative coefficient value for temperature and windvelocity which means lower temperature and lower windvelocity will cause higher patient count of emergency admission in the local area. The positive coefficients of $PM_{2.5}$, Humidity, and sunshine hours show that higher $PM_{2.5}$, Humidity, and Sunshine hours may cause more respiratory patient count. The coefficient of rainvolume is 0 which means there is no relation between people’s health and rainvolume variable. The dummy variable value of year 2014 and 2015 shows that people’s respiratory health is getting worse as years goes by. In case of t-statistics value, only the estimation of $PM_{2.5}$, Temperature, sunshine hours, year dummy and constant value is significant enough different from zero.

This OLS model may lack some critical point that people’s respiratory diseases may be caused by $PM_{2.5}$ value days or weeks before. Evidence has shown that the effect of $PM_{2.5}$ pollution will not reach zero until 5 days after exposure[21] (Joel Schwartz, 2000). Given the distribution of sensitivities likely in the general population, this time-lag effect is biologically plausible. In the next section, I will use OLS with time dynamic feature to solve this problem.

4.2 OLS model with dynamic feature

In this OLS model, we will estimate the impact of $PM_{2.5}$ level on health adding dynamic features. The impact of ambient air pollution on respiratory and cardiovascular related disease rate is not just determined by the pollution on the same day. The disease may be induced by the pollution days or weeks before the symptoms. First I assume 7 days as a forward displacement to test the time-lag effect in one week, then I assume 3-week as a forward displacement to test a longer time-lag effect. The estimation model is as following equation (3) - (5)

$$\mathrm{Health}_{dt} = \gamma_0 + \sum_{k=0}^{7} \gamma_k \mathrm{Pollution}_{d(t-7)} + \theta Z_{dt} + \varphi_{dt} + \epsilon_{dt} \qquad (3)$$

$$\mathrm{Health}_{dt} = \gamma_0 + \sum_{k=0}^{3} \gamma_k \mathrm{Pollution}_{d(t-3)} + \theta Z_{dt} + \varphi_{dt} + \epsilon_{dt} \qquad (4)$$

$$Z_{dt}: \mathrm{temperature}_{dt};\ \mathrm{humidity}_{dt}; \mathrm{windvelocity}_{dt}; \mathrm{sunshinehours}_{dt} \qquad (5)$$

Health_{dt} is the respiratory and cardiovascular related emergency records from the hospital in district d and time t; We can check the impact using the $PM_{2.5}$ level in 7 days before the symptoms which is $\mathrm{Pollution}_{d(t-7)}$; We can check the impact using the

[21] Joel Schwartz, May 2000, "The Distributed Lag between Air Pollution and Daily Deaths" Epidemiology, Vol.11, No.3

$PM_{2.5}$ level in 3 weeks before the symptoms which is $Pollution_{d(t-3)}$; φ_{dt} is the fixed effect. ε_{dt} is the error term; Z_{dt} is the weather condition variable in district d and time t.

From the OLS model with dynamic feature above, we can estimate how $PM_{2.5}$ pollution affect patient count with a 7-day lag effect and 3-week lag effect. The following table 4 compare those two results of dynamic OLS model with fixed month district and day of week effect.

In table 4, I found the t-statistic value of PM25-2week and PM25-3week is not significant enough. The time-lag effect is based on 1 week period. In the OLS 7-day model 4, I found the estimation has a better R-square value. The time-lag effect is significant enough until PM25-6 day which means 6 days before. The patient count is mostly affected by $PM_{2.5}$ level 2 days before and has least impact from 1 day before. All the other estimated coefficient is very similar to the OLS model without dynamic feature.

The dynamic feature OLS model solves the problem of time-lag effect, and we found the significant time-lag is around 5 days before the emergency record time. This provides a very important information for our estimation. Although we may have a very reasonable result from this updated model, we doubt that there is some endogenous problem between $PM_{2.5}$ variable and other weather condition variable.

Research from Shao-Hang Chu (2004)[22] shows that solar radiations and temperature will have impact on $PM_{2.5}$ level. In this way, I proposed a 2SLS model to verify if the endogeneity between $PM_{2.5}$ variable and weather condition variables will make a big difference.

4.3 2SLS model for robustness check

In 2SLS model, we will estimate the impact of $PM_{2.5}$ level on health using the dynamic feature of 7-day as a forward displacement. We argue that traffic volume data near the $PM_{2.5}$ monitor is a good instrumental variable of $PM_{2.5}$ level variable because of the following three reasons: firstly, they are correlated with the $PM_{2.5}$ level; secondly, they are uncorrelated with the weather condition variables; thirdly, they are excluded from the patient count variable. Thus, we use traffic volume data in the range of 500m of the $PM_{2.5}$ Monitor GPS location as our instrumental variable. The estimation model is shown in the following equations.

First stage: $$\mathrm{Pollution}_{dt} = \alpha_0 + \alpha_1 \mathrm{Trafficvolume}_{dt} + \epsilon_{dt} \quad (6)$$

Second stage: $$\mathrm{Health}_{dt} = \gamma_0 + \sum_{k=0}^{7} \gamma_k \mathrm{Pollution}_{d(t-7)} + \theta Z_{dt} + \varphi_{dt} + \epsilon_{dt} \quad (7)$$

$$Z_{dt}: \mathrm{temperature}_{dt};\ \mathrm{humidity}_{dt}; \mathrm{windvelocity}_{dt}; \mathrm{sunshinehours}_{dt} \quad (8)$$

Health_{dt} is the respiratory and cardiovascular related emergency records from the hospital in district d and time t; We can check the impact using the $PM_{2.5}$ level in

22 Shao-Hang Chu. PM2.5 episodes as observed in the speciation trends network. Atmospheric Environment, Volume 38, Issue 31, October 2004, Pages 5237-5246

the period of 7 days before the symptoms which is $\sum_{k=0}^{7} \gamma_k \text{Pollution}_{d(t-7)}$; φ_{dt} is the fixed effect. ε_{dt} is the error term; Z_{dt} is the weather condition in district d and time t. In the first stage, we have the traffic volume data from 2015.4.1 to 2015.4.30 as the instrumental variables.

From the 2SLS model above, we can estimate how $PM_{2.5}$ affect patient count using dynamic features while solving the endogeneity problem between $PM_{2.5}$ and weather condition. The following table 5 shows the result of 2SLS model estimation.

From the above result in table 5, we find the f-test statistics is larger or close to 10 which means the instrumental variables has a close relevance with the endogenous variable ($PM_{2.5}$ level). The result of 2SLS model is quite similar with OLS model but with a higher R-square value. In table 6, we compare the result of OLS model with dynamic features and 2SLS model to verify the robustness of OLS model.

From the above comparison, we find the estimation of $PM_{2.5}$ variable of both model is very similar in value of coefficient and t-statistics. Although 2SLS solve the endogeneity problem in $PM_{2.5}$ variables, it still don't make big difference in the result. 2SLS model only have 155 observations which cause a higher R-square value, but OLS model will have better statistics significance with a much larger sample of 2550 observations. I conclude that OLS model with 7-day dynamic feature is robust enough in the estimation.

4.4 OLS model with dynamic feature using medical expenses data

I also collected the respiratory and cardiovascular related medicine sales data in Hongkou district in period from 2013.1.18 to 2.15.6.30. I run the same model as OLS 7-day model using patient count data. From the result we can find how $PM_{2.5}$ data affect people's health in terms of medical expenses. The estimation model is as following:

$$\mathrm{Medicalexpense}_t = \gamma_0 + \sum_{k=0}^{7} \gamma_k \mathrm{Pollution}_{(t-7)} + \theta Z_t + \varphi_t + \epsilon_t \qquad (9)$$

$$Z_t: \mathrm{temperature}_t;\ \mathrm{humidity}_t; \mathrm{windvelocity}_t; \mathrm{sunshinehours}_t \qquad (10)$$

$\mathrm{Medicalexpense}_{dt}$ is the respiratory and cardiovascular related medicine sales data in the unit of ￥10,000 from the hospital in Hongkou district and time t; We can check the impact using the $PM_{2.5}$ level in the period of 7 days before the symptoms which is $\sum_{k=0}^{7} \gamma_k \mathrm{Pollution}_{(t-7)}$; φ_t is the fixed effect. ε_t is the error term; Z_t is the weather condition in Hongkou district and time t.

Using this model, we can estimate how $PM_{2.5}$ level affect the medical expenses in Hongkou District from year 2013 to 2015. The estimation result is shown in the following table 7.

From table 7, I found the coefficient of $PM_{2.5}$ variable is significant enough for 1-4 days before the current day. The temperature, humidity, sunshine hours variable is significant enough and have very similar value to the OLS model using patient count number. The R-square value is very good. The only difference from the result of patient count case is that the coefficient of $PM_{2.5}$ variable is not significant in 5 days before the current day. In the next section, I will analyze the result in details.

5 Empirical Results

5.1 Impact of month and year on patient count (hint from OLS model using patient count data and medicine sales data)

I use OLS model to find impact of month and year. In first case, I set fixed effect for district and year and set 11 dummy variables for February, March, April, May, June, July, August, September, October, November and December where January is the zero dummy. In the model using patient count data, if one month dummy variable is 1, that means this month will induce 1 more medicine expense comparing to January. In the model using medicine sales data, if one month dummy variable is 1, that means this month will induce ￥10,000 more medicine expenses in respiratory and cardiovascular related diseases. Using dummy variable value, we can find how each month affects people's health in terms of patient count and medicine expenses which is shown in Figure 8.

From Figure 8, we found December has the worst impact on respiratory and cardiovascular patient count while June has the lightest impact. Months in winter and spring tend to induce higher emergency respiratory diseases, and months in summer and autumn tend to have less emergency respiratory diseases. This result matches the conclusion from David M. Stieb et al. (2002)[23] that colder season tend to be more significant in the impact on people’s respiratory and cardiovascular health.

[23] David M. Stieb, Stan Judek & Richard T. Burnett(2002) Meta-Anlysis of Time-Series Studies of Air Pollution and Mortality: Effects of Gases and Particles and the Influence of Cause of Death, Age, and Season, Journal of the Air & Waste Management Association, 52:4, 470-484

In the case of year, I set fixed month and district effect. The zero dummy is year 2013. In the model using patient count data, if one year dummy variable equals to 1, that means this year induces one more emergency respiratory and cardivascular related diseases compared to year 2013. In the model using medicine sales data, if one month dummy variable is 1, that means this month will induce ￥10,000 more medicine expenses in respiratory and cardiovascular related diseases. The result of impact of year is in figure 9.

From the following figure 9, we found compared to year 2013, year 2014 and 2015 will induce 2.73 and 3.05 more respiratory disease. It shows that people's respiratory health is getting worse as time goes by.

5.2 Impact of $PM_{2.5}$ level (hint from OLS dynamic model using patient count data and medicine sales data)

I propose a case study that all the average $PM_{2.5}$ level has increased by 10μg/m3 . In this study, I assume all the other variables are fixed, then I estimate the average patient count increase and medicine expense increase using 2SLS model (7 days dynamic feature). If average PM2.5 level increases by 10 $\mu g/m^3$, it will cause averagely 8.3% more daily patient count in emergency records in respiratory and cardiovascular related diseases for all districts. In the case of Shanghai, I find some result about mortality increase caused by $PM_{2.5}$ pollution to compare. I found the mortality due to CVD ranges from 0.24(0.08,0.4) (95% CI) to 1.08(0.33, 1.83) (95% CI)[24] and

[24] Song, W.,Guo,H.,Dai,H.A.,2008.Time-seriesstudyontheeffectsofPM2.5 on daily mortality in Shanghai.

mortality due to RD ranges from 0.22(-0.08, 0.52)(95% CI)[25]. I think my result make up the blank of $PM_{2.5}$ impact on people's patient count in emergency admission.

In the case of medicine expenses, if average PM2.5 level increases by 10 $\mu g/m^3$， it will cause averagely 2.7% more daily medicine expenses(￥19,386) in that AAA hospital in respiratory and cardiovascular related diseases in Hongkou District. We can find $PM_{2.5}$ increase impact is different in daily patient count and daily medicine expenses. Here I come to a conclusion that daily patient count in emergency record has a higher reflection on $PM_{2.5}$ pollution than medicine expense. The reason I think it is because $PM_{2.5}$ pollution will cause more apparent symptoms like cough and rhinorrhea while they may disappear after the $PM_{2.5}$ level is lowered. The medicine expenses may have some time-effect after the pollution, and people tend to buy medicine after a while after the symptoms.

In the second case study, I use 2SLS model to test the impact of $PM_{2.5}$ in a dynamic feature of 7 days. I set fixed time and district effect and run 2SLS model for the duration 2015.4.1-2015.4.30. The estimation result shows the increase in patient count and medicine expenses on current day if PM2.5 level increases by 10 $\mu g/m^3$ in each day between $PM_{2.5}$ day to $PM_{2.5}$-7 day. The impact of $PM_{2.5}$ level in day t to day (t-7) is shown in figure 10.

Epidemiology 19 (6), S167. inShanghai.Epidemiology19(6),S167.

[25] Wong,C.M.,Thach,T.Q.,Chau,P.Y.,etal.,2010.Part4.Interactionbetweenair pollution and respiratory viruses: time-series study of daily mortality and hospital admissionsinHongKong.Res.Rep.HealthEff.Inst.154,283–362.

From figure 10, we can find the $PM_{2.5}$ level 2 days forward has the largest impact on people's respiratory health. The second highest is $PM_{2.5}$ level 5 days before and then the $PM_{2.5}$ on the same day. This phenomenon is very interesting, and give us a vivid explanation that air pollution has a time-lag effect on people's health. This result also shows that time-lag effect exists in medicine expenses which is smaller in the reflection of $PM_{2.5}$ pollution increase.

In the third case study, I set all the other variables fixed and $PM_{2.5}$ level increase by 20% for all days. Then I use 2SLS model (7 days dynamic feature) to estimate the average increase of patient count in emergency records in respiratory and cardiovascular related diseases for 4 different districts. In figure 11, we found different reflection of the local residents in each district.

From the result in figure 11, patient count in Jingan district increases by 10.4%, Hongkou district increases by 23.3%, Pudong district increases by 14.0%, and Huangpu district increases by 11.7%. This result shows that Hongkou district has the most sensitive reflection on $PM_{2.5}$ increase. The second highest is Pudong district. Huangpu district has the least sensitive reflection.

6. Conclusion

(1) How $PM_{2.5}$ level and other weather condition effect local resident's respiratory health? (2) Does $PM_{2.5}$ have time-lag effect on people's respiratory health outcomes? What is the time-lag effect of $PM_{2.5}$ level on people's respiratory health outcomes? (3)What is the impact of month(season) and year (time trend) on patient count and medicine expenses in respiratory and cardiovascular related diseases? (4) What is the daily welfare loss (medicine expenses) for people in Hongkou district be if $10\,\mu g/m^3$ more $PM_{2.5}$ level is generated? (5) What is the daily welfare loss(medicine expenses) for people in Hongkou district if a sudden burst of PM2.5 pollution is generated on each day?

My thesis did an empirical research on the impact of $PM_{2.5}$ and weather condition on local residents' respiratory health in terms of short-term exposure risk. My research present a quantitative analysis using rare data from Hospitals, Taxi Company and $PM_{2.5}$ level and weather condition data in Shanghai to empirically find the health impact of $PM_{2.5}$. In the case study, I finished a solid estimation of increased patient count induced by more air pollution ($PM_{2.5}$) and the estimated economy loss in terms of medicine. Here comes the conclusion:

(1) $PM_{2.5}$ level has very negative impact on people's health. The result is very significant. Some other weather condition like sunshine hours and temperature will also significantly affect people's respiratory health. Lower temperature and more sunshine hours will induce more respiratory disease.

(2) $PM_{2.5}$ has time-lag effect on people's respiratory health. The time-lag is about 5 days. Impact of $PM_{2.5}$ level 6 days before comes to be not significant enough. $PM_{2.5}$ level 2 days before has the greatest impact. The sequential for the impact value is (T-2>T-5>T>T-3>T-4>T-1).

(3) Colder months tend to induce more respiratory diseases while warmer months are the opposite. The time series analysis shows that people's respiratory health is getting worse as year goes by. Huangpu district tend to have more patient with respiratory diseases. The sequential for the impact value is (Huangpu District>Pudong District>Jingan District>Hongkou District). The reason may be caused by the difference in medical services, population and resident's lifestyle. This is an interesting topic for further research.

(4) In the case study, we find Hongkou District has the most sensitive reflection on higher $PM_{2.5}$ level. The sequential is (Hongkou District>Pudong District>Huangpu District>Jingan District). If average $PM_{2.5}$ level increase by $10 \mu g/m^3$, it will cause averagely 2.7% more daily medicine expenses(￥19,386） in respiratory and cardiovascular related diseases in Hongkou District, while it will cause averagely 8.3% more daily patient count in emergency records in respiratory and cardiovascular related diseases for all districts.

(5) In the case that a sudden burst of $PM_{2.5}$ is generated on each day, the daily medicine expenses will increase as plotted in figure 9.

From the above conclusion of the empirical result, I found not only $PM_{2.5}$ the only factor which may induce respiratory disease. Time, location, people’s lifestyle and their own health condition may also be critical reason for the illness. In this thesis, I focused on the effect of environment factors, and the most interesting found is the time distribution of the $PM_{2.5}$ time-lag effect on people’s respiratory health. This shed light on the insight of policy dealing with health protection from air pollution in China. The preliminary analysis of the welfare loss caused by more medicine uses in one urban district of Shanghai provides a very quantitative description of the negative impact of air pollution. Some extension such as the interaction effect between air pollution and district, air pollution and time, air pollution and population should be further analyzed in my model. The demographic data of the patient will be added into my model for a better description of people’s behavior in my study in the future.

References

1. Zongwei Ma et al., Estimating Ground-Level $PM_{2.5}$ in China Using Satellite Remote Sensing, American Chemical Society, 2014
2. The world bank, China Cost of pollution, Green edition, 2007
3. Furong Deng, Chinese Urban Transportation-related Air Pollution and Its Health Impact. 2015
4. Kan, H., Chen, R., Tong, S., 2012. Ambient air pollution, climate change, and population health in China. Environ. Int. 42, 10-19
5. Robert A. Rohde et al, Air pollution in China: Mapping of Concentrations and Sources, PLoS ONE 10(8): e0135749. doi:10.1371/journal. pone.0135749
6. Pope, C.A., Dockey, D.W.,2006. Health effects of the fine particulate air pollution: lines that connect. J. Air Waste Manag. Assoc. 56, 709-742
7. Samoli, E.,Analitis,A.,Touloumi,G.,etal.,2005.Estimatingtheexposure–response relationships between particulate matter and mortality within the APHEA multicity project. Environ Health Perspect113(1),88–95.
8. Lepeule, J.,Laden,F.,Dockery,D.,etal.,2012.Chronicexposureto fine particles and mortality: an extended follow-up of the Harvard Six Cities study from 1974 to 2009. Environ. Health Perspect.120(7), 965–970.
9. Krewski,D.,Jerrett,M.,Burnett,R.T.,etal.,2009.Extended follow-up and spatial analysis of the American cancer society study linking particulate air pollution and mortality. Res. Rep. Health Eff. Inst. 140, 5-114(discussion 115-36)
10. Kloog, I., Nordio, F., Zanobetti, A., et al., 2014. Short term effects of particle

exposure on hospital admissions in the mid-Atlantic States: a population estimate. PLoS One 9 (2), e88578.

11. C. Arden Pope III et al., Health effects of Fine Particulate Air Pollution: Lines that Connect, Journal of the Air and Waste Management Association, Volume 56, Issue 6, 2006
12. Raquel A Silva et al., Global premature mortality due to anthropogenic outdoor air pollution and the contribution of past climate change, Environmental Research Letters, 2013
13. C. Arden Pope III et al., Health effects of Fine Particulate Air Pollution: Lines that Connect, Journal of the Air and Waste Management Association, Volume 56, Issue 6, 2006
14. Lu, Feng, Dongqun Xu, Yibin Cheng, Shaoxia Dong, Chao Guo, Xue Jiang, and Xiaoying Zheng. 2015. Systematic review and meta-analysis of the adverse health effects of ambient PM 2.5 and PM 10 pollution in theChinese population." Environmental research, 136: 196-204.
15. Currie, Janet, and Matthew Neidell. 2004. “Air pollution and infant health: What can we learn from California’s recent experience.” National Bureau of Economic Research
16. Beatty, Timothy KM, and Jay P Shimshack. 2014. \Air pollution and children's respiratory health: A cohort analysis." Journal of Environmental Economics and Management, 67(1): 39-57.
17. Janke, Katharina. 2014. \Air pollution, avoidance behaviour and children's respiratory health: Evidence from England." Journal of health economics, 38: 23-

42.

18. Schwartz, Joel. 2000. \The distributed lag between air pollution and daily deaths." Epidemiology, 11(3): 320-326.

19. David Y.H. et al., PM2.5 in China: Measurements, sources, visibility and health effects, and mitigation, Particuology, 2014

20. David M. Stieb, Stan Judek & Richard T. Burnett(2002) Meta-Anlysis of Time-Series Studies of Air Pollution and Mortality: Effects of Gases and Particles and the Influence of Cause of Death, Age, and Season, Journal of the Air & Waste Management Association, 52:4, 470-484

21. Joel Schwartz, May 2000, "The Distributed Lag between Air Pollution and Daily Deaths" Epidemiology, Vol.11, No.3

22. Song, W.,Guo,H.,Dai,H.A.,2008.Time-seriesstudyontheeffectsofPM2.5 on daily mortality in Shanghai. Epidemiology 19 (6), S167. inShanghai.Epidemiology19(6),S167.

23. 1 Wong,C.M.,Thach,T.Q.,Chau,P.Y.,etal.,2010.Part4.Interactionbetweenair pollution and respiratory viruses: time-series study of daily mortality and hospital admissionsinHongKong.Res.Rep.HealthEff.Inst.154,283–362.

Appendix: Figures

Figure 1: Real-time Map of Air Pollution

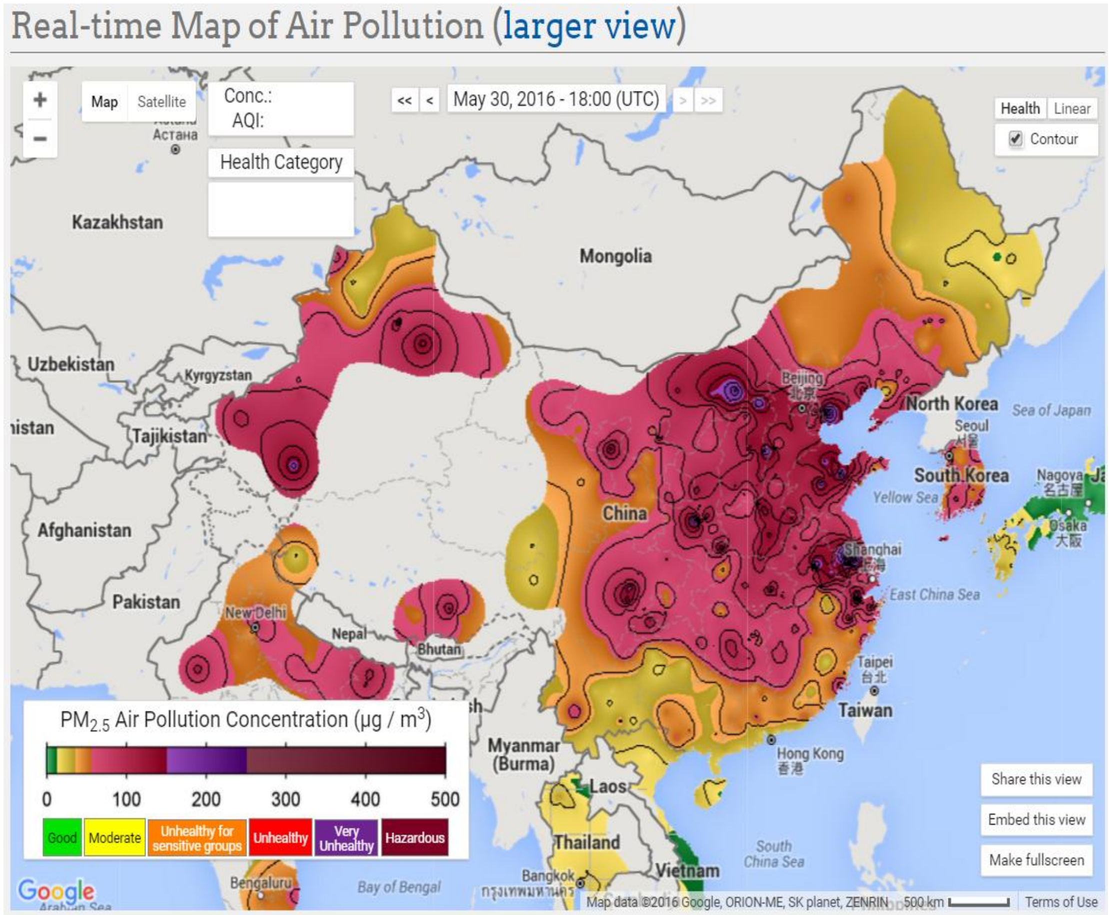


Figure 2: Traffic volume data in 500 meter range of PM2.5 monitor station

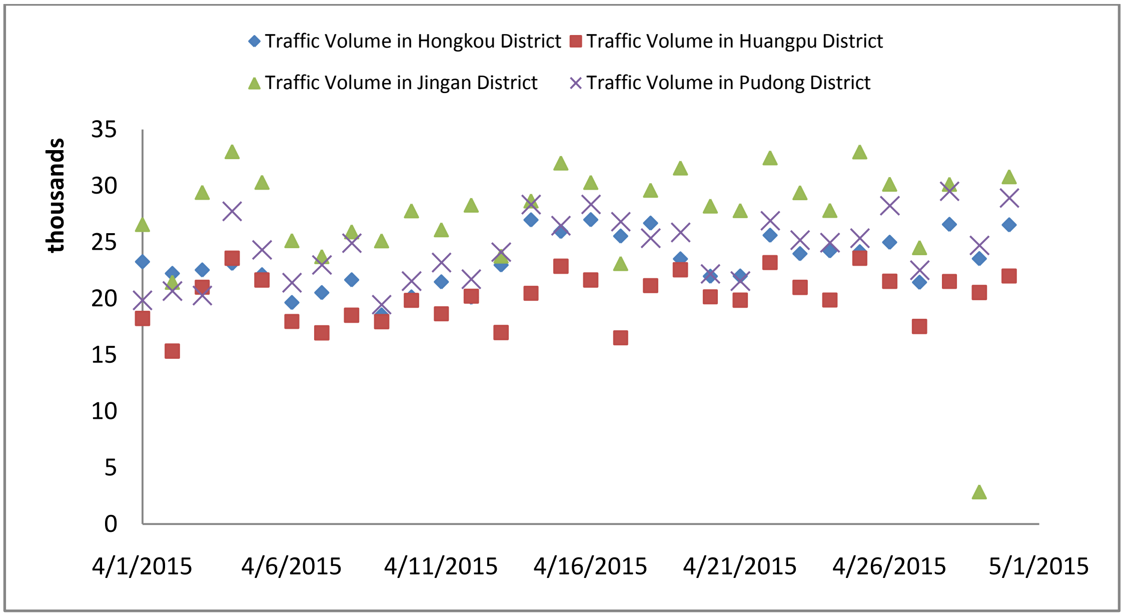


Figure 3: PM2.5 level in 4 different districts of Shanghai from 2013-2015

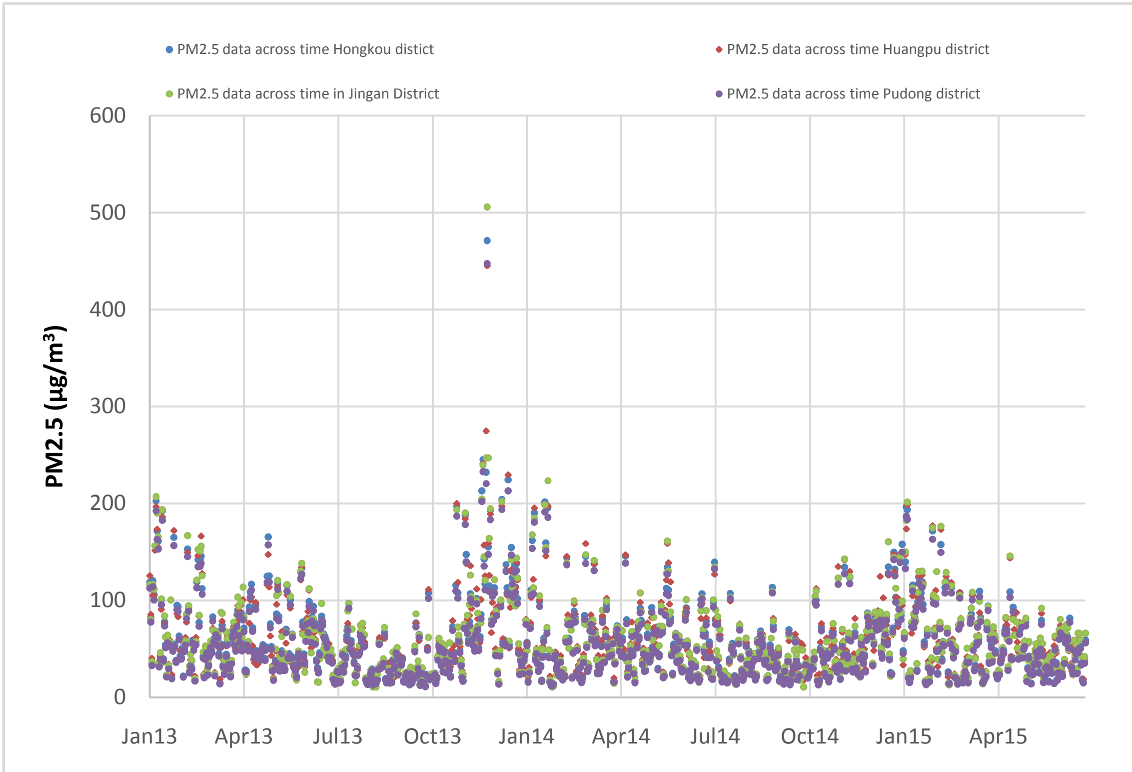


Figure 4: PM2.5 level data and traffic volume data in the range of April 2015

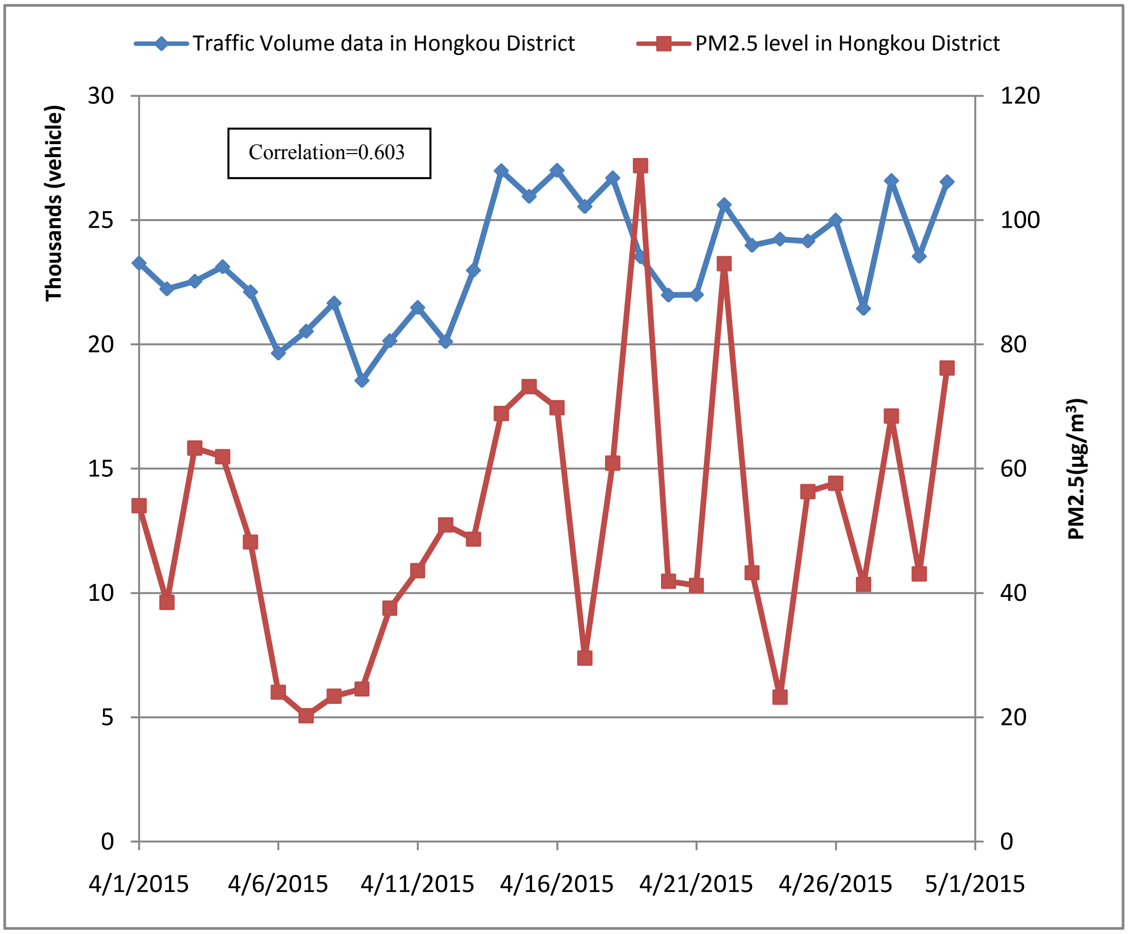


Figure 5. Location of PM2.5 monitor and Hospital in Four Urban District of Shanghai

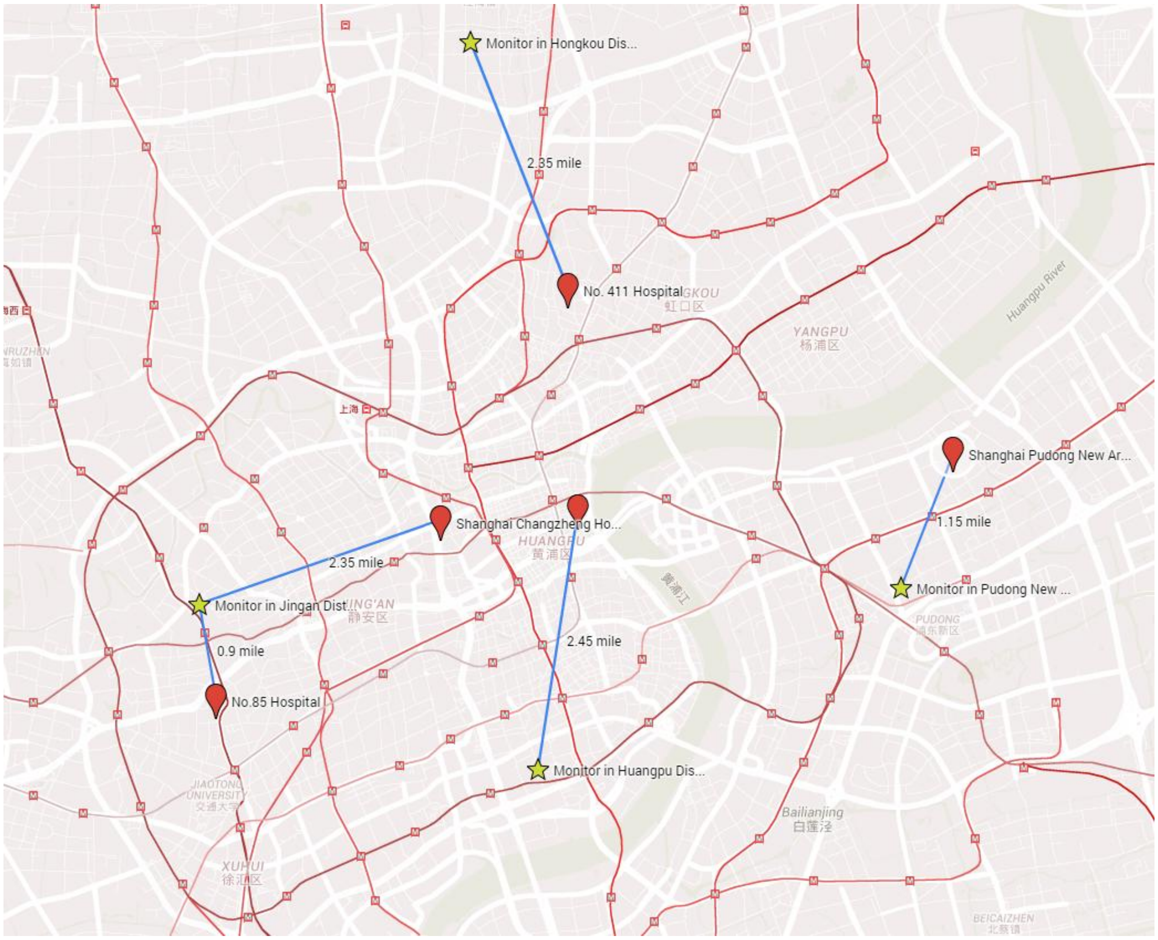


Figure 6. Patient Count Data in four different districts

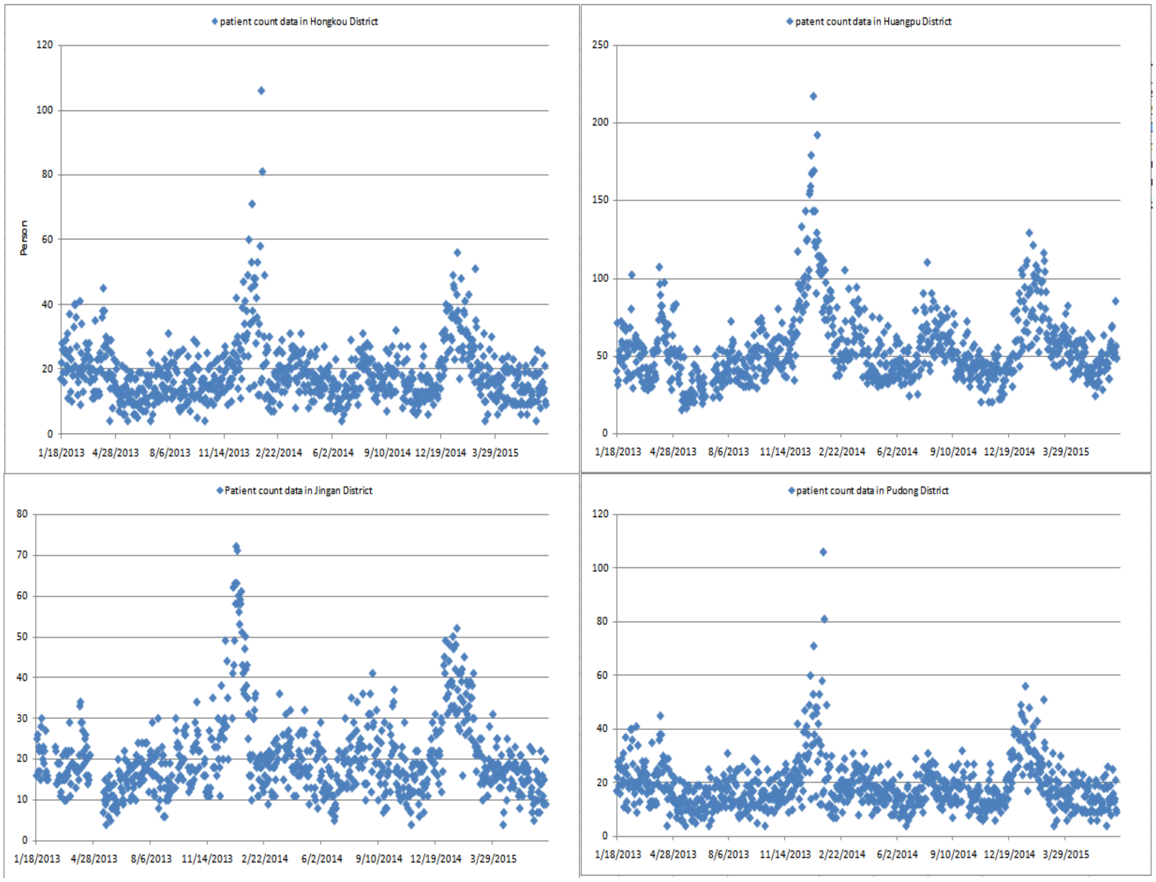


Figure 7. Medicine Expenses Data in Hongkou District

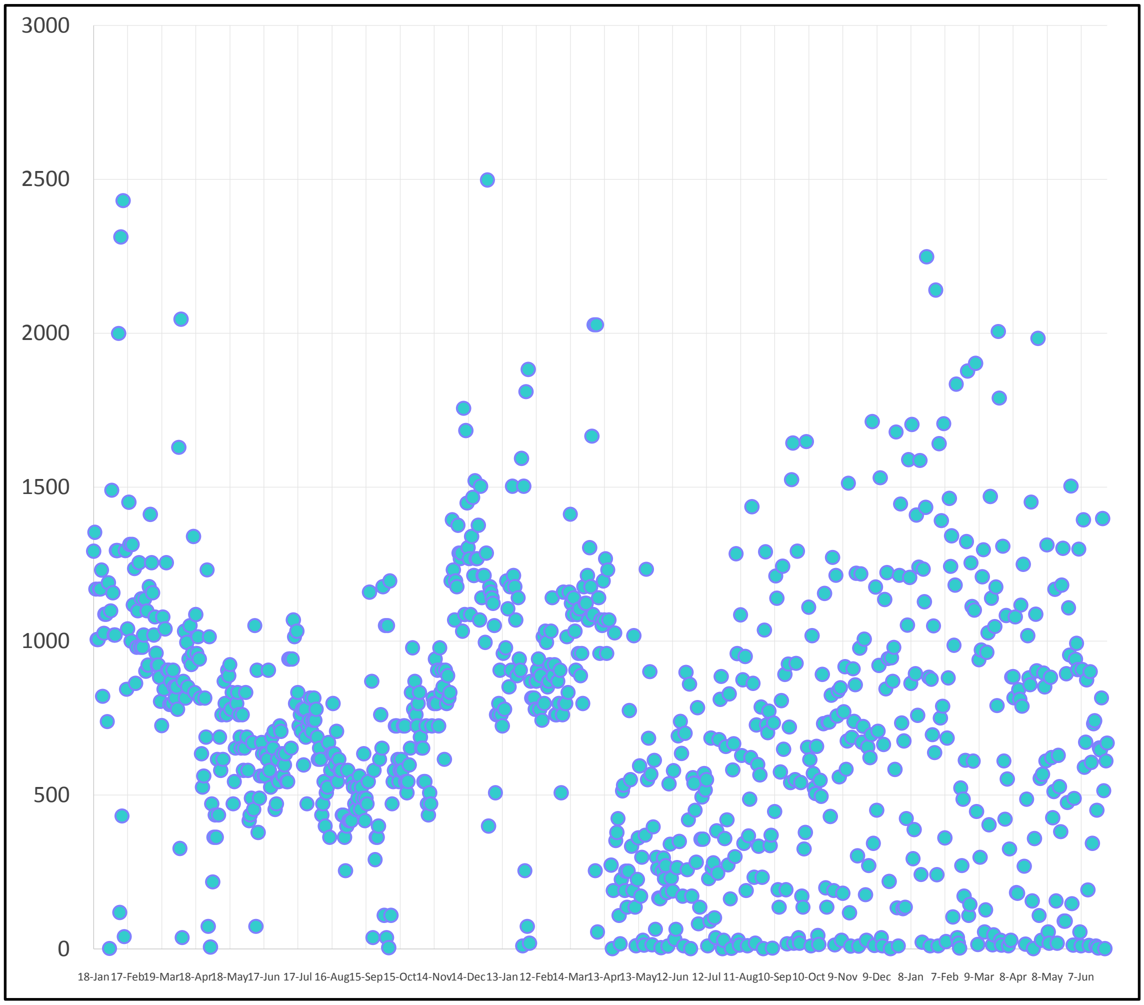
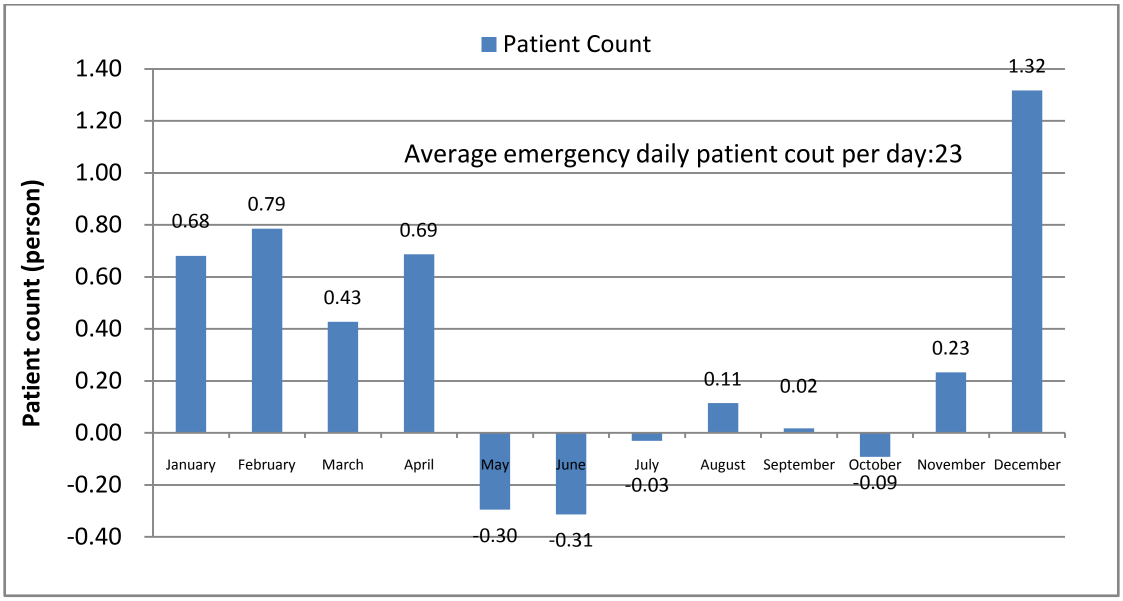


Figure 8: Impact of Month (difference of daily patient count or medicine expense caused by month compared to January)



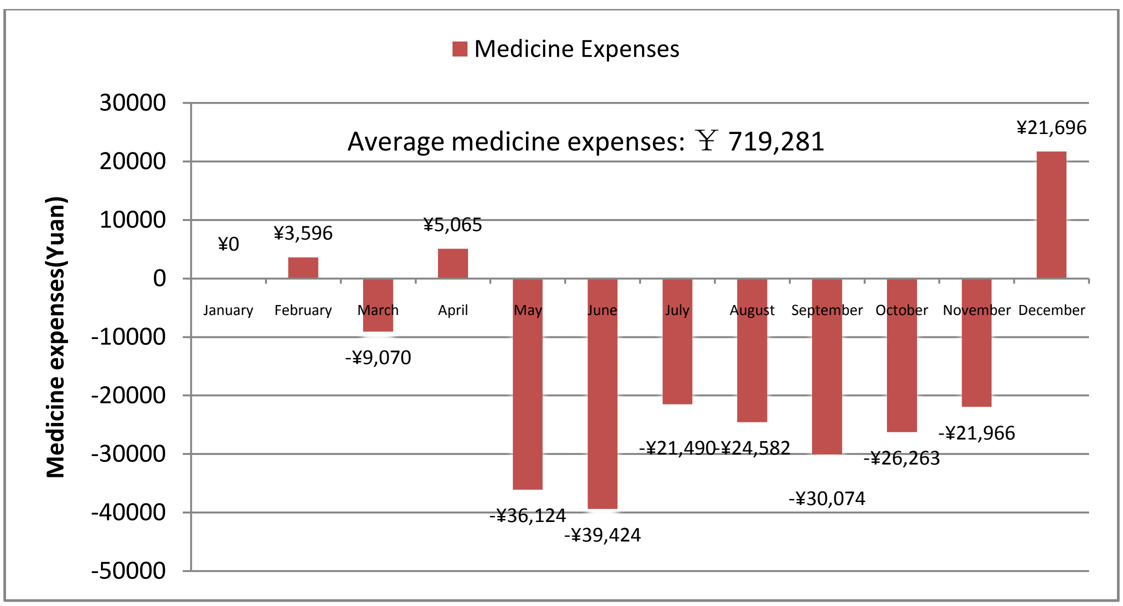
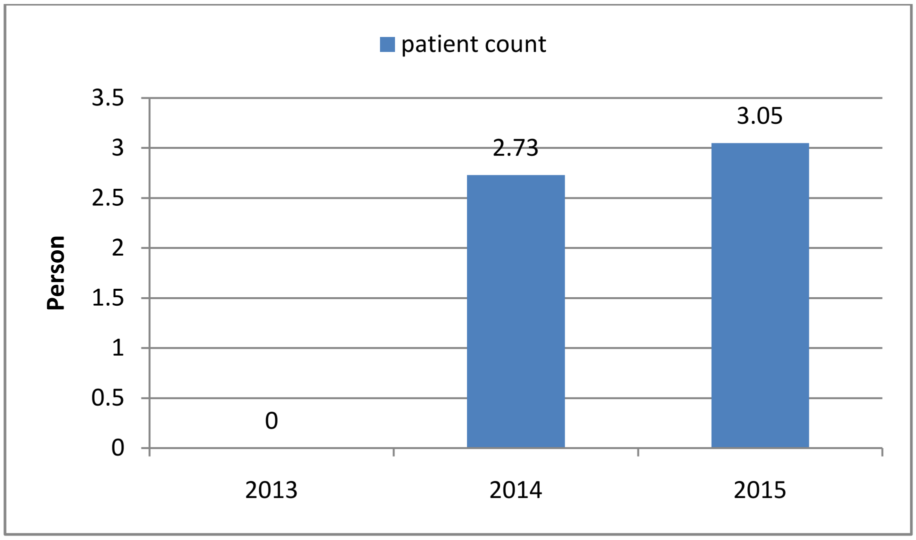


Figure 9: Impact of Year (difference of daily patient count or medicine expense caused by year compared to year 2013)



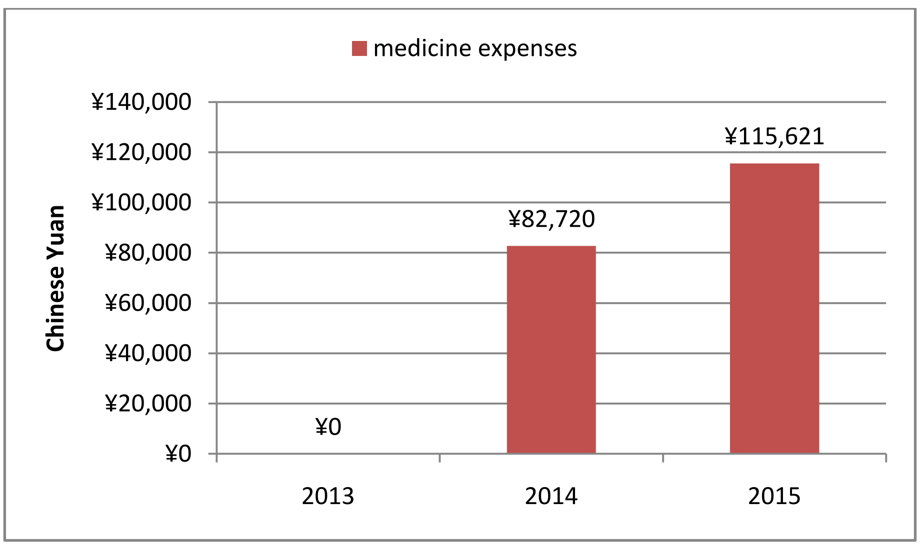


Figure 10: PM2.5 Impact on Respiratory and Cardiovascular Related Diseases (10㎍/m3 more PM2.5 level)

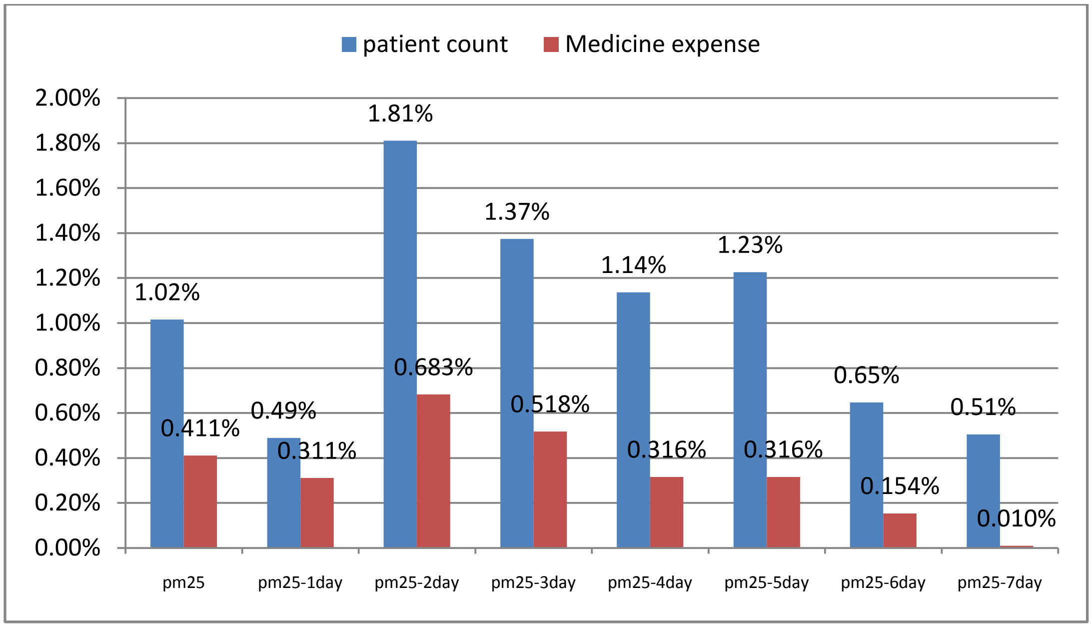
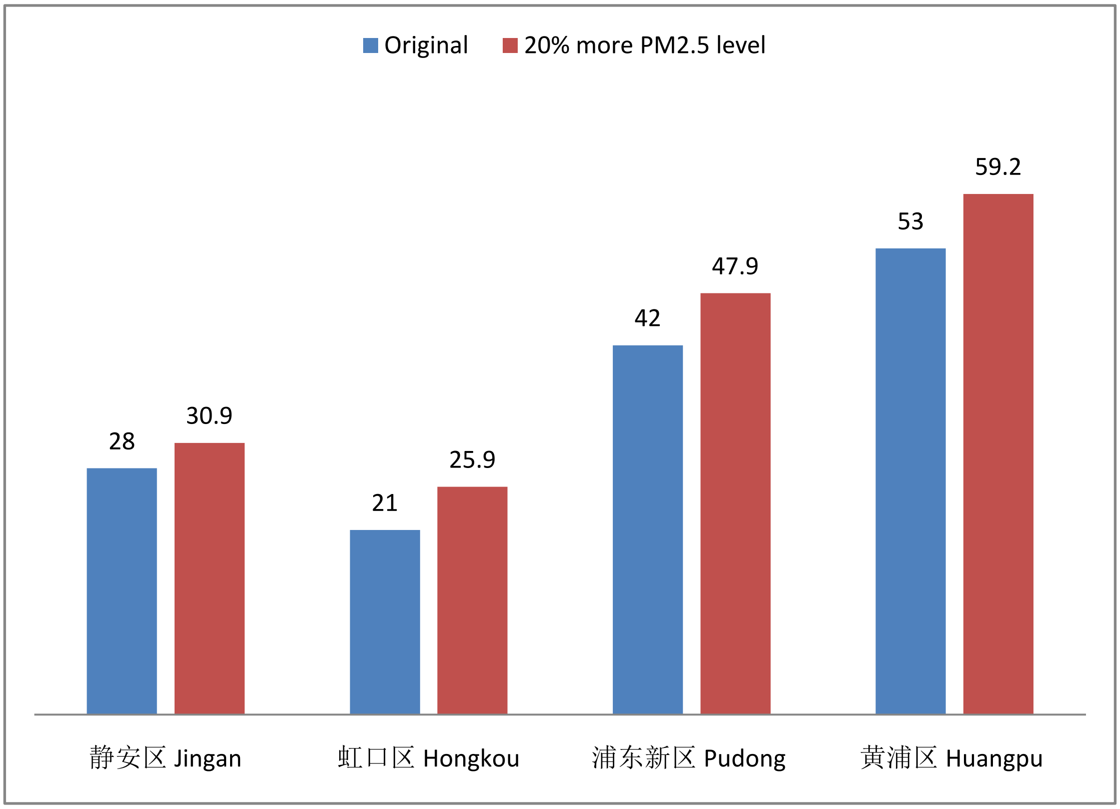


Figure 11: Impact analysis of PM2.5 level using 2SLS model



Appendix: Tables

Table 1. Summary Statistics of Weather Condition in Shanghai, 2013-2015

	Temperature(0.1 celsius)	Humidity(%)	Rainvolume(mm)	Windv(0.1m/s)	Sunh(0.1hour)
Mean	169.33	71.02	1973.21	27.14	46.44
Median	176	72	0	26	49
Standard Deviation	86.31	12.79	7710.27	9.56	39.14
Range	359	65	32700	61	127
Minimum	-9	31	0	6	0
Maximum	350	96	32700	67	127
Count	879	879	879	879	879

Note: the data comes from National Meteorological Information Center.

Table2. Summary Statistics of PM2.5 and patient count in Shanghai, 2013-2015

	$PM_{2.5}$ ($\mu g/m^3$)					Respiratory Diseases Patient Count				
District	Hongkou	Huangpu	Jingan	Pudong	All	Hongkou	Huangpu	Jingan	Pudong	All
Mean	56.19	57.49	58.36	53.26	57.32	18.96	55.18	20.53	42.28	31.83
Median	45.74	47.92	47.98	44.32	47.51	17	50	18	23	23
Standard Deviation	40.42	40.07	42.01	41.05	40.80	9.55	24.42	9.93	18.21	23.44
Range	459.42	433.80	495.18	416.23	495.18	102	202	68	156	213
Minimum	11.67	11.20	10.60	9.60	9.60	4	15	4	5	4
Maximum	471.09	445.00	505.78	468.02	505.78	106	217	72	225	225
Count	879	875	806	855	3415	879	875	806	855	3415

Note: the air pollution data come from China's Ministry of Environment and the emergency admission data come from the four hospitals under study.

Table 3: Models with only Contemporaneous Effect

	Model 1		Model 2		Model 3		Model 4	
Variables	Beta	t-statistic	Beta	t-statistic	Beta	t-statistic	Beta	t-statistic
pm25	0.019	1.325	0.019	2.036	0.019	2.388	0.021	2.507
temperature	-0.061	-10.142	-0.058	-14.586	-0.058	-14.626	-0.056	-13.894
humidity	0.051	1.021	0.041	1.240	0.043	1.340	0.044	1.407
rainvolume	0.000	0.207	0.000	0.168	0.000	0.112	0.000	0.106
windv	-0.003	-0.060	-0.002	-0.073	-0.004	-0.079	-0.003	-0.075
sunh	0.034	2.158	0.028	2.707	0.029	2.766	0.030	2.628
2014	2.607	2.466	2.734	3.893	2.724	3.926	2.634	3.730
2015	2.702	2.119	3.087	3.640	3.053	3.637	2.987	3.819
constant	30.641	6.499	-1.755	-3.484	-1.884	-3.776	-1.956	-3.964
year dummy	yes	yes	yes	yes	yes	yes	yes	yes
district fixed effect	no	no	yes	yes	yes	yes	yes	yes
month fixed effect	no	no	no	no	yes	yes	yes	yes
day of week fixed effect	no	no	no	no	no	no	yes	yes
Observation	2550	2550	2550	2550	2550	2550	2550	2550
rsquare	0.0814	0.0814	0.1601	0.1601	0.1689	0.1689	0.1691	0.1691

Note: the dependent variable is patient count data . All the regressions are estimated using OLS.

Table 4: Model with Dynamic Effects

Variables	OLS 7-day model 4 Beta	t-statistic	Variables	OLS 3-week model 4 Beta	t-statistic
pm25	0.019	3.206	pm25	0.016	1.325
pm25-1day	0.009	2.044	pm25-1week	0.024	2.053
pm25-2day	0.034	4.830	pm25-2week	0.013	1.454
pm25-3day	0.026	3.169	pm25-3week	0.009	1.954
pm25-4day	0.022	2.918	temperature	-0.061	-10.142
pm25-5day	0.023	2.197	humidity	0.051	1.021
pm25-6day	0.012	1.292	rainvolume	0.000	0.207
pm25-7day	0.010	1.253	windv	-0.003	-0.060
temperature	-0.083	-13.514	sunh	0.034	2.158
Humidity	0.042	1.885	2014	2.607	2.466
Rainvolume	0.000	0.113	2015	2.702	2.119
Windv	-0.053	-0.081	constant	-3.641	6.499
Sunh	0.029	2.929	year dummy	yes	yes
2014	2.765	2.435	district fixed effect	yes	yes
2015	3.222	1.992	day of week fixed effect	yes	yes
constant	-5.884	6.609	month fixed effect	yes	yes
year dummy	yes	yes	*	*	*
day of week fixed effect	yes	yes	*	*	*
district fixed effect	yes	yes	*	*	*
month fixed effect	yes	yes	*	*	*
Observation	2550	2550	Observation	2550	2550
R^2	0.1723	0.1723	R^2	0.0814	0.0814

* p value<0.05 if |t-statistic|>1.96

Note: the dependent variable is patient count data . All the regressions are estimated using OLS.

Table 5: 2SLS Model Estimation Result with Dynamic Feature

Variables	2sls Model with Dynamic Features		First Stage Estimation result		
	Second Stage Beta	t-statistic	Traffic volume Variables	First stage IV beta	F-statistics
pm25	0.0238	2.8891	Trafficvolume	0.0034	10.326
pm25-1	0.0121	2.2665	Trafficvolume-1	0.0028	11.258
pm25-2	0.0328	4.8965	Trafficvolume-2	0.0032	9.651
pm25-3	0.0219	3.1234	Trafficvolume-3	0.0031	10.695
pm25-4	0.0201	2.9586	Trafficvolume-4	0.0035	9.663
pm25-5	0.0246	2.1651	Trafficvolume-5	0.0033	10.659
pm25-6	0.0122	1.4324	Trafficvolume-6	0.0032	11.216
pm25-7	0.0098	1.2352	Trafficvolume-7	0.0029	12.654
temperature	-0.1223	-14.9856	*	*	*
humidity	0.0412	1.6987	*	*	*
rainvolume	0	0.1254	*	*	*
windv	-0.0563	-0.0895	*	*	*
sunh	0.0269	2.8867	*	*	*
constant	-4.1655	-3.9116	*	*	*
day of week fixed effect	yes	yes	*	*	*
month fixed effect	yes	yes	*	*	*
district fixed effect	yes	yes	*	*	*
Observation	155	155	*	*	*
R^2	0.5498	0.5498	*	*	*
* p value<0.05 if |t-statistic|>1.96			* P value<0.0016 if f-statistic>10		

Note: the dependent variable is patient count data . All the regressions are estimated using OLS.

Table 6: Comparison between OLS model and 2SLS model

Variables	2sls Model with Dynamic Features Second Stage Beta	t-statistic	Variables	OLS 7-day model 4 Beta	t-statistic
pm25	0.0238	2.8891	pm25	0.019	3.206
pm25-1	0.0121	2.2665	pm25-1day	0.009	2.044
pm25-2	0.0328	4.8965	pm25-2day	0.034	4.830
pm25-3	0.0219	3.1234	pm25-3day	0.026	3.169
pm25-4	0.0201	2.9586	pm25-4day	0.022	2.918
pm25-5	0.0246	2.1651	pm25-5day	0.023	2.197
pm25-6	0.0122	1.4324	pm25-6day	0.012	1.292
pm25-7	0.0098	1.2352	pm25-7day	0.010	1.253
temperature	-0.1223	-14.9856	temperature	-0.083	-13.514
humidity	0.0412	1.6987	humidity	0.042	1.885
rainvolume	0	0.1254	rainvolume	0.000	0.113
windv	-0.0563	-0.0895	windv	-0.053	-0.081
sunh	0.0269	2.8867	sunh	0.029	2.929
*	*	*	2014	2.765	2.435
*	*	*	2015	3.222	1.992
*	*	*	constant	-5.884	6.609
constant	-4.1655	-3.9116	year dummy	yes	yes
day of week fixed effect	yes	yes	day of week fixed effect	yes	yes
month fixed effect	yes	yes	district fixed effect	yes	yes
district fixed effect	yes	yes	month fixed effect	yes	yes
Observation	155	155	Observation	2550	2550
R^2	0.5498	0.5498	R^2	0.1723	0.1723

* p value<0.05 if |t-statistic|>1.96   * p value<0.05 if |t-statistic|>1.96

Note: the dependent variable is patient count data . All the regressions are estimated using OLS.

Table 7: OLS model with dynamic feature using medical expense data in Hongkou district

	OLS 7-day model expense	
Variables	Beta	t-statistic
pm25	0.030	4.328
pm25-1day	0.022	2.453
pm25-2day	0.049	3.280
pm25-3day	0.037	3.802
pm25-4day	0.023	2.189
pm25-5day	0.023	1.623
pm25-6day	0.011	1.326
pm25-7day	0.001	1.234
temperature	-0.124	-13.132
humidity	0.056	2.698
rainvolume	0.000	0.236
windv	-0.053	-0.026
sunh	0.057	3.263
2014	3.652	2.695
2015	4.633	2.365
constant	-8.977	7.632
year dummy	yes	yes
day of week fixed effect	yes	yes
district fixed effect	yes	yes
month fixed effect	yes	yes
Observation	2550	2550
R^2	0.1963	0.1963

Note: the dependent variable is medicine expenses data . All the regressions are estimated using OLS.