Three Essays in Spatial Econometrics

by

Mingjin Xia

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

Department of Economics

University of Manitoba

Winnipeg

Abstract

This dissertation consists of three essays in the field of spatial econometrics. Incorporating spatial interaction effects into the model can improve the explanatory power of econometric models. Additionally, employing spatial econometrics is an effective strategy for addressing the spatial dependence of variables, which could introduce endogeneity. In Chapter 1, the impact of $PM_{2.5}$ on sleeplessness in Chinese cities is examined using novel Weibo data and spatial econometric models. Despite no direct effect of PM2.5 on sleeplessness, a significant negative impact is observed in adjacent cities during severe pollution in winter, indicating spatial spillover effects. Chapter 2 investigates the Real-Time Air Pollution Monitoring Policy's influence on corporate responsibility in China, revealing spatial spillover effects of the policy. While no direct treatment effect of the policy on CSR or CER scores within treated cities is found, a significant indirect treatment effect of the policy on untreated neighboring cities is noted, possibly due to supply chain alterations. Chapter 3 investigates the impact of the minimum wage on employment in Chinese cities, while accounting for endogeneity arising from spatial heterogeneity, and spatial dependencies among variables. The findings indicate a notable positive direct effect and a considerable negative indirect effect of the minimum wage on employment rates.

Acknowledgements

First, I would like to express my deepest gratitude to my supervisor Professor Umut Oguzoglu. Without his guidance and encouragement, this dissertation could not have been finished. I am also very thankful to Professor Julia Witt and Professor Soodeh Saberian for their invaluable advice and insightful suggestions.

Thanks are also extended to all other professors and colleagues who provide useful comments and suggestions on my thesis.

Finally, I am deeply thankful for my family and friends. It is their love and support that keeps me going.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Spatial econometrics encompasses a broad spectrum of economic studies, including real estate economics, environmental and agricultural economics, and industrial organization (Bockstael, 1996; Anselin & Rey, 1997; Anselin, 2003; Pinkse & Slade, 2010). Incorporating spatial interaction effects into the model enhances our comprehension of the spatial structure of phenomena, thereby improving the explanatory power of econometric models (Tosetti, Santos, Moscone, & Arbia, 2018). Additionally, employing spatial econometrics and spatial weighted matrix regression is an effective strategy for addressing the spatial dependence of omitted variables, which could introduce endogeneity (Zangger, 2019). Furthermore, spatial econometric models can analyze both the direct and indirect effects (spatial spillover) of dependent variables on independent variables (Zangger, 2019). Conventional models typically assume the indirect effect to be zero. Therefore, I employ spatial econometric models to investigate three questions based on Chinese data.

Chapter 1 assesses the influence of air pollution on sleeplessness in a panel of Chinese cities. To investigate the effect of air pollution on sleep, I collected daily posts containing keywords of sleeplessness from a popular Chinese social media platform, Weibo, between March 2018 and February 2019. Keywords of sleeplessness are "Shimian" and "Shuibuzhao," which means "sleepless," "cannot fall asleep," "losing sleep," or "insomnia". I integrated the daily city-level measure of Weibo users' sleeplessness with the daily average $PM_{2.5}$ concentrations and meteorological data including daily average temperature, humidity, wind speed, and precipitation), creating a comprehensive daily city-level dataset. Methodologically, this paper employs a spatial econometric model, specifically the spatial Durbin model (SDM), for the analysis. I also used winter data to analyze the effect of $PM_{2.5}$ on Weibo users' sleeplessness, controlling for city, weekend, and time fixed effects, and meteorological factors.

Two main contributions are made in Chapter 1. Firstly, a novel measure of sleeplessness at the city and daily level was constructed by collecting posts containing keywords related to sleeplessness from Weibo. In contrast to sleeplessness measures generated by questionnaires and surveys (Costa-i Font, Fleche, & Pagan, 2021; Dickinson, McEvoy, & Bruner, 2022), this novel measure is easier to implement by collecting social media posts containing reference to sleeplessness than by conducting surveys to obtain a sample. It is also superior to survey data, which may not be representative of the target population (Vaske, 2011). Secondly, utilizing spatial econometrics regression is a robust approach for addressing spatial dependence among variables and mitigating the endogeneity issues caused by omitted variables.

Chapter 2 studies the impact of the real-time air pollution monitoring (RAPM) policy on corporate social responsibility (CSR) and corporate environmental responsibility (CER) of listed companies in China. Ashenfelter and Card (1984) emphasized that the standard Difference-in-Differences (DID) model serves as a benchmark to estimate the treatment effect of a policy. However, Delgado and Florax (2015) argue that the spatial correlations among socioeconomic variables result in the violation of one of the assumptions of the DID model, namely the Stable Unit Treatment Value Assumption. The DID estimation becomes biased and inefficient as it neglects spatial correlations of treatment and variables in different cities. They suggest that a spatial DID (SDID) model can be employed to control for spatial correlations in treatments and social interactions in treatment responses. Hence, I employs standard DID, Spatial DID, decomposed SDID (DSDID), Spatial Durbin Error Term DID (SDEM-DID), and decomposed SDEM-DID (DSDEM-DID) models, incorporating two distinct spatial weighted matrices (800 kilometers (KM) binary contiguity and 800KM inverse distance spatial weighted matrices), to analyze the treatment effects of the RAPM policy on city-level average CSR and CER scores from 2010 to 2016 across a panel of Chinese cities. Implementation of the RAMP policy commenced in 74 pilot cities at the beginning of 2013 in China.

This chapter makes two significant contributions. First, it conducts the analysis at the city level. While much of the existing literature has primarily focused on the firm level in the study of CSR and CER, this paper recognizes the significance of cities as central hubs for economic and social activities. Consequently, there is a growing importance in investigating CSR and CER at the city level (Y.-C. Chen, Hung, & Wang, 2018; G. Zhang, 2023). Second, this chapter employs spatial econometric models to account for spatial correlations in treatment, variables, and error terms. Spatial econometric models help mitigate endogeneity issues arising from the violation of the SUTVA and allow for the analysis of the policy's average direct and indirect treatment effects (ADTE and AITE) on treated cities and neighboring cities of treated cities.

In Chapter 3, I utilize city-level data and the Spatial Durbin Model (SDM) to examine the impact of the minimum wage on the employment rate. This analysis involves city-level panel data encompassing 263 Chinese cities from 2004 to 2011. All variables are transformed into logarithmic form, enabling the assessment of the minimum wage elasticity of the employment rate. The SDM model proves effective in addressing the spatial dependence of variables (Zangger, 2019). Failure to account for the spatial dependence of variables could introduce bias into the estimations. In China, the spatial dependence of minimum wage and employment arises from competition between cities regarding minimum wage levels and the migration of the labor force.

Chapter 2

Can Air Pollution Increase Sleeplessness? A Spatial Analysis from a Panel of Chinese Cities

Abstract

This paper employs spatial panel econometric methods to investigate the effect of air pollution on sleeplessness in a panel of Chinese cities. Declining air quality is an increasing problem in China, and it is associated with poor sleep, which may lead to mental health problems, traffic accidents, and lower productivity. However, unobserved spatial factors make the identification difficult, and traditional measures of sleeplessness based on survey data are often inadequate. Following Heyes and Zhu (2019), I used daily posts from the Chinese social media Weibo to capture sleeplessness in a city, and then employ various spatial econometric models to control for endogeneity stemming from unobserved spatial dependence. I find that the pollutant $PM_{2.5}$ displays both global and local spatial autocorrelations, underscoring the need for spatial methods. Empirical results based on data from all seasons, controlling for city, weekend, and time fixed effects and weather factors, suggest that $PM_{2.5}$ has no direct or indirect effect on sleeplessness. However, during the winter months when pollution is the most severe, $PM_{2.5}$ in a city has a significant negative effect on sleeplessness in the city's adjacent cities. Quantitatively, if $PM_{2.5}$ increases by 1 unit in the winter, the number of sleeplessness posts in that city may not increase, but the number of sleeplessness posts in neighboring cities will increase by 1.930.

2.1 Introduction

Sleep is a pivotal determinant of human well-being, health, and societal functioning. Inadequate sleep can result in reduced productivity, diminished labor effectiveness, and increased interpersonal conflict. Furthermore, it is linked to heightened mortality risks associated with fatal car accidents, strokes, cancer, or cardiovascular diseases (Gibson & Shrader, 2018; Hafner, Stepanek, Taylor, Troxel, & Van Stolk, 2017; Smith, 2016; Dickinson et al., 2022). Additionally, air pollution has a detrimental impact on the quality of human sleep (Heyes & Zhu, 2019; Strøm-Tejsen, Zukowska, Wargocki, & Wyon, 2016; Lawrence et al., 2018; Billings et al., 2019).

There are two channels that explain how air pollution can negatively affect human sleep quality. Firstly, air pollution can inflame the human respiratory tract and nervous system, increase the prevalence of allergic symptoms, and contribute to chronic diseases. These factors can subsequently lead to sleep-disordered breathing (SDB) (Zanobetti et al., 2010; Sánchez et al., 2019; Nuvolone, Petri, & Voller, 2018; Calderón-Garcidueñas, Leray, Heydarpour, Torres-Jardón, & Reis, 2016; Gilles et al., 2018; Patella et al., 2018; Akinseye et al., 2015; Y. Sun et al., 2018). Secondly, heavy air pollution might impede residents' outdoor actives, consequently escalating feelings of depression and anxiety, ultimately contributing to increase sleeplessness (Hartescu, Morgan, & Stevinson, 2015; B. Zheng et al., 2017; He, Luo, & Zhang, 2022).

Empirical and data barriers influence studying the effects of air pollution on sleep encounters and are evident in the exist literature. Empirical challenges arise due to spatial dependence of air pollution[1] and potential omitted variables, leading to endogeneity issues (Fu et al., 2021; Zangger, 2019). In cases where omitted variables exhibit spatial dependence, many studies adopt a strategy of utilizing variables from neighboring cities as instrumental variables[2] (Zangger, 2019). Unfortunately, finding effective instrumental variables is challenging (Fingleton & Le Gallo, 2010) and these instruments face limitations[3]. Collecting data on sleeplessness is inherently challenging, often leading studies to rely on questionnaires[4] for participants' self-

[1] Fu, Viard, and Zhang (2021) argued that a city's air pollution can have a spatial spillover effect on neighboring cities due to wind currents, this is known as the spatial dependence of air pollution.

[2] For instance, wind speed, wind direction, and air pollution levels of the target city's neighboring cities are used to create an exogenous air pollution instrumental variable for the target city (Heyes & Zhu, 2019; Williams & Phaneuf, 2019).

[3] Using neighboring cities' variables as instruments may not adequately capture the economic or demographic characteristics of a specific city (Fowler, Frey, Folch, Nagle, & Spielman, 2020). Additionally, even if these instruments consider spatial diffusion, they may fail to estimate spatial spillover effects between neighboring cities and the target city. Behaviors between these areas could mutually influence each other (Chaix, Merlo, Subramanian, Lynch, & Chauvin, 2005; Xu, 2014; Zangger, 2019).

[4] Responses are used to score a sleep quality index, such as the Pittsburgh Sleep Quality In-

reported sleep quality. The survey method takes time, is costly, and inconvenient, and it may not be representative of the target population (Vaske, 2011).

To investigate the effect of air pollution on sleep, I collected daily posts containing keywords of sleeplessness[5] from a popular Chinese social media platform, Weibo, between March 2018 and February 2019. I integrated the daily city-level measure of Weibo users[6] sleeplessness with the daily average $PM_{2.5}$ concentrations and meteorological data including daily average temperature, humidity, wind speed, and precipitation), creating a comprehensive daily city-level dataset. Methodologically, this paper employs a spatial econometric model, specifically the spatial Durbin model (SDM), for the analysis. I also used winter[7] data to analyze the effect of $PM_{2.5}$ on Weibo users' sleeplessness, controlling for city, weekend, and time fixed effects, and meteorological factors. Because in winter, China experiences a significant increase in air pollution, particularly in the northern regions, due to the winter heating system[8], which results in higher concentrations of $PM_{2.5}$ during winter compared to other seasons (Cai, Nan, Zhao, Jiao, & Pan, 2020; Fan et al., 2020). Additionally, the prevailing wind direction during winter is from the north to the south (X. Chen & Ye, 2019), whereas it is reversed in the summer. This northern wind direction in winter has the potential to carry $PM_{2.5}$ from northern regions to the south, establishing a spatial dependence of $PM_{2.5}$ between cities.

dex (PSQI) induced by Buysse, Reynolds III, Monk, Berman, and Kupfer (1989), or simply the arithmetic mean of sleep scores to measure sleep quality (Costa-i Font et al., 2021; Amez, Vujic, Abrath, & Baert, 2021; Dickinson et al., 2022).

[5] Keywords of sleeplessness are "Shimian" and "Shuibuzhao," which means "sleepless," "cannot fall asleep," "losing sleep," or "insomnia".

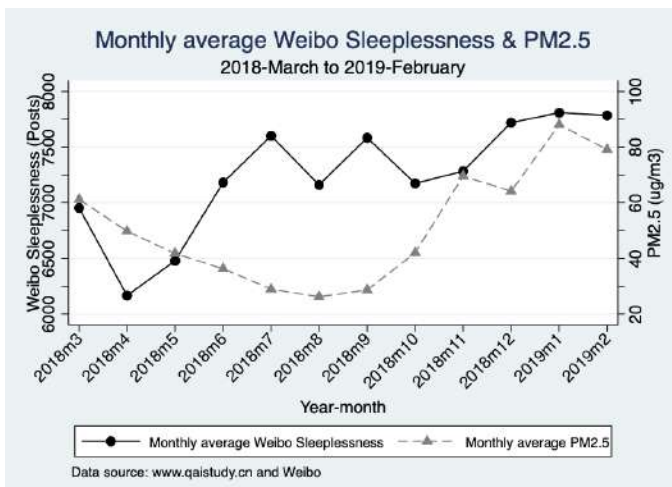
[6] Weibo users might not represent the general Chinese population, the adverse effects of air pollution on Weibo users could extend to non-users (Heyes & Zhu, 2019). Weibo users are, generally younger, more educated, and with higher incomes than the general population (Chiu, Lin, & Silverman, 2012; Y. Han, Li, Xiao, Li, & Zhu, 2021), and more likely to afford air purifiers, potentially reducing the effects of air pollution. Younger Weibo users may also be less vulnerable to air pollution compared to elderly individuals. If air pollution affects the population of Weibo users, non-users might experience the negative effects to a greater extent.

[7] Between November 2018 and February 2019.

[8] China's winter heating system is implemented in the northern regions based on a demarcation line known as the Qingling-Huaihe line, represented by the red line in Figure 1.2. This heating network utilizes water pipelines that connect boilers to the radiators of homes and offices (Fan, He, & Zhou, 2020). The system relies on inefficient and coal-based boilers, which are provided by the local government either at zero price or heavily subsidized. Consequently, residents are only required to pay a low fee for heating services. Typically, the heating season spans from November 15th to March 15th in most northern Chinese cities, extending from October to April in certain cities experiencing particularly cold winters.

Two main contributions are made in this paper. Firstly, a novel measure of sleeplessness at the city and daily level was constructed by collecting posts containing keywords related to sleeplessness from Weibo. In contrast to sleeplessness measures generated by questionnaires and surveys (Costa-i Font et al., 2021; Dickinson et al., 2022), this novel measure is easier to implement by collecting social media posts containing reference to sleeplessness than by conducting surveys to obtain a sample. It is also superior to survey data, which may not be representative of the target population (Vaske, 2011). The measurement of sleeplessness in this paper bears resemblance to that employed by Heyes and Zhu (2019). However, this study encompasses a more extensive array of cities within a given year. This broader coverage is aimed at strengthening control over the varying levels of $PM_{2.5}$ across provinces, considering seasonal changes and external effects [9].

Figure 2.1: Variations of monthly average Weibo users sleeplessness and $PM_{2.5}$



[9]For example, centralized winter heating system, the provinces containing 56 cities included in this study display notable variations in the adoption of centralized heating.

Figure 2.2: Winter average $PM_{2.5}$ and sleeplessness maps

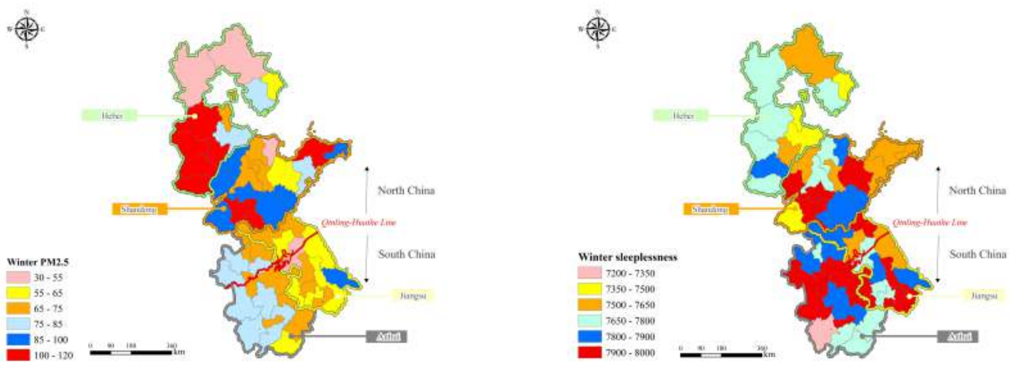


Figure 1.1 illustrates that $PM_{2.5}$ is more severe in winter, primarily due to the use of coal in winter heating system (Cai et al., 2020). In Figure 1.2, the monthly average $PM_{2.5}$ and sleeplessness in winter are presented. The concentration of $PM_{2.5}$ is higher in northern China compared to southern China. Interestingly, despite the higher $PM_{2.5}$ levels in the north, the sleeplessness population of Weibo users in south China is higher. This discrepancy could be attributed to the movement of $PM_{2.5}$ carried by wind from northern China to southern China (X. Chen & Ye, 2019).

Secondly, utilizing spatial econometrics regression in conjunction with a binary contiguity spatial weighted matrix[10] is a robust approach for addressing spatial dependence among variables and mitigating the endogeneity issues caused by omitted variables. This method can also overcome the limitations associated with instrumental variables, as discussed by Chaix et al. (2005), Fowler et al. (2020), Zangger (2019), and Xu (2014). The spatial weighted matrix is particularly valuable in capturing the magnitude of spatial spillover effects between neighboring cities.

The paper highlights four primary findings. First, the global and local spatial Moran's I tests reveal significant spatial autocorrelation in $PM_{2.5}$, underscoring the justification for utilizing spatial models. The detailed results of these tests for monthly average $PM_{2.5}$ can be found in Table 1.3, and a visual representation is provided in Figure 1.3 (located in the Appendix). Notably, the paper's results demonstrate a higher level of robustness compared to prior studies, attributed to its

[10]The elements in the matrix are set to 1 if two cities share a common border and 0 otherwise. Detailed explanations can be found in Section 4.

effective handling of spatial dependence among $PM_{2.5}$ levels in various cities.

The second finding underscores a significant indirect effect of $PM_{2.5}$ on sleeplessness. Specifically, a one-unit increase in $PM_{2.5}$ within a city results in an impactful increase in sleeplessness among Weibo users in neighboring cities by three individuals. Notably, this effect is observed before accounting for province-by-season fixed effects and meteorological factors. This contrasts with prior studies (Heyes & Zhu, 2019; Strøm-Tejsen et al., 2016; Lawrence et al., 2018; Billings et al., 2019), which predominantly focused on the direct effects of $PM_{2.5}$ on sleep, so this paper provides additional empirical evidence on the spatial spillover effects of $PM_{2.5}$ on individual sleeplessness. Importantly, in this study, the direct effect of $PM_{2.5}$ on sleeplessness within a city is found to be statistically insignificant.

The third finding is the estimation using winter data and controlling meteorological factors reveals that a statistically significant positive indirect effect of $PM_{2.5}$ on the sleeplessness of Weibo users. Additionally, an insignificant direct effect estimates of $PM_{2.5}$ on the sleeplessness of Weibo users is observed. Quantitatively, a one-unit increase in $PM_{2.5}$ in a city leads to a two-person increase in the sleeplessness of Weibo users in that city's neighboring cities. However, the sleeplessness of Weibo users in that city itself does not appear to be affected.

The fourth finding suggests that the estimates from the Spatial Durbin Model (SDM) is more accurate than those from the Spatial Error Model[11] (SEM) and the Spatial Autoregressive Model[12] (SAR). The Lagrange Multiplier (LM) tests, as detailed in Table 1.5, consistently support the SDM model.

Different spatial weight matrices are used for robustness check. The results show the direct and indirect effects of $PM_{2.5}$ on sleeplessness are insignificant during the full-season and significant in the winter, respectively, thus supporting findings in this paper.

Lastly, an examination was conducted to assess the impact of AQI[13] and addi-

[11] A drawback of SEM model is it lacks the ability to account for spatial spillover effects, because it solely controls for spatial dependence in error terms (Halleck Vega & Elhorst, 2015).

[12] The drawbacks of the SAR model include a consistent ratio between direct and indirect effects for all variables and the inability to discern whether a significant indirect effect is due to omitted spatially correlated explanatory variables or spatially correlated dependent variables (Halleck Vega & Elhorst, 2015).

[13] In China, the Air Quality Index (AQI) provides a comprehensive measure of a city's air quality.

tional air pollutants including PM_{10}, SO_2, CO, and O_3 on Weibo users' sleeplessness during winter. According to the results from the SDM model, both AQI and PM_{10} exhibit a noteworthy positive indirect effect on the sleeplessness of Weibo users but their influences are relatively smaller compared to the impact of $PM_{2.5}$.

The remainder of the paper is organized as follows: Section two reviews the relevant literature. Section three describe the data that were used and the empirical methodology. Section four presents and discusses the results. Section five contains my conclusions.

2.2 Related Literature

This section summarizes the existing literature about the effect of air pollution on sleep and spatial econometric models.

The effect of air pollution on sleep

In the study by Heyes and Zhu (2019), the researchers collected posts containing keywords related to sleeplessness from Weibo, constructing a daily city-level panel dataset. Their dataset encompassed 19 first-tier Chinese cities over a 2-year time span. To address endogeneity concerns, they created an instrumental variable for air pollution based on wind direction and wind speed. The findings suggested that a one-standard-deviation increase in the air quality index or $PM_{2.5}$ led to a respective increase of 11.6% and 12.8% in sleeplessness.

In the research conducted by Strøm-Tejsen et al. (2016), two field intervention experiments were employed in student dormitory rooms to assess the relationship between air quality (measured by CO_2 levels) and sleep quality, as well as next-day performance. The manipulation of electric heaters and bedroom ventilation led to varying CO_2 levels in different rooms. Mathematical analysis of the experimental results revealed that sleep quality and next-day performance improved when CO_2 levels were lower or when bedroom air quality was enhanced.

Using children's daily sleep quality based on cross-sectional survey data from 2012 and the average daily air pollution index in seven northeastern Chinese cities, Lawrence et al. (2018) demonstrated that air pollutants PM_1 and $PM_{2.5}$ (measured

A lower AQI indicates better air quality (Yang & Zhong, 2022).

in ug/m^3) have a more detrimental effect on children's sleep quality than other air pollutants such as SO_2, NO_2, O_3, and CO. If PM_1 and $PM_{2.5}$ were to increase by 11.6 ug/m^3, females' odds of sleep disorder would rise by 19% to 49% and 18% to 44%, respectively. For males, the odds of sleep disorder would increase by 28% to 47% and 26% to 42%, respectively.

After accounting for demographic and socioeconomic factors, Billings et al. (2019) investigated the long-term exposure of seniors (aged 59 to 77 years) to 1- and 5-year measures of air pollutants $PM_{2.5}$ and NO_2 in six U.S. cities. The findings revealed that with a 5 ug/m^3 increase in $PM_{2.5}$ exposure, the odds of sleep apnea would rise by 60%. In essence, higher $PM_{2.5}$ levels posed a greater risk to the elderly during sleep.

Spatial econometric model

Some studies consider the spatial dependence of air pollution and employ spatial econometric methods to examine the adverse effects of air pollution on individual health. X. Chen, Shao, Tian, Xie, and Yin (2017) constructed a Chinese city-level panel dataset comprising 116 cities over a span of 7 years to utilize the Spatial Durbin Model (SDM) to conduct their analysis. Their findings suggested that a direct effect of every ten thousand tons of sulfur dioxide emissions increased local lung cancer mortality and respiratory disease mortality by 350 and 300 people, respectively. Additionally, the indirect effect was observed to increase lung cancer mortality and respiratory disease mortality to 2,170 and 15,430 people in neighboring cities, respectively. This indicates a significant spatial spillover effect of air pollution, wherein an increase in sulfur dioxide emissions in one city could lead to higher mortality rates in neighboring cities than in the city itself. In this paper, the researchers utilized the inverse distance matrix for all cities, indicating that they did not set a distance threshold for a city's neighbors. This implies that a city's neighboring cities encompass all the other cities.

In the study, Feng, Cheng, Shen, and Sun (2019) utilized Chinese provincial panel data spanning from 2004 to 2013 to perform a spatial panel empirical analysis, demonstrating the significant adverse impact of air pollution on public health. Their findings indicated that a 1% increase in $PM_{2.5}$ was associated with a 0.38% increase in the number of people treated in hospitals per ten thousand. Notably, this negative effect of $PM_{2.5}$ on individual health was more pronounced when considering the spatial dependence of public health, emphasizing the importance of accounting for spatial factors in understanding the impact of air pollution.

The spatial dependence of air pollution has been a focus of several studies. In a cross-sectional analysis of 152 Chinese cities, Ma, Ji, and Fan (2016) explored the impact of spatial and economic factors on $PM_{2.5}$ pollution. They found that urbanization, population density, vehicle quantity, energy intensity, and refined oil price influenced $PM_{2.5}$ levels, with higher values of these factors leading to increased $PM_{2.5}$. Additionally, the spatial clustering of these factors contributed to spatial spillover effects, indicating that an increase in $PM_{2.5}$ in one city could affect neighboring cities.

In a study on the spatial correlation of firm-level CO_2 emissions, Cole, Elliott, Okubo, and Zhou (2013) used Japanese firm-level cross-sectional data in 2006 and spatial econometric methods. Their findings revealed that a firm's CO_2 emissions are spatially correlated, and various factors such as capital-to-labor ratio, size, research and development expenditures, and advertising expenditure influenced these emissions. The estimated coefficients considering spatial correlation were higher than those obtained without considering spatial effects, emphasizing the importance of accounting for spatial correlations to avoid biased estimates.

Examining the impact of foreign direct investment on air pollution in China, Tang, Li, and Yang (2016) utilized SAR and SEM models based on provincial-level panel data from 2000 to 2012. They concluded that a 1 percent increase in foreign direct investment led to a 0.0235 percent increase in air pollution. Moreover, their study confirmed significant spatial dependence of air pollution between provinces in China.

Compared to the above papers, this study used the spatial Durbin model and city- and daily-level data to investigate the effect of $PM_{2.5}$ on sleeplessness.

2.3 Data

This study explores the impact of the air pollutant $PM_{2.5}$ on sleep in 56 cities across mainland China, situated in four provinces: Anhui, Hebei, Jiangsu, and Shandong.

Notably, Hebei and Shandong have winter heating system during winter, whereas Anhui and Jiangsu do not. Consequently, the effects of air pollution on residents' health vary among these provinces. Residents in provinces with public heating systems are exposed to more severe air pollution during winter, primarily attributed to coal burning (Y. Chen, Ebenstein, Greenstone, & Li, 2013).

Following Heyes and Zhu (2019), a daily, city-level measure of sleeplessness data was obtained from Weibo social media posts. This data was then merged with the average daily air pollution data and meteorological data to create a daily panel dataset covering the period from March 2018 to February 2019. The dataset encompasses a total of 20,440 observations.

Sleeplessness

In the study by Heyes and Zhu (2019), it is argued that the sleeplessness index derived from keywords, specifically "Shimian" and "Shuibuzhao," serves as an appropriate measure. These Chinese terms are considered equivalent to English expressions such as "sleepless," "cannot fall asleep," "losing sleep," and "insomnia," among others. The use of Chinese characters, being glyphs and affirmative in meaning, helps reduce duplication and complications associated with expressing common concepts related to sleeplessness in alphabetic languages like English. The authors assert that while there may be other terms to express insomnia, the focus on these specific keywords does not introduce bias but rather enhances the efficiency of the estimation. Therefore, extracting daily nightly post counts (from 23:00 to 7:00) that include the keywords "Shimian" and "Shuibuzhao" from the widely used Chinese social media platform Weibo is considered a meaningful index for sleeplessness[14].

Air pollution

This research is centered on examining the impact of the air pollutant $PM_{2.5}$ on sleeplessness. $PM_{2.5}$ is chosen for investigation due to its capacity to travel longer distances and disperse over broader geographical areas compared to other pollutants like CO, SO_2, and NO_x (S. Chen, Guo, & Huang, 2018; Chan & Yao, 2008). Moreover, it can penetrate into human lung easier than others (W. Wang et al., 2022). Real-time air pollution data for $PM_{2.5}$ is sourced from the website www.qaistudy.cn, which aligns with the daily average pollution data provided by the

[14] The specific procedure for extracting Weibo sleeplessness data is provided in the Appendix.

Chinese Ministry of Environmental Protection (EMP). The unit of measurement for $PM_{2.5}$ is expressed in micrograms per cubic meter of air (ug/m^3). The data on $PM_{2.5}$ levels is collected from monitors distributed across various locations within a city, with different numbers of monitors in each city. The measurement of $PM_{2.5}$[15] levels for a city is obtained by calculating the simple arithmetic average across all monitors within that city.

Meteorological factors

Considering the impact of meteorological conditions on sleep quality is crucial. Meteorological factors, including temperature, humidity, wind speed, and precipitation, have been shown to affect individuals' sleep quality (G. Zheng, Li, & Wang, 2019; Obradovich, Migliorini, Mednick, & Fowler, 2017; Doherty, Youn, Haltiner, & Watson, 2010). For example, temperatures exceeding 36°C can lead to a reduction in sleep duration and increased difficulty in falling asleep (G. Zheng et al., 2019).

Daily average temperature (°C), humidity (%), wind speed (km/h), and precipitation (mm) are included as moderating variables to consider the influence of weather conditions on sleep. The weather data were sourced from stations registered by the World Meteorological Organization (WMO) and collected by the U.S.-based National Oceanic and Atmospheric Administration (NOAA).

---

[15] The measurement of $PM_{2.5}$ may be subject to errors due to variations in the number and locations of monitors used for detection in different cities (Heyes & Zhu, 2019). This implies that the severity of air pollution within a city can differ across various areas. Such variations could introduce attenuations in the estimations of the models. To address this concern, Heyes and Zhu (2019) computed the correlation of $PM_{2.5}$ levels between monitors at different locations in Beijing. Their findings revealed a high correlation in $PM_{2.5}$ levels among monitors at different locations, exceeding 0.9. This high correlation helps mitigate the measurement error associated with $PM_{2.5}$ levels.

Table 2.1: Summary statistics

Variable	Observations	Mean	Std. dev	Min	Max
Weibo posts	20440	7238.474	978.915	392	9868
$PM_{2.5}$	20440	51.140	38.026	4.091	400.833
Humidity	20440	70.300	16.159	15.678	98.772
Precipitation	20440	2.676	6.160	0.000	105.653
Wind speed	20440	2.316	0.945	0.507	8.808
Temperature	20440	14.975	10.446	-18.818	32.734

Table 1.1 presents summary statistics for the variables under consideration, with a total of 20,440 observations. We observe that the Weibo posts related to sleeplessness range from 392 to 9,868 posts. In the case of $PM_{2.5}$ levels, the range extends from 400.833 ug/m^3 to 4.091 ug/m^3. The mean PM2.5 value is 51.140 ug/m^3, indicating a relatively low average.

However, a comparison with data from the Statista database[16] reveals that the average $PM_{2.5}$ level in China in 2018 was 49.3 ug/m^3. This figure aligns closely with the $PM_{2.5}$ data presented in this paper, suggesting a comparable scenario.

2.4 Empirical Strategy and Results

Following the notation in Heyes and Zhu (2019), the model can be expressed as follows:

$$S_{it} = \alpha + P_{it}\beta + M_{it}\gamma + c_i + \zeta_t + \epsilon_{it} \qquad (1)$$

where S_{it} represents the nightly sleeplessness in city i on day t. P_{it} is the daily average $PM_{2.5}$ pollution concentration in city i on day t. M_{it} consists of vectors of weather controls, including average temperature, average humidity, average precipitation, and average wind speed. c_i and ζ_t are space and time-specific effects that control for time and space-invariant city characteristics, respectively. ϵ_{it} is the error term. The primary focus is on the estimated coefficient β, which signifies the

[16] You can access the Statista database through the following link: https://www.statista.com/statistics/1048610/china-annual-pm25-particle-levels/.

effect of $PM_{2.5}$ on sleeplessness. It indicates how much sleeplessness of Weibo users increases by β due to one additional unit of air pollution, $PM_{2.5}$.

The air pollution levels of cities are likely to exhibit spatial correlation (Feng et al., 2019; X. Chen et al., 2017; Cole et al., 2013; Ma et al., 2016; Tang et al., 2016). Cole et al. (2013) highlights various ways in which spatial correlations can impact urban air pollution. First, the error term ϵ_{it} could be spatially correlated due to the omission of spatially correlated explanatory variables. Second, cities in close proximity may share the same local characteristics, such as the strength of air pollution regulations. Third, air pollution tends to exhibit spatial clustering when cities are closely located. Fourth, explanatory variables may be spatially correlated due to similar characteristics of other nearby cities.

Spatial econometrics encompasses a broad spectrum of economic studies, including real estate economics, environmental and agricultural economics, and industrial organization (Bockstael, 1996; Anselin & Rey, 1997; Anselin, 2003; Pinkse & Slade, 2010). Incorporating spatial interaction effects into the model enhances our comprehension of the spatial structure of phenomena, thereby improving the explanatory power of econometric models (Tosetti et al., 2018). Additionally, employing spatial econometrics and spatial weighted matrix regression is an effective strategy for addressing the spatial dependence of omitted variables, which could introduce endogeneity (Zangger, 2019). The spatial weighted matrix reflects the extent of spatial spillover effects between neighboring cities.

Spatial econometrics commonly employs three primary models: the spatial autoregressive model, the spatial error model, and the spatial Durbin model.

The spatial autoregressive model (SAR) is:

$$Y = \rho WY + \alpha l_n + X\beta + \epsilon \qquad (2)$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

Where Y denotes an $N * 1$ vector as the dependent variable for every region (observation) in the sample ($i = 1, 2, ..., N$). l_n is an $N * 1$ vector of value ones associated with constant term parameter α, X denotes an $N * K$ matrix as the independent variables for every region (observation) in the sample ($i = 1, 2, ..., N$)

that is associated with the parameter β, and ϵ is an $N * 1$ vector of disturbance terms. W represents an $N * N$ spatial weighted matrix, WY is a spatially lagged dependent variable and denotes the endogenous interaction effects between the dependent variables. A parameter ρ measures the strength of spatial dependence between dependent variables of different regions; it is a $K * 1$ vector.

The spatial error model (SEM) is:

$$Y = \alpha l_n + X\beta + \mu \qquad (3)$$

$$\mu = \lambda W\mu + \epsilon$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

Where $Y, \alpha l_n, X\beta$, and ϵ contain the same variables as the SAR model. $W\mu$ represents spatially autocorrelated error terms, signifying the interaction effects between the disturbance terms of different observations. The scalar parameter λ gauges the strength of dependence between cities; it is a $K * 1$ vector.

The spatial Durbin model (SDM) is:

$$Y = \rho WY + \alpha l_n + X\beta + WX\theta + \epsilon \qquad (4)$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

Where WX is a spatially lagged independent variable, representing the exogenous interaction effects between independent variables. The parameters ρ and θ quantify the strength of spatial dependence between regions and are $K * 1$ vectors. All other variables' contents are the same as in the SAR model.

If no spatial dependence exists between variables, the equation can be written as:

$$Y = \alpha l_n + X\beta + \epsilon \qquad (5)$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

This equation represents a standard linear regression model (OLS).

The binary spatial contiguity and spatial inverse distance matrices are commonly employed in spatial econometric model regressions. W represents a first-order spatial weighted matrix and is time-invariant. In this matrix, the rows correspond to observation i, and the columns correspond to observation j, reflecting the impact of neighboring observation j on observation i (J. LeSage & Pace, 2009). Consider four cities labeled C1, C2, C3, and C4. C1 is a neighbor of C2 and C3. C2 is a neighbor of C1 and C3. C3 is a neighbor of C1, C2, and C4. C4 is a neighbor of C3.

First, we generate a first-order binary contiguity matrix C:

$$C = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

In matrix C, elements equal to 1 indicate that two regions share a common border, while 0 indicates that they do not. For instance, a value of 1 in row 1 and columns 2 and 3 signifies that C1 shares a border with both C2 and C3. All other elements in row 1 have a value of zero. Elements in rows 2, 3, and 4 carry the same implications as row 1, signifying the shared borders between the corresponding cities. Notably, diagonal elements in the matrix are set to zero, signifying that regions are not considered neighbors of themselves.

The matrix is typically normalized to a spatially weighted matrix with row sums of 1 to create a spatial lag or linear combination of values from neighboring observations. This spatial weighted matrix is:

$$W = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

WY and WX can be understood as the weighted average of the surrounding dependent variables and the weighted average of the surrounding independent variables. $W\mu$ is a weighted average of the surrounding error terms.

The spatial inverse distance matrix is constructed based on the distance between regions. The elements of the spatial inverse distance matrix are:

$$W_{ij} = \frac{1}{d_{ij}}, i \neq j; 0, i = j$$

The construction of a spatial inverse distance matrix is the same as that of the binary contiguity spatial weighted matrix. The diagonal elements of the spatial inverse distance matrix are also zero. The off-diagonal elements represent the inverse distance between cities, where d_{ij} is the distance between city i and city j. A spatial inverse matrix ensures that cities are considered neighbors based on their distance, capturing spatial spillover effects from cities without shared borders to other cities.

Many other spatial weighted matrices exist in spatial econometric research, such as the k-nearest adjacent matrix and the threshold inverse distance matrix. Additionally, some matrices are constructed based on economic standards or the coordinates of observations.

Spatial econometric models encompass three spatial interaction effects and two effects (J. LeSage & Pace, 2009). The three spatial interaction effects are endogenous interaction effects, exogenous effects, and interaction effects between error terms. Endogenous and exogenous interaction effects gauge whether the dependent and explanatory variables, respectively, and error terms of neighboring cities are spatially dependent. Two effects are direct and indirect effect. The direct effect signifies the impact of a change in an explanatory variable on the dependent variable in the same unit. In contrast, the indirect effect measures the spatial spillover effect of variables in a city to its neighboring cities, which is typically constrained to zero in standard econometric models. Indirect effects are further categorized into global and local indirect effects. Global indirect effects represent the impact of a change in a city's explanatory variable that spills over to all other areas, including those not directly connected by a spatial weighted matrix. Additionally, J. LeSage and Pace (2009) argue that the global indirect effect is transmitted to the area of origin of the effect. On the other hand, local indirect effects specifically measure the spatial spillover effects among areas connected according to a weighted spatial matrix.

The SDM model is utilized to introduce direct and indirect effects due to its inclusion of both endogenous and exogenous interaction effects. The reduced form of the SDM model (4) can be expressed through the following calculation:

$$Y = (I - \rho W)^{-1}(\alpha l_n + X\beta + WX\theta + \epsilon)$$

Taking the partial derivatives of $E(Y)$ of the above equation with respect to the K_{th} explanatory variables of X in the city (observation) 1 up to N, we obtain:

$$\begin{bmatrix} \frac{\partial E(Y)}{\partial X_{1k}} & \dots & \frac{\partial E(Y)}{\partial X_{nk}} \end{bmatrix} = (I - \rho W)^{-1} \begin{bmatrix} \beta_k + W_{11}\theta_k & W_{12}\theta_k & \cdots & W_{1n}\theta_k \\ W_{21}\theta_k & \beta_k + W_{22}\theta_k & \cdots & W_{2n}\theta_k \\ \vdots & \vdots & \cdots & \vdots \\ W_{n1}\theta_k & W_{n2}\theta_k & \cdots & \beta_k + W_{nn}\theta_k \end{bmatrix}$$

Since the diagonal elements of the spatial weighted matrix W are zero, the $W_{11}, W_{22}, ..., W_{nn}$ are all equal to 0. The above partial derivatives matrix can be re-written as:

$$\begin{bmatrix} \frac{\partial E(Y)}{\partial X_{1k}} & \dots & \frac{\partial E(Y)}{\partial X_{nk}} \end{bmatrix} = (I - \rho W)^{-1} \begin{bmatrix} \beta_k & W_{12}\theta_k & \cdots & W_{1n}\theta_k \\ W_{21}\theta_k & \beta_k & \cdots & W_{2n}\theta_k \\ \vdots & \vdots & \cdots & \vdots \\ W_{n1}\theta_k & W_{n2}\theta_k & \cdots & \beta_k \end{bmatrix} = (I - \rho W)^{-1}(\beta_k I_n + \theta_k W)$$

Where W is the spatial weighted matrix, and I_n is an identity matrix. The expression of the spatial multiplier matrix is an infinite series as follows:

$$(I - \rho W)^{-1} = I + \rho W + \rho^2 W^2 + \rho^3 W^3 + \cdots$$

W represents the first-order contiguity neighbors. The matrix W^2 shows the second-order contiguity neighbors, which are the neighbors to the first-order neighbors. For example, if C1 has only one contiguity neighbor, C2. C2 has two contiguity neighbors C1 and C3. Then the first-order neighbor of C1 is C2, and the second-order neighbors of C1 are C1 and C3. The matrix W^3 shows the third-order contiguity neighbors, and so on.

Since the diagonal elements of the first matrix term I on the right-hand side of the above equation are 1, the non-diagonal elements are zero, and this term represents a direct effect of a change in X. However, the diagonal elements of the second matrix term ρW on the right-hand side of the above equation are assumed to be zero, and this term represents an indirect effect of a change in the explanatory variables. All other terms on the right-hand side of the above equation represent higher-order indirect effects.

J. LeSage and Pace (2009) defined the direct effect and the indirect effect measured by the diagonal elements and the non-diagonal elements of the matrix $(I - \rho W)^{-1}(\beta_k I_n + \theta_k W)$. They also suggest calculating the average sums of diagonal and non-diagonal elements of the matrix $(I - \rho W)^{-1}(\beta_k I_n + \theta_k W)$ to obtain the average direct and indirect effects, respectively.

Based on the analysis of direct and indirect effects above, the direct and indirect effects of the other models can be derived. SEM and OLS models only have a direct effect, which can be obtained through β_k. The SAR model includes endogenous interaction effects WY but no exogenous interaction effects WX. The direct effect and indirect effect can be obtained directly by $(I - \rho W)^{-1}\beta_k I_n$. The SDM model includes both endogenous and exogenous interaction effects, and the direct effect and indirect effect can be obtained directly by $(I - \rho W)^{-1}(\beta_k I_n + \theta_k W)$.

Although the SEM model accounts for spatial dependence in the disturbance process, it does not provide an indirect effect because $E(\epsilon) = 0$. Based on the definition of the local and global indirect effects, the SAR model estimates global indirect effects, and the SDM model estimates both global and local indirect effects. However, the SAR model includes a consistent ratio between direct and indirect effects for all variables and the inability to discern whether a significant indirect effect is due to omitted spatially correlated explanatory variables or spatially correlated dependent variables (Halleck Vega & Elhorst, 2015), the indirect effects of the SAR model are only global. The SEM model lacks the ability to account for spatial spillover effects, because it solely controls for spatial dependence in error terms (Halleck Vega & Elhorst, 2015).

J. LeSage and Pace (2009) recommended the widespread use of the SDM model in spatial analysis because it controls for both exogenous and endogenous spatial effects and has no prior restrictions on the ratio between direct and indirect effects. Moreover, the SDM model can control the endogeneity caused by the spatial dependence inherent in omitted variables and exogenous independent variables (J. P. LeSage & Pace, 2008). The process of how the SDM model controls this endogeneity is shown as follows:

$$Y = X\beta + O \qquad (6)$$

$$O = \rho W O + X\gamma + \epsilon \qquad (7)$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

Assume that Y is influenced by an exogenous variable X and an omitted variable O, and that O is not only spatially dependent but also correlated with X. Rearranging equation (7) yields the following:

$$O = (I - \rho W)^{-1}(X\gamma + \epsilon) \qquad (8)$$

Substituting equation (8) into (6) and rearranging yields the SDM model (9):

$$Y = \rho W Y + X(\beta + \gamma) - \rho W X \beta + \epsilon \qquad (9)$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

Spatial econometrics has evolved from estimating cross-sectional data to panel data. J. P. Elhorst (2017) concluded that spatial panel econometric models have two advantages. First, they can control for space-specific effects. Because applying spatial models does not imply that all unobserved space-specific and time-invariant variables can be controlled for. For instance, cities near the ocean will have different economic and political backgrounds than inland cities. Second, spatial panel econometrics can control for time-specific effects caused by space-invariant variables, such as the adjustment of government policies that can significantly affect the functioning of the economy.

The extension of the spatial cross-sectional model to the spatial panel model for N observations over T periods is obtained by adding a subscript t, and by introducing space-specific effects (r_i) and time-specific effects (ζ_t) that control for unobserved time-invariant and space-invariant variables, as follows:

The panel SAR model is:

$$Y_{it} = \rho W Y_{it} + \alpha l_n + X_{it}\beta + r_i + \zeta_t + \epsilon_{it}$$

$$\epsilon_{it} \sim N(0, \sigma^2 I_n)$$

The panel SEM model is:

$$Y_{it} = \alpha l_n + X_{it}\beta + r_i + \zeta_t + \mu_{it}$$

$$\mu_{it} = \lambda W \mu_{it} + \epsilon$$

$$\epsilon_{it} \sim N(0, \sigma^2 I_n)$$

The panel SDM model is:

$$Y_{it} = \rho W Y_{it} + \alpha l_n + X_{it}\beta + W X_{it}\theta + r_i + \zeta_t + \epsilon_{it}$$

$$\epsilon_{it} \sim N(0, \sigma^2 I_n)$$

Where i is the city of observation ($i = 1, 2, 3, ..., N$) and t is the time period ($t = 1, 2, 3, ..., T$).

The spatial panel models used in this research can be written as:

The panel SAR model is:

$$S_{it} = \rho W S_{it} + \alpha + P_{it}\beta + M_{it}\gamma + r_i + \zeta_t + \epsilon_{it}$$

$$\epsilon_{it} \sim N(0, \sigma^2 I_n)$$

The panel SEM model is:

$$S_{it} = \alpha + P_{it}\beta + M_{it}\gamma + c_i + \zeta_t + \mu_{it}$$

$$\mu_{it} = \lambda W \mu_{it} + \epsilon$$

$$\epsilon_{it} \sim N(0, \sigma^2 I_n)$$

The panel SDM model is:

$$S_{it} = \rho W S_{it} + \alpha + P_{it}\beta + M_{it}\gamma + W P_{it}\theta_1 + W M_{it}\theta_2 + c_i + \zeta_t + \epsilon_{it}$$

$$\epsilon_{it} \sim N(0, \sigma^2 I_n)$$

W is a spatial weighted matrix. θ_1 and θ_2 represent the effects of neighboring cities' air pollution and meteorological factors on sleeplessness. The main interest of the coefficient is β and θ_1. The space- and time-specific factors are treated as fixed effects. Maximum likelihood (ML) estimators are extended to include fixed effects. The response parameters of the fixed effects model can be estimated by first concentrating out the fixed effects (called demeaning). The resulting equation can be estimated using the ML estimation (J. P. Elhorst, 2014).

Before delving into the spatial econometric analysis, three tests will be conducted. First, the global and local Moran's I tests[17] (Anselin, 1995) are utilized to examine the spatial autocorrelation of the average monthly $PM_{2.5}$ across months.

Second, classical and robust panel Lagrange multiplier (LM) tests are employed to assess whether spatial econometric models offer a more suitable description of the data compared to a model without spatial interaction effects.

Third, Lagrange multiplier tests (Burridge, 1981) are applied to determine whether the Spatial Durbin Model (SDM) can be simplified to Spatial Autoregressive (SAR) or Spatial Error Model (SEM) models. A binary contiguity spatial weighted matrix is adopted for tests and the estimation of spatial econometric models.

Additionally, two different spatial weighted matrices (5 neighbors and 200KM inverse distance spatial weighted matrices) are used for robustness checks. This choice is informed by the understanding that the magnitude of spatial spillover effects of $PM_{2.5}$ decreases as a city has more neighbors or as the distance between a city and its neighbors increases. Hence, these two additional spatial weighted matrices are adopted only for robustness checks. This approach aligns with Tobler's first law, as articulated by Sui (2004), stating that "everything is related to everything else, but nearby things are more related than distant things" (p. 236).

[17] All results of panel global and local Morans' I tests were generated in Stata using Maurizio Pisati's tools for spatial analysis.

2.5 Results

1.5.1 Empirical Results

Table 2.2: The OLS estimation

Dependent variable: sleeplessness					
	Pooled OLS	City FEs	Temporal controls		Meteorological controls
	(1)	(2)	(3)	(4)	(5)
$PM_{2.5}$	2.667*** (0.000)	2.722*** (0.000)	2.667*** (0.000)	0.696** (0.013)	-0.068 (0.804)
Precipitation	N	N	N	N	-1.385 (0.323)
Wind speed	N	N	N	N	-12.944 (0.164)
Temperature	N	N	N	N	1.764 (0.386)
Humidity	N	N	N	N	10.614*** (0.000)
Observation	20440	20440	20440	20440	20440
City FEs	N	Y	Y	Y	Y
Weekends FEs	N	N	Y	Y	Y
Province by season FEs	N	N	N	Y	Y
Meteorological controls	N	N	N	N	Y

Notes: *** indicates significance at the 1% confidence level, ** at 5%, and * at 10%.

All standard errors of the estimation clustered at city level. Table 1.2 presents the coefficient estimates of the OLS model described by equation (1) for $PM_{2.5}$. The dependent variable is the daily count of Weibo posts related to sleeplessness, while the independent variable is the daily average of the air pollutant $PM_{2.5}$. In Column (1), the estimation is conducted without controlling for fixed effects and meteorological variables. The coefficient for $PM_{2.5}$ is positively associated with sleeplessness, and it is statistically significant at the 1% level. This coefficient implies that a 1 ug/m^3 increase in $PM_{2.5}$ corresponds to an increase of 3 Weibo posts about sleeplessness, assuming no influence from other variables. In the context of Weibo users, it can be interpreted as follows: a 1 ug/m^3 increase in $PM_{2.5}$ is linked to approximately three more Weibo users experiencing sleeplessness if there are no other effects.

In Column (2) of Table 1.2, the OLS regression includes controls only for city fixed effects, effectively eliminating unobserved time-invariant effects such as city population and Weibo penetration. The coefficient estimate for $PM_{2.5}$ remains significant at the 1% level and positive. Columns (3) and (4) present coefficient estimates for $PM_{2.5}$ after accounting for weekend and province-by-season fixed effects. Despite this control, the significant positive effect of $PM_{2.5}$ on Weibo users' sleeplessness persists. The coefficients are lower compared to the model without time-fixed effects, indicating significant seasonally and weekend variations in sleep patterns. However, in Column (5), after controlling for meteorological factors, the coefficients for $PM_{2.5}$ lose significance.

Table 1.12 (Appendix) illustrates the impact of $PM_{2.5}$ and meteorological factors on sleeplessness. In this table, I controlled for meteorological factors to ascertain which ones might confound the effects of $PM_{2.5}$. First, I controlled for daily average precipitation, and both $PM_{2.5}$ and precipitation showed significant positive effects on Weibo users' sleeplessness. Second, after controlling for daily average wind speed, $PM_{2.5}$ and precipitation still exhibited significant positive effects on Weibo users' sleeplessness, while wind speed had a large significant negative effect. Keeping other variables constant, a 1 km/h increase in daily average wind speed resulted in a decrease of 36 individuals experiencing sleeplessness among Weibo users. This implies that higher wind speed is associated with a reduction in human sleeplessness. Subsequently, I accounted for the daily average temperature, which revealed an insignificant impact on Weibo users' sleeplessness. $PM_{2.5}$, precipitation, and wind speed still positively and significantly affect sleeplessness among Weibo users.

However, after controlling for daily average humidity, the effects of $PM_{2.5}$, precipitation, and wind speed on Weibo users' sleeplessness became insignificant and the effect of temperature is still insignificant. Humidity independently showed significant and positive effects on Weibo users' sleeplessness. This suggests that humidity confound the effect of $PM_{2.5}$ on Weibo users' sleeplessness.

Table 2.3: Global Moran's I test of $PM_{2.5}$

$PM_{2.5}$ global Moran's I test:		
Month	Moran's Index	P-value
March 2018	0.100***	0.000
April 2018	0.033*	0.051
May 2018	0.131***	0.000
June 2018	0.134***	0.000
July 2018	0.018	0.277
August 2018	0.055***	0.001
September 2018	0.079***	0.000
October 2018	0.018	0.269
November 2018	0.307***	0.000
December 2018	0.137***	0.000
January 2019	0.060***	0.000
February 2019	0.236***	0.000

Notes: *** indicates significance at the 1% confidence level, ** at 5%, and * at 10%.

Table 1.3 presents the results of the spatial global Moran's I test for the monthly average $PM_{2.5}$ in each month. All Moran's indices are significant at the 1% level, except for April 2018 at the 10% level, and July 2018 and October 2018, which are insignificant. These results imply that $PM_{2.5}$ levels in different cities exhibit significant spatial autocorrelation for most months.

Figure 1.3, displayed in the Appendix, illustrates the spatial local Moran's I test for the monthly average $PM_{2.5}$ in March, August, November, and January. The figure indicates that significant local autocorrelation is present in $PM_{2.5}$ levels between different cities, aligning with the findings from the global Moran's I test.

Table 2.4: Diagnostic checking between spatial models and the general OLS model

Diagnostic checking		
Null hypothesis: no spatial autocorrelation		
Spatial weighted matrix: contiguity		
Spatial error model:	Statistic	P-value
Lagrange multiplier	12135.272***	0.000
Robust Lagrange multiplier	0.334	0.563
Spatial autoregression model:	Statistic	P-value
Lagrange multiplier	12199.934***	0.000
Robust Lagrange multiplier	65.017***	0.000
Notes: *** indicates significance at the 1% confidence level, ** at 5%, and * at 10%.		

Table 1.4 presents classical and robust LM tests to diagnose whether spatial autocorrelation is present in the regression. Concerning the spatial error check, the classical LM significantly rejects the null hypothesis of no spatial error autocorrelation, while the robust LM test does not reject the null hypothesis. The classical LM statistic is 12135.272 and is significant at the 1% level, whereas the robust LM statistic is 0.334 and is insignificant. Meanwhile, both classical and robust LM tests significantly reject the absence of spatial autocorrelation of the spatial lag, with the classical (12199.934) and robust LM (65.017) statistics both being significant at the 1% level. This suggests that the SAR model is more appropriate for analyzing the effects of $PM_{2.5}$ on sleeplessness when employing spatial econometrics. All of these spatial autocorrelation test results lead to the conclusion that it is necessary to adopt spatial econometrics to study the effects of $PM_{2.5}$ on sleeplessness.

Table 2.5: Spatial model tests between SAR, SEM, and SDM

Spatial model tests		
Spatial weighted matrix: Contiguity		
Null hypothesis: SDM model can be simplified to the SEM model, H0: $\theta + \lambda\beta = 0$		
	Statistic	P-value
Lagrange multiplier	41.26***	0.000
Null hypothesis: SDM model can be simplified to the SAR model, H0: $\theta = 0$		
	Statistic	P-value
Lagrange multiplier	13.99***	0.002
Notes: *** indicates significance at the 1% confidence level, ** at 5%, and * at 10%.		

All the LM results in Table 1.5 indicate that the SDM model cannot be simplified to the SAR or SEM models. All ML statistics (41.26 for the SEM model, 13.99 for the SAR model) are significant at the 1% level. Therefore, the SDM model is used to analyze the effect of $PM_{2.5}$ on sleeplessness of Weibo users. Nevertheless, the results of the SAR and SEM models are also presented, which can be compared with the results of the SDM model.

Table 2.6: Full-season data SDM model estimation

Dependent variable: sleeplessness	SDM	SDM	SDM	SDM	SDM
Matrix	Contiguity	Contiguity	Contiguity	Contiguity	Contiguity
	(1)	(2)	(3)	(4)	(5)
$PM_{2.5}$	0.100 (0.805)	0.056 (0.897)	0.059 (0.892)	-0.036 (0.921)	-0.134 (0.723)
Precipitation	N	N	N	N	3.059 (0.194)
Wind speed	N	N	N	N	-12.103 (0.478)
Temperature	N	N	N	N	-30.568** (0.038)
Humidity	N	N	N	N	-3.870 (0.177)
$W * PM_{2.5}$	1.091** (0.019)	1.138** (0.018)	1.122** (0.020)	0.410 (0.328)	-0.013 (0.976)
W*Precipitation	N	N	N	N	-4.557** (0.047)
W*Wind speed	N	N	N	N	4.415 (0.851)
W*Temperature	N	N	N	N	31.583** (0.035)
W*Humidity	N	N	N	N	9.837*** (0.000)
ρ	0.619*** (0.000)	0.619*** (0.000)	0.615*** (0.000)	0.519*** (0.000)	0.506*** (0.000)
$PM_{2.5}$					
Direct effect	0.380 (0.342)	0.332 (0.435)	0.338 (0.429)	0.039 (0.913)	-0.138 (0.709)
Indirect effect	2.850*** (0.000)	2.801*** (0.000)	2.838*** (0.000)	0.723 (0.242)	-0.209 (0.744)
Total effect	3.230*** (0.000)	3.133*** (0.000)	3.176*** (0.000)	0.762 (0.227)	-0.347 (0.590)
Observation	20440	20440	20440	20440	20440
City FEs	N	Y	Y	Y	Y
Weekends FEs	N	N	Y	Y	Y
Province by season FEs	N	N	N	Y	Y
Meteorological controls	N	N	N	N	Y

Notes: *** indicates significance at the 1% confidence level, ** at 5%, and * at 10%.

Table 1.6 presents the regression results for the SDM model using full-season data and the binary contiguity spatial weighted matrix. The spatial autocorrelation coefficients (0.619, 0.619, 0.615, 0.519, and 0.506) of are significant at the 1% level in columns (1) to (5). This implies that the global spatial autocorrelation of $PM_{2.5}$ is highly significant. $W * PM_{2.5}$ represents the local spatial autoregression of $PM_{2.5}$, and it is insignificant after controlling for province-by-season fixed effects and meteorological factors. Due to the global and local spatial autocorrelation of the explanatory variables, the coefficients in the SDM model do not indicate the marginal effects of the explanatory variables. The direct and indirect effects should be estimated to explain the effects of the explanatory variables on sleeplessness. The estimates of the independent variables differ from their direct effects because of the feedback effects of the global spatial effects and, to some extent, the local spatial effects.

The estimates in column (1) of Table 1.6 present the SDM model without controlling for fixed effects and meteorological factors. The direct effect of $PM_{2.5}$ on sleeplessness is insignificant, and its value is lower than its coefficient estimates in the non-spatial model. This suggests that a city's $PM_{2.5}$ has an insignificant effect on that city's sleeplessness of Weibo users. In the SDM model, the impact of $PM_{2.5}$ on Weibo users' sleeplessness is partitioned into direct and indirect effects, accounting for the spatial dependencies of $PM_{2.5}$. Consequently, $PM_{2.5}$ in a given city may influence the sleeplessness of neighboring cities through wind-mediated transport, yielding a significant impact on Weibo users in those cities. However, the direct effect of $PM_{2.5}$ within the originating city is insignificant. Concurrently, the sleeplessness of Weibo users in the originating city is influenced by $PM_{2.5}$ from its neighboring cities, transported by the wind. The estimates of $PM_{2.5}$ in the non-spatial model are both significant and positive, likely due to the inclusion of indirect effects from neighboring cities. Thus, in the non-spatial model, which cannot distinguish between the direct and indirect effects of $PM_{2.5}$.

The indirect effects of $PM_{2.5}$, initially set to zero in the non-spatial model, in the SDM model is 2.85 and is significant at the 1% level, appearing to be 7.5 times stronger than the direct effect. In other words, if the $PM_{2.5}$ in a certain city increases by 1 ug/m^3, the sleeplessness of Weibo users in neighboring cities will increase by three people when holding other variables constant.

After accounting for city and weekend fixed effects in the SDM model, $PM_{2.5}$ demonstrates a similar impact on Weibo users' sleeplessness, as depicted in columns (2) and (3), in comparison to the outcomes in column (1). When considering the province-by-season fixed effects, both the direct and indirect effects of $PM_{2.5}$ on Weibo users' sleeplessness, as indicated in column (4), are deemed insignificant. Moreover, the estimate of $PM_{2.5}$ after adjusting for meteorological factors remains statistically insignificant. These findings suggest that unobserved seasonal disruptions in different provinces and meteorological factors exert a more noticeable influence on Weibo users' sleep.

Figure 1.1 displays a line chart illustrating the monthly average $PM_{2.5}$ and Weibo sleeplessness posts. The Weibo sleeplessness posts exhibit a continuous increasing trend throughout the observed period. Notably, from April 2018 to July 2019, there is a particularly sharp increase in Weibo sleeplessness posts. During this period, temperatures rise, and $PM_{2.5}$ levels decrease. However, after August, $PM_{2.5}$ levels begin to rise, and Weibo sleeplessness posts show a fluctuating upward trend.

In consideration of the impact of the Chinese winter heating system, which operates from November to February of the following year, $PM_{2.5}$ levels are higher during these months compared to others (Q. Xiao, Ma, Li, & Liu, 2015). The winter heating system may represent an unobserved seasonal shock in different provinces, especially since this policy is implemented exclusively in the northern region of China and not in the southern region. This study encompasses data from two northern provinces and two southern provinces.

Table 1.13 (in the Appendix) displays SDM model estimations for each season. Direct effects of $PM_{2.5}$ are found to be statistically insignificant in every season. Indirect effects of $PM_{2.5}$ are significant only in spring and winter, with significances at the 10% and 1% levels, respectively. Consequently, the analysis will specifically focus on the winter season (the period between November 2018 and February 2019) to investigate the effect of $PM_{2.5}$ on sleeplessness.

Table 2.7: Winter data SDM estimation

Dependent variable: sleeplessness	SDM	SDM	SDM	SDM	SDM
Matrix	Contiguity	Contiguity	Contiguity	Contiguity	Contiguity
	(1)	(2)	(3)	(4)	(5)
$PM_{2.5}$	-0.031 (0.937)	-0.027 (0.948)	-0.021 (0.958)	-0.021 (0.958)	-0.019 (0.964)
Precipitation	N	N	N	N	-0.248 (0.976)
Wind speed	N	N	N	N	-29.829 (0.309)
Temperature	N	N	N	N	-60.907** (0.019)
Humidity	N	N	N	N	-6.747 (0.196)
$W * PM_{2.5}$	0.872* (0.050)	0.909** (0.043)	0.894** (0.046)	0.894** (0.046)	0.967* (0.055)
W*Precipitation	N	N	N	N	11.555 (0.235)
W*Wind speed	N	N	N	N	6.858 (0.833)
W*Temperature	N	N	N	N	38.474 (0.143)
W*Humidity	N	N	N	N	12.113** (0.034)
ρ	0.541*** (0.000)	0.541*** (0.000)	0.539*** (0.000)	0.539*** (0.000)	0.506*** (0.000)
$PM_{2.5}$					
Direct effect	0.139 (0.710)	0.144 (0.718)	0.152 (0.703)	0.152 (0.703)	0.147 (0.723)
Indirect effect	1.765*** (0.003)	1.772*** (0.003)	1.809*** (0.002)	1.8099*** (0.002)	1.782*** (0.008)
Total effect	1.904*** (0.001)	1.916*** (0.000)	1.961*** (0.001)	1.961*** (0.001)	1.930*** (0.006)
Observation	6720	6720	6720	6720	6720
City FEs	N	Y	Y	Y	Y
Weekends FEs	N	N	Y	Y	Y
Time FEs	N	N	N	Y	Y
Meteorological Controls	N	N	N	N	Y
Notes: *** indicates significance at the 1% confidence level, ** at 5%, and * at 10%.					

Table 1.7 displays estimates derived from the SDM model using winter data,

incorporating successive controls for city, weekend, and time fixed effects, as well as meteorological factors. Since this dataset pertains to the winter season, the estimation does not account for seasonal fixed effects.

In this setup, $PM_{2.5}$ exhibits an insignificant direct effect and a significant (all at 1% level) indirect effect on Weibo users' sleeplessness, and estimates are positive. Column (5) displays the estimates of SDM model after controlling city, weekend, time fixed effects, and meteorological factors. The estimates can be explained as, during winter, when $PM_{2.5}$ increases by ug/m^3, Weibo users' sleeplessness may not increase in that city, but the sleeplessness of Weibo users in neighboring cities increases by two persons holding the other variables constant.

Tables 1.9 and 1.10 (in the Appendix) present estimations of the SAR, SEM, and SDM models utilizing the binary contiguity spatial weighted matrix based on full-season data and winter data, respectively. All models incorporate controls for city, weekend, time fixed effects, and meteorological factors.

Table 1.9 displays the regression results of the spatial econometric models using full-season data. $PM_{2.5}$ exhibits insignificant direct and indirect effects in the SAR and SDM models. The coefficient estimates for the SEM model represent only direct effects and are also insignificant.

The estimates of the spatial econometric models for the winter data are presented in Table 1.10. In the SAR model, $PM_{2.5}$ demonstrates significant and positive direct and indirect effects on the sleeplessness of Weibo users. Holding other variables constant, if the $PM_{2.5}$ of a city increases by 1 ug/m^3, the sleeplessness of Weibo users in that city and its neighboring cities would each increase by one person. In contrast, in the SDM model, $PM_{2.5}$ has an insignificant direct effect on the sleeplessness of Weibo users. However, the coefficient estimation of $PM_{2.5}$'s indirect effect is three times that of the SAR model. In the SEM model, the effect of $PM_{2.5}$ on Weibo users' sleeplessness remains insignificant.

The observed differences in direct and indirect effects between the SAR and SDM models arise from a limitation of the SAR model, which maintains a consistent ratio between direct and indirect effects for all variables. In the SDM model, encompassing both endogenous and exogenous spatial interactions of dependent and independent variables, the analysis extends to both global and local indirect effects.

The direct effect of $PM_{2.5}$, even strengthened by the feedback effect generated by the global indirect effect, on the original city's sleeplessness of Weibo users remains insignificant. This indicates that the indirect effect of $PM_{2.5}$ contains a higher ratio than the direct effect of $PM_{2.5}$ in the SDM model. However, the SAR model assumes an equal ratio for both direct and indirect effects, leading to the observed differences in estimations between the two models.

1.5.2 Robustness check

For robustness checks, two distinct spatial weighted matrices (5 neighbors and 200KM inverse distance spatial weighted matrices) are employed. Table 1.8 presents the direct and indirect effects of $PM_{2.5}$ estimated by the SDM model using different spatial weighted matrices. When analyzing the full-season data, $PM_{2.5}$ exhibits insignificant direct and indirect effects on sleeplessness, even after controlling for meteorological factors.

During the winter, the direct effect of $PM_{2.5}$ is insignificant, but the indirect effects of $PM_{2.5}$ significantly and positively influences Weibo users' sleeplessness. In other words, in winter, a 1 ug/m^3 increase in $PM_{2.5}$ may not lead to an increase in Weibo users' sleeplessness in that city. However, the sleeplessness of Weibo users in neighboring cities increases by two individuals if no other effects are considered. This observation supports the findings presented in this paper.

Table 2.8: Robustness check

Model: SDM				
Dependent variable: sleeplessness	Full-season data	Full-season data	Winter data	Winter data
Spatial weighted matrix:	5 neighbors	200KM inverse distance	5 neighbors	200KM inverse distance
PM2.5	(1)	(2)	(3)	(4)
Direct effect	-0.354	-0.149	0.077	0.070
	(0.320)	(0.692)	(0.824)	(0.855)
Indirect effect	-0.446	-1.083	1.701**	1.785**
	(0.596)	(0.219)	(0.046)	(0.030)
Total effect	-0.800	-1.232	1.777**	1.854**
	(0.401)	(0.207)	(0.043)	(0.031)
	0.649***	0.659***	0.588***	0.608***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	20440	20440	6720	6720
Additional contorls:				
City FEs	Y	Y	Y	Y
Weekends FEs	Y	Y	Y	Y
Time FEs	Y	Y	Y	Y
Meteorological Controls	Y	Y	Y	Y
Notes: *** indicates significance at the 1% confidence level, ** at 5%, and * at 10%.				

1.5.3 The effects of the other air pollutants

Other air pollutants can also impact human sleep quality (Cheng et al., 2019; Zhou et al., 2023). Consequently, I conducted an analysis to explore the effects of AQI and additional air pollutants including PM_{10}, SO_2, CO, and O_3, on Weibo users' sleeplessness during winter. The analysis utilized SDM models with a binary contiguity spatial weight matrix and controlled for city, weekend, and time fixed effects, and meteorological factors. The results are presented in Table 1.11 (in the Appendix).

AQI and PM_{10} exhibit significant positive indirect effects on Weibo users' sleeplessness, although these effects are smaller than those observed for $PM_{2.5}$. Conversely, the direct and indirect effects of SO_2, CO, and O_3 on Weibo users' sleeplessness are found to be statistically insignificant. Consequently, while AQI and PM_{10} may influence the sleeplessness of Weibo users in neighboring cities, the impact of $PM_{2.5}$ is more pronounced.

2.6 Conclusion

In this study, a measure of daily nighttime sleeplessness is combined with daily data on $PM_{2.5}$ and meteorological factors (including temperature, humidity, wind speed, and precipitation) to create a daily city-level dataset spanning from March 2018 to February 2019. The objective is to investigate the impact of air pollution, specifically $PM_{2.5}$, on sleeplessness using this daily city-level dataset. The sleeplessness measure in this research is derived by aggregating posts containing keywords related to sleeplessness from the widely used Chinese social media platform, Weibo. This innovative measure of sleeplessness is not only easier to implement but also more representative compared to traditional survey data.

This paper employs a Spatial Durbin Model (SDM) with a binary contiguity spatial weighted matrix to address endogeneity resulting from omitting variables with spatial dependence. The analysis also includes controls for city and weekend fixed effects. Additionally, two other spatial weighted matrices (5 neighbors and 200KM inverse distance matrices) are used for robustness checks.

There are four main findings in this paper. First, the global and local spatial Moran's I test indicates that $PM_{2.5}$ exhibits significant spatial autocorrelation, strengthening the rationale for using spatial econometric models. The results are not only plausible but also more robust compared to a general model, especially when spatial autocorrelation of $PM_{2.5}$ is present, and it is controlled by the SDM model.

Second, $PM_{2.5}$ demonstrates no direct or indirect effect on sleeplessness when analyzing full-season data and controlling for city, weekend, provinces-by-season fixed effects, and meteorological factors.

Third, focusing on winter data (when there is greater severity of $PM_{2.5}$ due to winter heating system), if $PM_{2.5}$ increases by one unit, the sleeplessness of Weibo users in that city may not increase, but the sleeplessness of Weibo users in neighboring cities will increase by two people, holding other variables constant. The robustness checks also support this finding. However, this does not imply that the sleeplessness of Weibo users in the target city is not affected by $PM_{2.5}$. The impact of $PM_{2.5}$ on the target city stems from the PM2.5in neighboring cities, driven by wind patterns—illustrating the spatial dependence of $PM_{2.5}$. Consequently, $PM_{2.5}$ from neighboring cities is transported to the target city via the wind, influencing sleeplessness in the target city.

The fourth finding pertains to the selection of spatial models. The SDM model appears to be more suitable than either the SEM or the SAR model. The results of the LM tests consistently support the SDM model.

Finally, the effects of other air pollutants on Weibo users' sleeplessness in winter were also analyzed. The results of the SDM model indicate that both AQI and PM_{10} have a significant positive indirect effect on the sleeplessness of Weibo users. However, the magnitudes of the effects of AQI and PM_{10} are smaller than that of $PM_{2.5}$.

Here are two policy suggestions. First, the government could could analyze Weibo data to gauge residents' air pollution concerns in various areas, enabling tailored policy responses based on regional differences.

Second, media information could serve as an indicator of residents' mental health. Medical institutions could leverage this public data for diagnosing mental health conditions, which might be more efficient and effective than using traditional questionnaires.

2.7 Appendix

Figure 2.3: Local Moran's I tests of monthly average $PM_{2.5}$

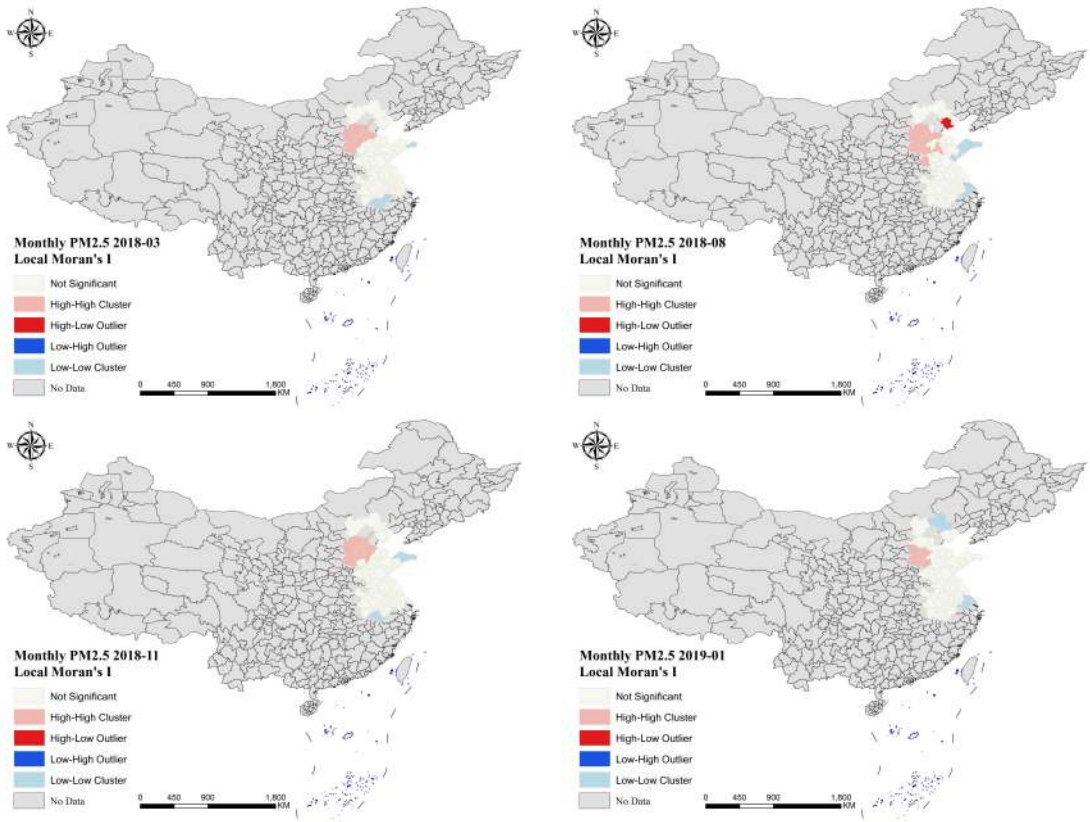


Table 2.9: SAR, SEM, and SDM models with full-season data

Dependent variable: sleeplessness	SAR	SEM	SDM
Matrix	Contiguity	Contiguity	Contiguity
	(1)	(2)	(3)
$PM_{2.5}$	-0.224	-0.121	-0.157
	(0.390)	(0.753)	(0.709)
Precipitation	0.456	1.779	2.999
	(0.726)	(0.401)	(0.211)
Wind speed	-53.537***	-76.769***	-4.739
	(0.000)	(0.000)	(0.806)
Humidity	-7.456***	-16.074***	-20.426
	(0.000)	(0.000)	(0.251)
Temperature	5.502***	9.210***	-4.060
	(0.000)	(0.000)	(0.199)
$W * PM_{2.5}$	N	N	-0.203
			(0.665)
W*Precipitation	N	N	-3.316
			(0.192)
W*Wind speed	N	N	-58.402**
			(0.031)
W*Humidity	N	N	12.578
			(0.475)
W*Temperature	N	N	10.469***
			(0.001)
ρ	0.591***	N	0.588***
	(0.000)		(0.000)
λ	N	0.592***	N
		(0.000)	
$PM_{2.5}$			
Direct effect	-0.240	N	-0.205
	(0.423)		(0.322)
Indirect effect	-0.281	N	-0.691
	(0.427)		(0.744)
Total effect	-0.521	N	-0.895
	(0.425)		(0.235)
Observation	20440	20440	20440
City FEs	Y	Y	Y
Weekends FEs	Y	Y	Y
Time FEs	Y	Y	Y
Meteorological Controls	Y	Y	Y

Notes: *** indicates significance at the 1% confidence level, ** at 5%, and * at 10%.

Table 2.10: SAR, SEM, and SDM models with winter data

Dependent variable: sleeplessness	SAR	SEM	SDM
Matrix	Contiguity	Contiguity	Contiguity
	(1)	(2)	(3)
$PM_{2.5}$	0.596**	0.154	-0.019
	(0.027)	(0.175)	(0.964)
Precipitation	9.135**	13.948**	-0.248
	(0.011)	(0.020)	(0.976)
Wind speed	-25.718***	-40.323**	-29.829
	(0.000)	(0.042)	(0.309)
Temperature	-21.038***	-37.720***	-60.907**
	(0.000)	(0.000)	(0.019)
Humidity	4.707***	7.184***	-6.747
	(0.000)	(0.003)	(0.196)
$W * PM_{2.5}$	N	N	0.967*
			(0.055)
W*Precipitation	N	N	11.555
			(0.235)
W*Wind speed	N	N	6.858
			(0.833)
W*Temperature	N	N	38.474
			(0.143)
W*Humidity	N	N	12.113**
			(0.034)
ρ	0.511***	N	0.506***
	(0.000)		(0.000)
λ	N	0.512***	N
		(0.000)	
$PM_{2.5}$			
Direct effect	0.657**	N	0.147
	(0.029)		(0.723)
Indirect effect	0.590**	N	1.782***
	(0.034)		(0.008)
Total effect	1.247**	N	1.930***
	(0.030)		(0.006)
Observation	6720	6720	6720
City FEs	Y	Y	Y
Weekends FEs	Y	Y	Y
Time FEs	Y	Y	Y
Meteorological Controls	Y	Y	Y

Notes: *** indicates significance at the 1% confidence level, ** at 5%, and * at 10%.

Table 2.11: The effects of AQI, PM_{10}, SO_2, CO,and O_3 on Weibo users' sleeplessness in winter

Model: SDM					
Dependent variable: sleeplessness					
Spatial weighted matrix: binary contiguity					
	AQI	PM_{10}	SO_2	CO	O_3
Direct effect	0.121	-0.072	-0.774	0.488	0.724
	(0.753)	(0.765)	(0.747)	(0.702)	(0.495)
Indirect effect	1.630***	1.608***	-11.842	2.644	-2.753
	(0.002)	(0.002)	(0.545)	(0.259)	(0.176)
Total effect	1.751***	1.536***	-12.616	3.133	-2.029
	(0.004)	(0.007)	(0.554)	(0.169)	(0.318)
Observation	6720	6720	6720	6720	6720
City FEs	Y	Y	Y	Y	Y
Weekends Fes	Y	Y	Y	Y	Y
Time FEs	Y	Y	Y	Y	Y
Meteorological controls	Y	Y	Y	Y	Y
Notes: *** indicates significance at the 1% confidence level, ** at 5%, and * at 10%.					

Table 2.12: Meteorological factors' effects on sleeplessness

Dependent variable: sleeplessness				
	Meteorological controls			
	(1)	(2)	(3)	(4)
$PM_{2.5}$	0.878***	0.686***	0.724**	-0.068
	(0.002)	(0.016)	(0.009)	(0.804)
Precipitation	7.528***	8.477***	8.515***	-1.385
	(0.000)	(0.000)	(0.000)	(0.323)
Wind speed	N	-36.743***	-34.459***	-12.944
		(0.000)	(0.000)	(0.164)
Temperature	N	N	3.193	1.764
			(0.121)	(0.386)
Humidity	N	N	N	10.614***
				(0.000)
City FEs	Y	Y	Y	Y
Weekends FEs	Y	Y	Y	Y
Province by season FEs	Y	Y	Y	Y

Notes: *** indicates significance at the 1% confidence level, ** at 5%, and * at 10%.

Table 2.13: Seasonal SDM model estimation

Model: SDM				
Dependent variable: sleeplessness	Summer	Spring	Fall	Winter
Matrix	Contiguity	Contiguity	Contiguity	Contiguity
PM2.5	(1)	(2)	(3)	(4)
Direct effect	-0.135	-1.151	-1.327	0.147
	(0.883)	(0.471)	(0.381)	(0.723)
Indirect effect	3.107*	-2.079	0.620	1.782***
	(0.097)	(0.673)	(0.777)	(0.008)
Total effect	2.972	-3.230	-0.707	1.930***
	(0.126)	(0.521)	(0.683)	(0.006)
ρ	0.541***	0.522***	0.226***	0.506***
	(0.000)	(0.000)	(0.000)	(0.000)
City FEs	Y	Y	Y	Y
Weekends FEs	Y	Y	Y	Y
Time FEs	Y	Y	Y	Y
Meteorological Controls	Y	Y	Y	Y
Notes: *** indicates significance at the 1% confidence level, ** at 5%, and * at 10%.				

2.8 Procedure of Python

Figure 2.4: Procedure of Python

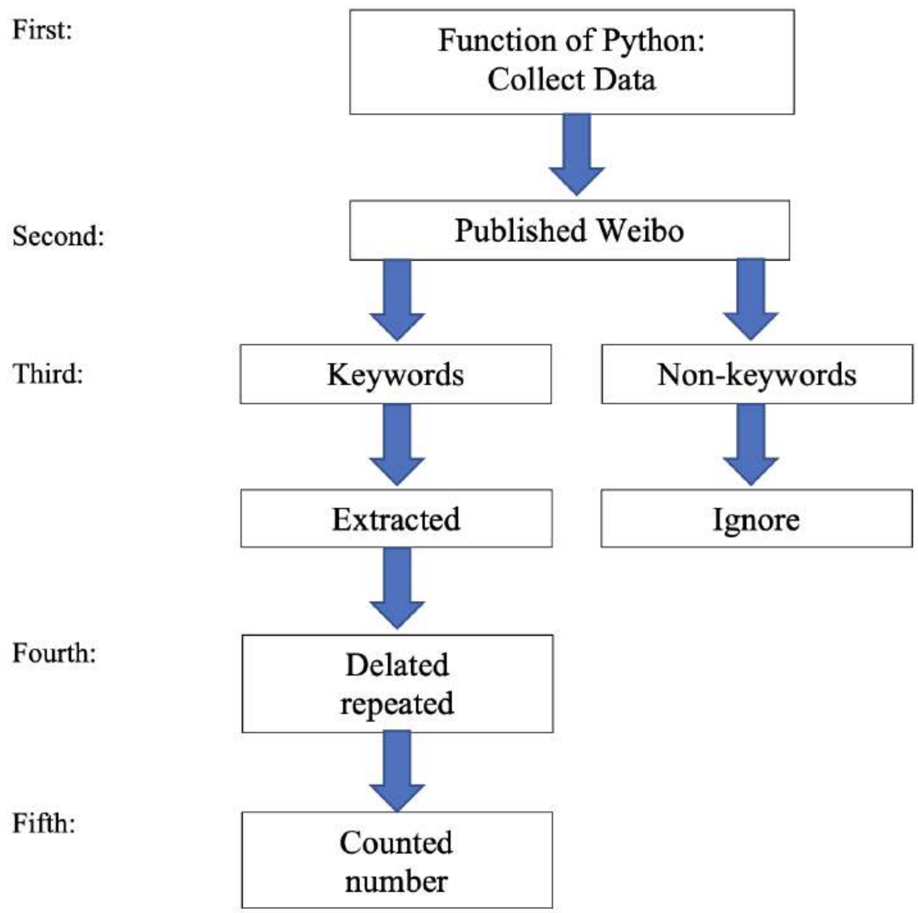


Figure 1.4 illustrates the Python procedure I used to scrape data on sleeplessness from Weibo. First, I integrate Python code with Weibo account cookies and a proxy tunnel. Weibo account cookies can be obtained after logging into my Weibo account. The cookies serve various functions, including storing Weibo users' login status, confirming Weibo user identification, storing users' preferences, and allowing Weibo to collect user information. In my research, I use cookies to maintain my login status when extracting sleeplessness data using Python. However, cookies typically expire after 10 to 15 hours. Consequently, I need to log in to my Weibo account again to obtain new cookies. Generally, a city's data can be fully extracted before the cookies expire, and one or two cookies can be used for data extraction. If using

two cookies, it is necessary to ensure that two Weibo accounts are available.

The reason for using a proxy tunnel is its ability to handle concurrent requests. This means the proxy tunnel allows Python to simultaneously utilize multiple proxy servers, enabling the extraction of sleeplessness data from multiple pages within the same timeframe. This capability enhances the efficiency of the data extraction process.

Furthermore, I integrate a MongoDB database into the Python code. Given the substantial volume of extracted data, this MongoDB database serves as a repository to prevent system overload. Subsequently, the extracted data can be exported and saved in an Excel spreadsheet.

Second, Weibo provides advanced search functionality, allowing me to conduct research on published Weibo posts based on three factors: keywords, period, and the city where Weibo users are located.

Figure 2.5: Weibo example

Figure 1.5 provides an example of Weibo posts obtained through advanced search. In this example, the keyword is "Shuibuzhao," the period spans from 23:00 on January 1, 2015, to 07:00 on January 2, 2015, and the Weibo users are located

in Beijing. Upon clicking the search button with these specified factors, published Weibo posts containing the keywords will be displayed on the screen.

Third, I input keywords, the specified period, and the city code into the Python code. The Python script then extracts published Weibo posts containing the specified keywords within the given period and location of the Weibo users. Once the extraction is complete for one keyword, I proceed to enter another keyword for the subsequent extraction. The procedure for extracting data for different periods and cities follows a similar process as with keywords. In my research, I utilized two keywords, namely "Shuimian" and "Shuibuzhao," with the period set from 23:00 of one day to 07:00 of the following day. Weibo posts without the specified keywords are disregarded in the extraction process.

Fourth, I conduct data cleaning on the extracted Weibo posts. Given that some users may post multiple Weibo messages with the same keywords throughout the night, I eliminate duplicated Weibo posts from the same username on the same day to enhance the dataset's quality.

Fifth, I proceed to tally the cleaned published Weibo posts for each city and every day. Subsequently, I amalgamate the counts of these published Weibo posts to compile a daily sleeplessness dataset.

2.9 Python codes

```
import os
import re
import time
import requests
from lxml import etree
from pandas.core.frame import DataFrame
# 禁用安全请求警告
from pymongo import MongoClient
from requests.adapters import HTTPAdapter
from selenium import webdriver
from urllib3.exceptions import InsecureRequestWarning
import threading
import random

requests.packages.urllib3.disable_warnings(InsecureRequestWarning)
city1 = {
    13.5:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11], ["石家庄", "唐山", "秦皇
岛", "邯郸", "邢台", "保定",
                                               "张家口", "承德", "沧州",
"廊坊", "衡水"]],
    13:
        [[2], ["唐山"]],
    14:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 23], ["太原", "大同", "阳泉",
"长治", "晋城", "朔州",
                                               "晋中", "运城", "忻州",
"临汾", "吕梁"]],
    15:
        [[1, 2, 3, 4, 5, 6, 7, 22, 25, 26, 28, 29], ["呼和浩特", "包头",
"乌海", "赤峰", "通辽", "鄂尔多斯",
                                                     "呼伦贝尔", "兴安
盟", "锡林郭勒盟", "乌兰察布盟",
                                                     "巴彦淖尔盟", "阿拉
善盟"]],
    21:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14], ["沈阳", "大
连", "鞍山", "抚顺", "本溪", "丹东",
                                                           "锦州", "营
口", "阜新", "辽阳", "盘锦", "铁岭",
                                                           "朝阳", "葫芦
岛"]],
    22:
        [[1, 2, 3, 4, 5, 6, 7, 8, 24], ["长春", "吉林", "四平", "辽源",
"通化", "白山",
```

```
                                        "松原", "白城", "延边朝鲜族自治
州"]],
    23:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 27], ["哈尔滨", "齐齐哈
尔", "鸡西", "鹤岗", "双鸭山", "大庆",
                                                       "伊春", "佳木斯",
"七台河", "牡丹江", "黑河", "绥化",
                                                       "大兴安岭"]],
    32:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], ["南京", "无锡",
"徐州", "常州", "苏州", "南通",
                                                       "连云港", "淮安",
"盐城", "扬州", "镇江",
                                                       "泰州", "宿迁"]],
    33:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11], ["杭州", "宁波", "温州",
"嘉兴", "湖州", "绍兴",
                                               "金华", "衢州", "舟山",
"台州", "丽水"]],
    34:
        [[1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18],
["合肥", "芜湖", "蚌埠", "淮南", "马鞍山", "淮北",

"铜陵", "安庆", "黄山", "滁州", "阜阳", "宿州",

"巢湖", "六安", "亳州", "池州", "宣城"]],
    35:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9], ["福州", "厦门", "莆田", "三明",
"泉州", "漳州",
                                       "南平", "龙岩", "宁德"]],
    36:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11], ["南昌", "景德镇", "萍乡",
"九江", "新余", "鹰潭",
                                               "赣州", "吉安", "宜春",
"抚州", "上饶"]],
    38:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17], ["济
南", "青岛", "淄博", "枣庄", "东营", "烟台",
                                                                  "潍
坊", "济宁", "泰安", "威海", "日照", "莱芜",
                                                                  "临
沂", "德州", "聊城", "滨州", "菏泽"]],
    37:
```

```
        [[2], ["青岛"]],
    41:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17],
["郑州", "开封", "洛阳", "平顶山", "安阳", "鹤壁",

"新乡", "焦作", "濮阳", "许昌", "漯河", "三门峡",

"南阳", "商丘", "信阳", "周口", "驻马店"]],
    42:
        [[1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13], ["武汉", "黄石", "十
堰", "宜昌", "襄樊", "鄂州",
                                                            "荆门", "孝感", "荆
州", "黄冈", "咸宁", "随州",
                                                            "恩施土家族苗族自治
州"]],
    43:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 31], ["长沙", "株
洲", "湘潭", "衡阳", "邵阳", "岳阳",
                                                                   "常德", "张家
界", "益阳", "郴州", "永州", "怀化",
                                                                   "娄底", "湘西
土家族苗族自治州"]],
    44:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 12, 13, 14, 15, 16, 17, 18, 19,
20, 51, 52,
          53], ["广州", "韶关", "深圳", "珠海", "汕头", "佛山",
                "江门", "湛江", "茂名", "肇庆", "惠州", "梅州",
                "汕尾", "河源", "阳江", "清远", "东莞", "中山",
                "潮州", "揭阳", "云浮"]],
    45:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12], ["南宁", "柳州", "桂
林", "梧州", "北海", "防城港",
                                                           "钦州", "贵港", "玉
林", "百色", "贺州", "河池",
                                                           ]],
    51:
        [[1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19,
20, 32, 33,
          34], ["成都", "自贡", "攀枝花", "泸州", "德阳", "绵阳",
                "广元", "遂宁", "内江", "乐山", "南充", "眉山",
                "宜宾", "广安", "达州", "雅安", "巴中", "资阳",
                "阿坝", "甘孜", "凉山"]],
    52:
        [[1, 2, 3, 4, 22, 23, 24, 26, 27], ["贵阳", "六盘水", "遵义", "安
```

```
顺", "铜仁", "黔西南",
                                                     "毕节", "黔东南", "黔南"]],
    53:
        [[1, 3, 4, 5, 6, 23, 25, 26, 27, 28, 29, 31, 32, 33, 34, 35],
["昆明", "曲靖", "玉溪", "保山", "昭通",

"楚雄", "红河", "文山", "思茅", "西双版纳",

"大理", "德宏", "丽江", "怒江", "迪庆", "临沧"]],
    61:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10], ["西安", "铜川", "宝鸡", "咸
阳", "渭南", "延安",
                                                    "汉中", "榆林", "安康", "商
洛"]],
    62:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 24, 26, 29, 30], ["兰州", "嘉峪
关", "金昌", "白银", "天水", "武威",
                                                                    "张掖", "平
凉", "酒泉", "庆阳", "定西", "陇南",
                                                                    "临夏", "甘
南"]],
    64:
        [[1, 2, 3, 4, 5], ["银川", "石嘴山", "吴忠", "固原", "中卫"]],
    11.5:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 28,
29, 1000], ["东城区", "西城区", "崇文区", "宣武区",

"朝阳区", "丰台区", "石景山区", "海淀区",

"门头沟区", "房山区", "通州区", "顺义区",

"昌平区", "大兴区", "怀柔区", "平谷区",

"密云县", "延庆县",'北京不限']],
    11:
        [[1000], ['北京不限']],
    50:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 22, 23,
24, 25,
          26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40,
41, 42, 43, 81,
          82, 83, 84], ["万州区", "涪陵区", "渝中区", "大渡口区", "江北区",
                         "沙坪坝区", "九龙坡区", "南岸区", "北碚区", "万盛
区",
```

```
                    "双桥区", "渝北区", "巴南区", "黔江区", "长寿区",
                    "綦江县", "潼南县", "铜梁县", "大足县", "荣昌县",
                    "璧山县", "梁平县", "城口县", "丰都县", "垫江县",
                    "武隆县", "忠县", "开县", "云阳县", "奉节县",
                    "巫山县", "巫溪县", "石柱土家族自治县",
                    "秀山土家族苗族自治县", "酉阳土家族苗族自治县",
                    "彭水苗族土家族自治县", "江津市", "合川市",
                    "永川市", "南川市"]],
    31:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,
18, 19, 20,
          30], ["黄浦区", "卢湾区", "徐汇区", "长宁区", "静安区",
                "普陀区", "闸北区", "虹口区", "杨浦区", "闵行区",
                "宝山区", "嘉定区", "浦东新区", "金山区", "松江区",
                "青浦区", "南汇区", "奉贤区", "崇明县"]],
    12:
        [[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 21, 23,
25], ["和平区", "河东区", "河西区", "南开区", "河北区",

"红桥区", "塘沽区", "汉沽区", "大港区", "东丽区",

"西青区", "津南区", "北辰区", "武清区", "宝坻区",

"宁河县", "静海县", "蓟县"]],
    46:
        [[1, 2, 90], ["海口", "三亚", "其他"]],
    63:
        [[1, 21, 22, 23, 25,
          26, 27, 28], ["西宁", "海东", "海北", "黄南", "海南",
                        "果洛", "玉树", "海西"]],
    54:
        [[1, 21, 22, 23, 24, 25,
          26], ["拉萨", "昌都", "山南", "日喀则", "那曲",
                "阿里", "林芝"]],
    65:
        [[1, 2, 21, 22, 23, 28, 29, 27, 30, 31, 32, 40, 42, 43], ["乌鲁
木齐", "克拉玛依", "吐鲁番", "哈密", "昌吉",
                                                                  "博尔
塔拉", "巴音郭楞", "阿克苏", "克孜勒苏",
                                                                  "喀
什", "和田", "伊犁", "塔城", "阿勒泰"]]}
# appKey = "ckI1NVprakJVUFFjOXloNDpONWR3TWdkaXdhV2Fkakt1"
# ip_port = 'secondtransfer.moguproxy.com:9001'
# proxy = {"http": "http://" + ip_port, "https": "https://" + ip_port}
```

```
'''proxyHost = "http-dyn.abuyun.com"
proxyPort = "9020"
proxyUser = "HF41558568A8365D"
proxyPass = "74046DB3F993AD04"
proxyMeta = "http://%(user)s:%(pass)s@%(host)s:%(port)s" % {
    "host": proxyHost,
    "port": proxyPort,
    "user": proxyUser,
    "pass": proxyPass,
}
proxies = {
    "http": proxyMeta,
    "https": proxyMeta,
}'''
appKey = "d3NEQ3A3M3c5Um5xeXczMTpjYjZFU293Q0t6TzJaYjBl"
ip_port = 'secondtransfer.moguproxy.com:9001'
proxy = {"http": "http://" + ip_port, "https": "https://" + ip_port}
headers = {
    "Proxy-Authorization": 'Basic ' + appKey,
    'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64)
AppleWebKit/537.36 (KHTML, like Gecko) '
                  'Chrome/86.0.4240.111 Safari/537.36 ',
}
cookies = [
    {'Cookie': 'SINAGLOBAL=200690801012.7706.1587320839895;
_ga=GA1.2.1916098359.1587321770;
__gads=ID=77798fbad7acd8db:T=1616514891:R:S=ALNI_Ma6UZIJ-
t_0U_wiPKQivfYdvRoF-w; UOR=,,login.sina.com.cn;
SUBP=0033WrSXqPxfM725Ws9jqgMF55529P9D9W5A2ds7sXxY0oiUb_I55UsF5JpX5KMhU
gL.Fo-4So.ReK-0eKB2dJLoIXeLxKMLB.2LBKzLxKqL1KnL1-
qLxK.L1KMLB-2LxK.L1KMLB-2LxK-LBo5L12qLxK-LBo2L1h2LxK-LB-BLBK.p1K.E;
httpsupgrade_ab=SSL; wvr=6; wb_view_log_5969815314=1280*7201.5;
ALF=1654392034; SSOLoginState=1622856039;
SCF=At30u5O1kJgn3YCw4reSlHF6TTVpydwTRFoV_YjGKr3vcP-
MXgRTNKR4XbJkA7UXeAkR69pfAntloyHA5seR1Eo.;
SUB=_2A25NvqE4DeRhGeNH7VsZ8SvPyjiIHXVuzZXwrDV8PUNbmtAfLWH_kW9NSoa4t1z0
0lVj6UKqqEUzsYzNi1JtgHe7; _s_tentry=login.sina.com.cn;
Apache=562594257917.9099.1622856044727;
ULV=1622856044736:49:2:2:562594257917.9099.1622856044727:1622849945243
;
webim_unReadCount=%7B%22time%22%3A1622856047997%2C%22dm_pub_total%22%3
A5%2C%22chat_group_client%22%3A0%2C%22chat_group_notice%22%3A0%2C%22al
lcountNum%22%3A41%2C%22msgbox%22%3A0%7D'

},
    {'Cookie': 'SINAGLOBAL=4775764169397.394.1572481252068;
_ga=GA1.2.1801845056.1581628166;
__gads=ID=d72e1cd39fe17c1e-2214d94e32c700e3:T=1618034134:S=ALNI_MYNIu4
PLp4_r-2e_5JPhloQ7RJazA;
UOR=www.bing.com,data.weibo.com,login.sina.com.cn;
```

```
SUBP=0033WrSXqPxfM725Ws9jqgMF55529P9D9WFWLlLfUNFZiJHzOAq7xc045JpX5KMhU
gL.FoMfSonpShM7eK.2dJLoI0eLxKqLB-qL12qLxK-LBKBLBK.LxK-L1KeL1hnLxK-
L1KeL1hnLxK-L1KeL1hnEeh54eh.t; ALF=1654385921;
SSOLoginState=1622849923;
SCF=Auy6FzxnCvbMOrn06OCBlBI7n3ZBTTmLZBnX_n2i7twq8ZmLyra-
BK69KexVepoMjaTTdpzLJ0seI43jLaTRi00.;
SUB=_2A25NvsnUDeRhGeFL7VoQ9CnMyjWIHXVuzbwcrDV8PUNbmtAfLW3MkW9Nfdx40GKQ
iKl-z9XbQxAaJ4NYSzgxY4oA; wvr=6; wb_view_log_7568147019=1280*7201.5;
_s_tentry=login.sina.com.cn; Apache=6764926375959.015.1622849984429;
ULV=1622849984625:46:1:1:6764926375959.015.1622849984429:1618522569798
; httpsupgrade_ab=SSL;
webim_unReadCount=%7B%22time%22%3A1622856084593%2C%22dm_pub_total%22%3
A0%2C%22chat_group_client%22%3A0%2C%22chat_group_notice%22%3A0%2C%22al
lcountNum%22%3A12%2C%22msgbox%22%3A0%7D'

}
]
request = requests.session()
request.mount('http://', HTTPAdapter(max_retries=3))
request.mount('https://', HTTPAdapter(max_retries=3))
sem = threading.Semaphore(5)


def weibo(url2, city3):
    with sem:
        page = 50
        count = 0
        while url2:
            print(url2)
            try:
                resopnse = request.get(url2, headers=headers,
verify=False, timeout=10, cookies=random.choice(cookies),
                                          proxies=proxy)
            except requests.exceptions.ConnectionError:
                break
            except requests.exceptions.InvalidURL:
                resopnse = request.get(url2, headers=headers,
verify=False, timeout=10, cookies=random.choice(cookies),
                                          proxies=proxy)
            if '抱歉，未找到' in resopnse.text:
                break
            else:
                parser = etree.HTML(resopnse.text)
                all_data = parser.xpath('//*[@class="card"]')
                for data in all_data:
                    dict1 = {}
                    name = data.xpath('./div[1]/div[2]/div[1]/div[2]/
a[1]')[0].text
                    try:
                        time1 = data.xpath('./div[1]/div[2]/p[2]/
```

```
a[1]')[0].text.replace('\\n', '').replace(' ',

'').replace(
                            '\n',
                            '')
                        ff = bool(re.search(r'\d', time1))
                        if ff:
                            time1 = data.xpath('./div[1]/div[2]/p[2]/
a[1]')[0].text.replace('\\n', '').replace(' ',

'').replace(
                                '\n', '')
                        elif '#' in time1:
                            time1 = data.xpath('./div[1]/div[2]/p[3]/
a[1]')[0].text.replace('\\n', '').replace(' ',

'').replace(
                                '\n', '')
                        else:
                            time1 = data.xpath('./div[1]/div[2]/p[3]/
a[1]')[0].text.replace('\\n', '').replace(' ',

'').replace(
                                '\n', '')
                    except:
                        time1 = data.xpath('./div[1]/div[2]/p[3]/
a[1]')[0].text.replace('\\n', '').replace(' ',

'').replace(
                            '\n',
                            '')
                        if '年' in time1:
                            try:
                                time1 = int(time1[0:4])
                            except:
                                time1 = data.xpath('./div[1]/div[2]/
p[3]/a[1]')[0].text.replace('\\n', '').replace(' ',

'').replace(
                                    '\n', '')
                                time2 = int(time1[0:4])
                                time3 = 2018 - time2
                                if time3 <= 0:
                                    time1 = time1
                                else:
                                    continue
                        else:
                            time1 = time1
                    text = data.xpath('./div[1]/div[2]/p//text()')
                    text = str(text).replace("'\\ue627'",
```

```
'').replace('\\xa0', '').replace("',", '').replace('\\n',

'').replace(
                        ' ', '').replace("'", '').replace('hg,', '',
20).replace("'L'", ' ').replace('\\u200b',

'   ').replace(
                        '[', '').replace(']', '').replace('\\ue40f', '
').replace('\\ue107', '   ')
                    if '展开全文' in text:
                        text1 = data.xpath('./div[1]/div[2]/p[2]//
text()')
                        text1 = str(text1).replace("'\\ue627'",
'').replace('\\xa0', '').replace("',", '').replace(
                            '\\n',
                            '').replace(
                            ' ', '').replace("'", '').replace('hg,',
'', 20).replace("'L'", ' ').replace('\\u200b',

'   ').replace(
                            '[',
                            '').replace(
                            ']', '').replace('\\ue40f', '
').replace('\\ue107', '   ').replace('来自Android客户端',

'').replace(
                            '来自iPhone客户端', '').replace('来自微博手机
版', '').replace('来自微博weibo.com', '')
                        dict1['微博内容'] = text1
                        '''try:
                            dict1['微博来源'] = data.xpath('.//
*[@class="from"]/a[2]')[0]
                        except:
                            dict1['微博来源'] = '暂无来源'''
                        #insertmongo(dict1)
                    else:
                        dict1['微博内容'] = text
                        dict1['微博来源'] = text.split(time1)[1]
                    dict1['用户名'] = name
                    dict1['时间'] = time1
                    dict1['城市'] = city3
                    print(dict1)
                    insertmongo(dict1)
            if '下一页' in resopnse.text:
                try:
                    url2 = 'https://s.weibo.com' + parser.xpath('//
*[@class="next"]/@href')[0]
```

```
                except:
                    break
                count = count + 1
                if count >= page:
                    break
            else:
                break
        '''if len(list1) == 0:
            print('你的cookies已经失效了，请重新登录微博获取')'''

    '''rr = DataFrame(list1)
    df = rr.drop_duplicates(subset=['微博内容', '用户名', '时间', '城市'],
keep='first')
    df = df.reset_index()
    xx1 = df.copy()
    try:
        xx1.sort_values(by="时间", axis=0, ascending=True,
inplace=True)
    except:
        print('你的cookies已经失效了，请重新登录微博获取')
    xx1.to_excel(r'F:\{}_{}_weibo.xlsx'.format(palce1, month),
index=False)'''


def insertmongo(dict):
    client = MongoClient("mongodb://localhost:27017/")
    collection = client['weibo']['weibo']
    collection.insert(dict)


if __name__ == "__main__":
    '''tt = input('请输入关键字:')
    start_time = input('请输入起始时间,格式为:2020-04-08-9:')
    end_time = input('请输入结束时间,格式为:2020-05-08-9:')
    page = int(input('请输入爬取的最大页数:'))
    palce1 = int(input('请输入第一个城市编码:'))
    palce2 = int(input('请输入第二个区域编码,如果你选择不限制地区，输入1000，如
果选择区域，从1开始选择:'))'''
    print('程序默认最大页数是50页')
    month = int(input('请输入月份:'))
    if month < 10:
        month = str(month)
        month = '0' + month
    else:
        month = month
    palce1 = int(input('请输入第一个城市编码:'))
    tty = [0, 1, 2, 3, 4, 5, 6, 23]
```

```
    #list3 =
['01','02','03','04','05','06','07','08','09','10','11','12','13','14'
,'15','16','17','18','19','20','21','22','23','24','25','26','27','28'
,'29','30','31']
    llu = ['睡不着', '失眠', '失眠睡不着']
    city_list = city1.get(palce1)
    # print(city_list[1])
    # print(type(city_list[0]))
    # print(city_list[0])
    list1 = []
    thred = []
    for city in city_list[0]:
            for rre in llu:
                for i in range(1, 10):
                    for u in tty:
                        url5 = 'https://s.weibo.com/weibo?q={}
&region=custom:{}:{' \
                               '}
&typeall=1&suball=1&timescope=custom:2018-{}-0{}-{}:2018-{}-0{}-{}
&Refer=g'.format(
                            rre, palce1, city, month, i, u, month, i,
u + 1)
                        ccy = city_list[0].index(city)
                        tt = threading.Thread(target=weibo,
args=(url5, city_list[1][ccy]))
                        thred.append(tt)
    '''for city in city_list[0]:
        for rre in llu:
            for i in range(10, 21):
                for u in tty:
                    url5 = 'https://s.weibo.com/weibo?q={}
&region=custom:{}:{' \
                           '}
&typeall=1&suball=1&timescope=custom:2014-{}-{}-{}:2014-{}-{}-{}
&Refer=g'.format(
                        rre, palce1, city, month, i, u, month, i, u +
1)
                    ccu = city_list[0].index(city)
                    # tt = threading.Thread(target=weibo, args=(url5,
city_list[1][ccu]))
                    xx = threading.Thread(target=weibo, args=(url5,
city_list[1][ccu]))
                    thred.append(xx)'''
    '''for city in city_list[0]:
        for rre in llu:
            for i in range(21, 31):
                for u in tty:
                    url5 = 'https://s.weibo.com/weibo?q={}
&region=custom:{}:{' \
                           '}
```

```
&typeall=1&suball=1&timescope=custom:2014-{}-{}-{}:2014-{}-{}-{}
&Refer=g'.format(
                        rre, palce1, city, month, i, u, month, i, u +
1)
                    ccu = city_list[0].index(city)
                    # tt = threading.Thread(target=weibo, args=(url5,
city_list[1][ccu]))
                    xx = threading.Thread(target=weibo, args=(url5,
city_list[1][ccu]))
                    thred.append(xx)'''

    for the in thred:
        the.start()
```

Chapter 3

The Effect of Chinese Real-time Air Pollution Monitoring Policy on Corporate Social Responsibility: A Spatial Difference-in-difference Analysis

Abstract

This study examines the impact of the real-time air pollution monitoring (RAPM) policy on the corporate social responsibility (CSR) and corporate environmental responsibility (CER) of listed companies in China, utilizing city-level data and employing various spatial models. The findings suggest that the RAPM policy may not directly affect the CSR or CER scores of treated cities. However, it is observed to decrease the CSR or CER scores of neighboring cities, particularly those that are untreated. Quantitatively, the estimated results from the decomposed spatial Durbin error DID model indicate that the implementation of the RAPM policy does not influence a city's average CSR or CER scores. Nevertheless, it leads to a decrease of -1.509 and -0.350 in the average CSR or CER scores of treated cities' untreated neighboring cities, respectively, while holding other variables constant. This study contributes to the existing literature by highlighting the spatial spillover effects of the RAPM policy. While previous studies focused on the potential positive impact of a policy on the CSR or CER scores of listed companies within the policy-implemented cities, they overlooked the spatially dependent treatment effects. It is noted that listed companies in treated cities may alter their supply chains or relocate manufacturing plants to cities without the RAPM policy, thereby reducing their environmental protection costs. Consequently, this may lead to a decline in the CSR or CER scores of listed companies in cities not covered by the policy.

3.1 Introduction

Since 2008, the Chinese government has mandated listed companies to disclose their corporate social responsibility (CSR) activities. This regulation aims to address social and environmental concerns, alleviate economic development disparities, and foster societal sustainability (Almunawar & Low, 2013; Y.-C. Chen et al., 2018; Kong, Cheng, & Jiang, 2021). For example, Y.-C. Chen et al. (2018) argued that there is a reduction in the amount of industrial wastewater discharge and sulfur dioxide (SO2) emissions in cities following mandatory CSR disclosure. The reason is that mandatory CSR disclosure imposes political and social pressure on companies, which increases companies' efforts in promoting CSR initiatives. Particularly, in China, CSR can be influenced through two main channels: government incentives (such as green financing, environmental blacklists, and CSR awards) and company characteristics (including company size, performance, age, board and CEO characteristics) (Chai, Xie, Yeh, Lan, & Cui, 2022; Y. Zhang, Wang, & Kwon, 2021; Y. Zhang & Zhao, 2022; G. Zhang, 2023; Swandari & Sadikin, 2016).

The assessment criteria for local government are not only economic performance indicators but also social responsibility, resident welfare, unemployment rate, and relative equity. Therefore, besides government encourages and companies' characteristics, government policy can also affect companies' CSR in China. (Kong et al., 2021; Chai et al., 2022; C. Zhang, Liu, Ge, Hao, & Hao, 2021; Y. Zhang & Zhao, 2022). For instance, Kong et al. (2021) argued that the government promotes CSR among companies by creating relevant policies and allocating resources. Y. Zhang and Zhao (2022) considered a real-time air pollution monitoring (RAPM) policy as a quasi-natural experiment and studied the effect of the RAPM policy on companies' CSR and CER scores using the difference-in-difference (DID) model.

In 2012, China introduced the RAMP policy, specifically designed to improve air quality. Its implementation commenced in 74 pilot cities at the beginning of 2013. The RAMP policy mandates local environmental bureaus in each city to publicly disclose their real-time air quality indexes (AQI). These indexes are computed using particulate matter (PM2.5 and PM10) and other air pollutants originating from sources such as traffic, chemical factories, or coal burning. The pollutants encompass sulfur dioxide (SO2), carbon monoxide (CO), nitrogen dioxide (NO2), and ozone (O3) (Cao, Chen, & McIntyre, 2021; Y. Zhang & Zhao, 2022). In contrast to previous air pollution regulations, the RAMP policy allows the central government to directly monitor the air quality of each city, ensuring more effective and accurate monitoring. Unlike earlier regulations that required local environmental bureaus to collect and report local air quality to the central government, the RAMP policy represents a departure from this approach. Under the previous regulation, local governments often manipulated local air quality data to protect local companies from incurring high environmental protection costs (Jin, Andersson, & Zhang, 2016).

In this regard, to better estimate the treatment effect of a policy, Ashenfelter and Card (1984) emphasized that the standard Difference-in-Differences (DID) model serves as a benchmark. The conditional average treatment effect (ATE) of a policy can be estimated under the assumptions of standard DID model (Delgado & Florax, 2015). These assumptions include: 1) Independence of Treatment Assignment, where the assignment of treatment is assumed to be independent of the dependent variables; 2) the Parallel Trends Assumption, positing that in the absence of treatment, both the treated and control groups would evolve with the path trend; and 3) the Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1978; Rubin, 1991), which assumes that the outcomes of one region are independent of the treatment effect of the other regions.

On the other hand, Delgado and Florax (2015) argue that the spatial correlations of socioeconomic variables lead to the violation of SUTVA. The DID estimation becomes biased and inefficient as it neglects spatial correlations of treatment and variables in different cities. They suggest that a spatial DID (SDID) model can be employed to control for spatial correlations in treatments and social interactions in treatment responses. In the SDID model, the Average Treatment Effect (ATE) is divided into the Average Direct Treatment Effect (ADTE) and the Average Indirect Treatment Effect (AITE). The ADTE is the treatment effect on the directly treated cities, while the AITE is the treatment effect on the neighbors of directly treated cities, representing the spatial spillover effects of the treatment.

Furthermore, J. P. Elhorst (2014) explained that the error term may be spatially autocorrelated, indicating that omitted determinants of dependent variables may exhibit spatial autocorrelation or there might be an unobserved shock following a spatial pattern. Therefore, in addition to the DID and SDID models, this paper also constructed a Spatial Durbin Error DID model (SDEM-DID) to control for the spatial autocorrelation of error terms.

At the same time, indirect treatment effects are different between treated and untreated neighboring cities. Chagas, Azzoni, and Almeida (2016) argue that a treated city's indirect treatment effect on its untreated neighbors is higher than on its treated neighbors. This is because the direct treatment effects on a treated city are more crucial than the indirect treatment effects from its treated neighbors. However, the indirect treatment effect on an untreated city arises only from its treated neighbors. Meanwhile, untreated cities do not generate any direct or indirect treatment effects because they have no policy implementation. Therefore, I also decomposed the AITE of the RAPM policy on treated cities' neighboring cities into effects on treated neighboring cities and effects on untreated neighboring cities.

In summary, this paper employs standard DID, SDID, decomposed SDID (DS-DID), SDEM-DID, and decomposed SDEM-DID (DSDEM-DID) models, incorporating two distinct spatial weighted matrices (800 kilometers (KM) binary contiguity and 800KM inverse distance spatial weighted matrices), to analyze the treatment effects of the Real-Time Air Pollution Monitoring (RAPM) policy on city-level average Corporate Social Responsibility (CSR) and Corporate Environmental Responsibility (CER) scores.

This paper makes two significant contributions. First, it conducts the analysis at the city level. While much of the existing literature has primarily focused on the firm level in the study of CSR and CER, this paper recognizes the significance of cities as central hubs for economic and social activities. Consequently, there is a growing importance in investigating CSR and CER at the city level (Y.-C. Chen et al., 2018; G. Zhang, 2023).

Second, the paper employs spatial econometric models to account for spatial correlations in treatment, variables, and error terms. Spatial econometric models help mitigate endogeneity issues arising from the violation of the Stable Unit Treatment Value Assumption (SUTVA) and allow for the analysis of the policy's AITE on the neighboring cities of treated cities.

The results indicate that the RAPM policy does not have a discernible impact on the average CSR or CER scores of treated cities. However, it does lead to a decrease in the CSR or CER scores of the untreated neighboring cities of the treated cities. Quantitatively, the estimated results from the DSDEM-DID model with an 800KM binary contiguity spatial weighted matrix, after controlling for other moderator variables, reveal that the implementation of the RAPM policy does not influence a city's average CSR or CER scores. Still, it does result in a decrease of -1.509 and -0.350 in the average CSR or CER scores of the untreated neighboring cities of the treated cities, with other variables held constant. This finding contrasts with existing studies, such as those by Kong et al. (2021), Chai et al. (2022), and Y. Zhang et al. (2021), which argued that a policy would increase the CSR or CER scores of listed companies located in the cities where the policy is implemented. However, these studies overlooked the spatial spillover effects of the policy.

In the eastern and central regions of China, where most listed companies are concentrated, the estimated results further corroborate that the implementation of the RAPM policy does not lead to an increase in the average CSR and CER scores of treated cities. Instead, it results in a decrease in the average CSR and CER scores of the untreated neighboring cities of the treated cities.

Placebo studies are employed for robustness checks, assuming the implementation of the RAPM policy at the beginning of 2012. The results reveal that both the ADTE and AITE of the RAPM policy on treated cities and their neighbors are all statistically insignificant. This supports the main findings presented in this paper.

The remainder of the paper is structured as follows: section two reviews the relevant literature, section three presents the data and explains the empirical methodology of the study, section four presents and discusses the results, and section five concludes.

3.2 Related Literature

This section aims to review the existing literature about the effect of policy on companies' CSR scores and the spatial difference-in-difference model.

The effect of policy on companies' CSR scores

Y. Zhang and Zhao (2022) examined the impact of the RAPM policy on firm-level CSR and CER scores using the standard Difference-in-Differences (DID) methodology. Their findings suggested significant positive and negative effects of the RAPM policy on the CSR and CER scores of state-owned enterprises (SOEs) and non-state-owned enterprises (non-SOEs) respectively. The overall treatment effects of the RAPM policy led to a reduction in companies' CSR and CER scores. This was attributed to the ability of SOEs to receive environmental subsidies from local governments.

Kong et al. (2021) employed a DID model to examine the impact of local governmental policy incentives on the CSR scores of local companies. Their dataset spanned five years, with varying policy incentives each year. Their regression results led to the conclusion that local governmental policy incentives have a significant positive effect on CSR scores for local companies. Moreover, they found that the influence of policy incentives is more pronounced for companies with low state ownership proportions, but those companies have strong political connections with local government.

Chai et al. (2022) highlighted that policies exert a more substantial influence on companies in China compared to Western countries. They utilized the Propensity Score Matching (PSM) DID model to investigate the impact of the National Civilized City Award (NCCA) on the CSR scores of local companies, using data from listed companies spanning 2012 to 2018 in China. The findings indicated a significant increase in CSR scores for local listed companies due to the NCCA campaign. This was attributed to local government pressure on companies to actively engage in CSR activities, thereby aiding the local government in securing the NCCA title.

Y. Zhang et al. (2021) similarly employed a PSM-DID model to examine the impact of the NCCA campaign on the CER scores of local listed companies. Their findings indicated that CER scores were higher in cities winning the NCCA title compared to non-NCCA cities. Furthermore, they observed that the effect was more pronounced for state-owned companies.

The conclusions drawn from these studies are inconsistent. While Y. Zhang and Zhao (2022) and Y. Zhang et al. (2021) argue that the policy effect is more significant for state-owned companies than for non-state-owned companies, Kong et al. (2021) reach the opposite conclusion, suggesting that the policy effect is greater for non-state-owned companies. This inconsistency could be attributed to the notion that the implementation of a policy is more effective at a city level rather than at company level. Consequently, there is a need to investigate the impact of the policy on the average CSR scores of companies at the city level.

Spatial difference in difference

Bardaka, Delgado, and Florax (2018) employed the SDEM-DID model to assess the causal effect of a 1994 urban rail station investment on gentrification[1]. The results indicated that the construction of an urban rail station in 1994 had a substantial direct treatment effect on the median housing values of homes located within 1 mile of the rail station. Quantitatively, this construction led to a significant 22.26% increase in the median housing values within that radius. Furthermore, the rail station's construction exhibited a notable indirect treatment effect on the median housing values of neighboring homes. Specifically, a 10% increase in the proportion of neighboring houses resulted in a 3.07% increase in the housing values of those neighbors. However, the construction of the rail station only had a significant direct treatment effect on median household income, with insignificant effects on educational attainment and managerial occupation.

Chagas et al. (2016) utilized a DSDEM-DID model to examine the impact of burning fields to facilitate sugarcane harvesting on hospitalizations for respiratory diseases in Sao Paulo. Their dataset encompassed 644 municipalities in Sao Paulo, spanning from 2002 to 2013. Initially using the standard DID model, they observed that burning fields increased respiratory hospitalizations by 0.81 cases per thousand. However, upon adopting the DSDEM-DID model, they found that the direct treatment effect of burning fields on hospitalizations was insignificant. Instead, the indirect treatment effect of burning fields on the hospitalizations of non-burning neighboring regions was substantial, measuring at 1.35 cases per thousand. The total treatment effect of burning fields on hospitalizations was calculated as 1.44 cases per thousand. Remarkably, the treatment effect on non-burning regions contributed 94% to the total treatment effect. This highlights the significance of considering the

[1] They used median household income, educational attainment, managerial occupation, and media housing values to define gentrification.

indirect treatment effect, which was overlooked in the standard DID model, and was found to be higher than the direct treatment effect.

Bardaka et al. (2018) and Chagas et al. (2016) did not explicitly focus on dissecting a policy's direct and indirect treatment effects. Nevertheless, both studies employed spatial DID models for their analyses. Their research outcomes underscored the substantial contribution of the indirect treatment effect to the overall treatment effect. Consequently, employing spatial econometric models becomes essential for scrutinizing the impact of policies on companies' CSR scores.

3.3 Data

The panel dataset comprises city-level data from the developed eastern, less developed northeastern and central, and the least developed western regions of China. In total, the dataset includes 237 cities spanning the years 2010 to 2016. Among these, 73 cities are identified as treated cities implementing the RAPM policy, a determination made by the central government. One treated city (Lhasa) was excluded from the analysis due to missing data.

The dependent variables in this analysis consist of city-level average CSR and CER scores. The firm-level data, encompassing CSR scores and company information, are sourced from the Hexun website. The CSR scores are constructed from five subgroup responsibility scores, which include shareholder responsibility, employee responsibility, supplier and customer responsibility, CER, and social responsibility. Hexun offers CSR scores and subgroup responsibility scores for listed companies in each city, with the weighted aggregation of different subgroup scores used to compute CSR scores. However, Hexun does not provide CER scores for certain companies. Therefore, I need to calculate CER scores for each company based on CSR scores.

The listed companies under consideration span various industries, and the computation of CER scores involves different weightings for companies in distinct sectors. Specifically, the CER scores of companies' core business categorizes into the primary, second, and tertiary industries contributing[2] 20%, 30%, and 10%, respec-

[2]Basing on definitions of National Bureau of Statistics of China, primary industry refers to agri-

tively, to the overall CSR score of each company. Consequently, the calculation of CER scores for all companies is performed by factoring in these industry-specific weightings in the overall CSR scores.

Subsequently, I retained the CSR and CER scores for all listed companies that remained listed throughout the period from 2010 to 2016. These scores were then associated with specific cities based on the longitude and latitude information from the companies' registered offices as provided in the company information dataset. Ultimately, the average CSR and CER scores at the city level were computed by considering the number of listed companies in each city. The unit of measurement for both CSR and CER scores is points.

Additionally, other moderator variables include the natural logarithm of the total city population (log population) and GDP (log GDP), serving as measures of city size and development. Larger and more developed cities often wield greater influence in enforcing environmental protection measures and impacting the CSR scores of listed companies (Tran, Fu, & Boehe, 2023). Following the approach of Chai et al. (2022), the second industry GDP as a proportion of the total GDP (GDP2Ratio) is also controlled for. Fetscherin, Voss, and Gugler (2010) argued that multinational enterprises can introduce global CSR practices to Chinese companies through their local suppliers, customers, and by advising government officials. To account for this influence, the natural logarithm of foreign direct investment (log FDI) is controlled for. However, due to missing foreign direct investment data for some cities, those cities were excluded from the analysis when controlling for log FDI. The total number of cities included in the analysis, accounting for log FDI data, is 222. All these moderator variables are sourced from the China Statistic Yearbooks.

culture, forestry, animal husbandry, fisheries (excluding service industries of agriculture, forestry, animal husbandry and fisheries); secondary industry refers to mining and quarrying (excluding mining support service activities), manufacturing (excluding repair of fabricated metal products, machinery and equipment), production and distribution of electricity, heating power, gas and water, construction; tertiary industry is service industry, and refers to the other industries except primary and secondary industry (excluding international organizations).

Table 3.1: Summary statistics without includes FDI

Variable	Observation	Mean	Std. dev.	Min	Max
CSR	1,659	25.50473	14.19606	-13.55	84.07
CER	1,659	6.756233	4.146008	-4.065	25.221
logpop	1,659	5.874443	0.7744424	2.970414	8.129175
logGDP	1,659	7.31575	0.9231874	4.663439	10.30518
GDP2Ratio	1,659	49.8706	10.35515	17.0211	89.70

Table 3.2: Summary statistics includes FDI

Variable	Observation	Mean	Std. dev.	Min	Max
CSR	1,554	25.66232	13.86105	-13.55	84.07
CER	1,554	6.769605	4.046111	-4.065	25.221
logPOP	1,554	5.967	0.6673602	3.78419	8.129175
logGDP	1,554	7.409652	0.864428	5.359883	10.30518
GDP2Ratio	1,554	49.75825	9.812305	17.0211	82.2417
logFDI	1,554	1.650578	1.135087	0.0015987	5.73417

Tables 2.1 and 2.2 present the summary statistics for various variables. In the FDI dataset, data from 15 cities were excluded due to the absence of FDI data. Nevertheless, the means of their city-level average CSR and CER scores closely resemble those in the dataset before removing these 15 cities.

Figure 3.1: The number of listed companies and treated cities

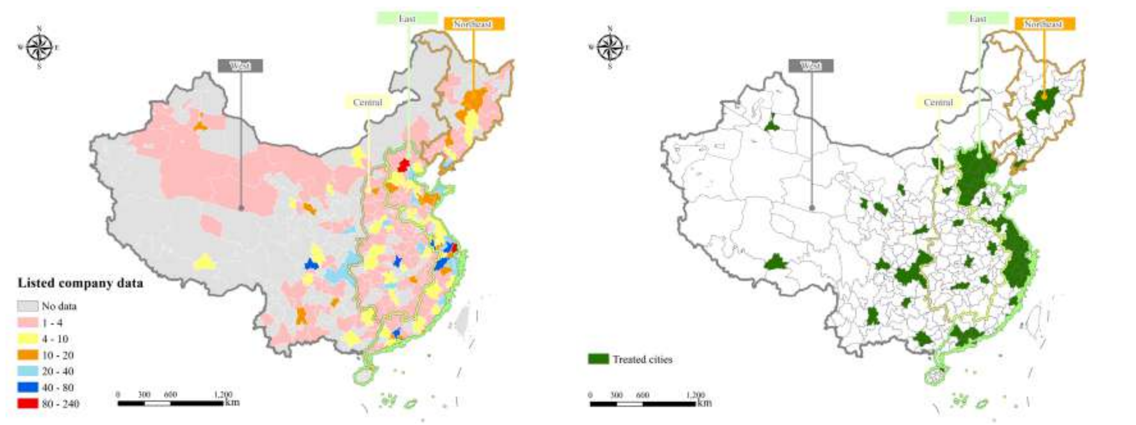


Figure 2.1 displays the number of treated cities and listed companies across different cities. The majority of listed companies and treated cities are concentrated in the eastern and central regions of China. Consequently, an additional analysis of the treatment effects of the RAPM policy on CSR and CER scores was conducted using data specifically from the eastern and central regions of China.

3.4 Empirical Strategy and Results

Standard DID model:

$$y_{it} = \alpha_0 + \alpha_1 DT + \alpha_2 x_{it} + c_i + \zeta_t + \varepsilon_{it}$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

Where y_{it} represents one of the following city-level outcomes in city i at time t: 1) the average Corporate Social Responsibility (CSR) score, and 2) the average Corporate Environment Responsibility (CER) scores. The variable D is a treatment group dummy variable, taking values of 1 and 0 to signify the treatment group (73 pilot cities after the beginning of 2013) and the control group, respectively. Similarly, the variable T is a time dummy variable, taking values of 1 and 0 to indicate post-treatment (after the beginning of 2013, when the RAPM policy started implementation) and pre-treatment, respectively.

The primary explanatory variable is the product of two dummy variables. The first dummy variable signifies the treatment or control group, taking the value of one for treatment and zero for control. The second dummy variable represents the period before or after the implementation of the RAPM policy, with a value of one for after and zero for before.

x_{it} denotes the other moderator variables, including city-level yearly log population (LogPOP), log GDP, second industry GDP weights of total GDP (GDP2Ratio), and log foreign direct investment (LogFDI). c_i and ζ_t are space and time specific effects that control for time and space invariant city characteristics, respectively.

Average treatment effect (ATE):

$$ATE = \{E[y|X = x, D = 1, T = 1] - E[y|X = x, D = 1, T = 0]\}$$

$$\begin{aligned} &- \{E[y|X = x, D = 0, T = 1] - E[y|X = x, D = 0, T = 0]\} \\ &= \alpha_1 \end{aligned}$$

ATE interpretation is RAPM policy lead a pilot city's average CSR score or average CER score increase or decrease by α_1.

Following (Delgado & Florax, 2015), setting spatial DID model for controlling spatial correlations in treatments and social interaction in treatment responses.

Spatial difference-in-difference (SDID) model:

$$\begin{aligned} y_{it} &= \alpha_0+\alpha_1(I + \rho W)DT+\alpha_3 x_{it}+\alpha_4 W x_{it}+c_i+\zeta_t+\varepsilon_{it} \\ &= \alpha_0+\alpha_1 DT+\alpha_1 \rho W DT+\alpha_3 x_{it}+\alpha_4 W x_{it}+c_i+\zeta_t+\varepsilon_{it} \\ &= \alpha_0+\alpha_1 DT+\alpha_2 W DT+\alpha_3 x_{it}+\alpha_4 W x_{it}+c_i+\zeta_t+\varepsilon_{it} \end{aligned}$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

W represents a spatially weighted matrix, where ρ quantifies the strength of spatial dependence between regions. The term $\alpha_2 WDT$ encapsulates spatial interaction in the treatment response, with $\alpha_2 = \alpha_1 * \rho$. Additionally, Wx_{it} signifies a spatially lagged independent variable, representing the exogenous interaction effects between independent variables.

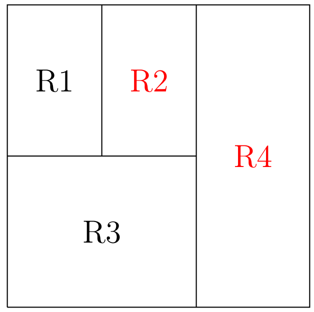
Average treatment effect (ATE):

$$\begin{aligned} ATE(wz) = &\{E[y|X = x, D = 1, T = 1, WD = wd] - E[y|X = x, D = 1, T = 0, WD = wd]\} \\ &- \{E[y|X = x, D = 0, T = 1, WD = wd] - E[y|X = x, D = 0, T = 0, WD = wd]\} \\ &= \alpha_1 + \alpha_3 wd \end{aligned}$$

$$ATE = \alpha_1 + \alpha_3 \overline{WD}$$

$\overline{WD}$ is the average proportion of treated neighbors.

Interpretation of WD:



Assuming regions R1, R2, R3, and R4, where R2 and R4 are treated groups. The neighborhood relationships are as follows: R1's neighbors are R2 and R3. R2's neighbors are R1, R3, and R4. R3's neighbors are R1, R2, and R4. R4's neighbors are R2 and R3.

The spatial weighted matrix C and treatment dummy variables vector D can be expressed as:

$$C = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \qquad D = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

The standardized spatial weighted matrix W and WD can be expressed as:

$$W = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix} \qquad WD = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/3 \\ 2/3 \\ 1/3 \end{bmatrix}$$

WZ_1 (the first element of WD) $= 1/2$, which means that one of the two neighbors of R1 is directly treated. Alternatively, it can be interpreted as the proportion of R1's treated neighbors, equal to 1/2. WZ_2, WZ_3, and WZ_4 have similar interpretations. Hence, $\overline{WD} = \frac{1/2+1/3+2/3+1/3}{4} = \frac{1}{2}$, which is the average proportion of treated neighbors for all units.

ATE can be decomposed into two components: the Average Direct Treatment Effect (ADTE) and the Average Indirect Treatment Effect (AITE). $ADTE = \alpha_1$, indicating that the implementation of the RAPM policy leads to an increase or decrease of α_1 in the average CSR or CER scores in the treated cities. $AITE = \alpha_1 \rho wd = \alpha_3 \overline{WD}$, meaning that the implementation of the RAPM policy leads to an increase or decrease of $\alpha_3 \overline{WD}$ in the average CSR or CER scores in the treated cities' neighboring cities.

Based on the literature of Chagas et al. (2016), I decompose the spatial weighted matrix:

$$W = W_{T,T} + W_{T,NT} + W_{NT,T} + W_{NT,NT}$$

In this equation, an element $W_{i,j}$ in W indicates the spatial effect of region-j on region-i, which includes all regions in treatment and control groups. The items on the right hand side of the W equation can be expressed by:

$$W_{T,T} = D_1 * W * D_1$$

$$W_{T,NT} = D_1 * W * D_2$$

$$W_{NT,T} = D_2 * W * D_1$$

$$W_{NT,NT} = D_2 * W * D_2$$

D_1 is an n*n matrix, the diagonal elements of D_1 are one if the city is a treated city. D_2's diagonal elements are one if the city is an untreated city. Using the above example includes R1, R2, R3, and R4. R1's neighbors are R2 and R3. R2's neighbors are R1, R3, and R4. R3's neighbors are R1, R2, and R4. R4's neighbors are R2 and R3. R2 and R4 are treated groups.

$$D_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad D_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$W_{T,T} = D_1 * W * D_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 1/2 & 1/2 & 1 \end{bmatrix} * \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \end{bmatrix}$$

In the context of the matrix elements, the intersection of the second row and the fourth column represents the scenario where one of R2's three neighbors is a treated region, implying an indirect treatment effect on R2. Similarly, the element at the intersection of the fourth row and the second column signifies that one of R4's two neighbors is a treated region, resulting in an indirect treatment effect on R4. Following similar rules, $W_{T,NT}$, $W_{NT,T}$, and $W_{NT,NT}$ can be calculated, and their elements have the same interpretation as $W_{T,T}$.

In summary, $W_{T,T}$ represents the spatial weighted matrix for treated cities, including only treated neighboring cities. $W_{T,NT}$ represents the spatial weighted matrix for treated cities, including only untreated neighboring cities. $W_{NT,T}$ represents the spatial weighted matrix for untreated cities, including only treated neighboring cities. $W_{NT,NT}$ represents the spatial weighted matrix for untreated cities, including only untreated neighboring cities. Each of these matrices provides information about the spatial relationships between treated and untreated cities based on their treatment status.

In decomposed SDID (DSDID) model, only $W_{T,T}$ and $W_{NT,T}$ are included because untreated cities have no direct or indirect treatment effects.

Decomposed spatial difference-in-difference (DSDID) model:

$$y_{it} = \alpha_0 + \alpha_1 DT + \alpha_2 W_{T,T} DT + \alpha_3 W_{NT,T} DT + \alpha_4 x_{it} + \alpha_5 W x_{it} + c_i + \zeta_t + \varepsilon_{it}$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

The ATE of the DSDID model can be decomposed into three components: $ADTE$, $AITE_1$, and $AITE_2$. $ADTE$=α_1 represents the average direct treatment effect. $AITE_1$=$\alpha_2\overline{W_{T,T}D}$ shows whether the RAPM policy leads to an increase or decrease in the average CSR or CER scores in a treated city's neighboring treated cities by $\alpha_2\overline{W_{T,T}D}$. Similarly, $AITE_2$=$\alpha_3\overline{W_{NT,T}D}$ indicates whether the RAPM policy leads to an increase or decrease in a treated city's neighboring untreated cities' average CSR or CER scores by $\alpha_3\overline{W_{NT,T}D}$.

To control for spatial autocorrelations in error terms, in addition to the DID and SDID models, this paper also constructed a spatial Durbin error DID model and a decomposed spatial Durbin error DID model.

Spatial Durbin error term DID (SDEM-DID) model:

$$y_{it} = \alpha_0+\alpha_1 DT+\alpha_2 WDT+\alpha_3 x_{it}+\alpha_4 Wx_{it}+c_i+\zeta_t+\mu_{it}$$

$$\mu_{it} = \lambda W\mu_{it} + \epsilon_{it}$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

Decomposed SDEM-DID (DSDEM-DID) model:

$$y_{it} = \alpha_0+\alpha_1 DT+\alpha_2 W_{T,T}DT+\alpha_3 W_{NT,T}DT+\alpha_4 x_{it}+\alpha_5 Wx_{it}+c_i+\zeta_t+\mu_{it}$$

$$\mu_{it} = \lambda W\mu_{it} + \epsilon_{it}$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

λ measure the strength of spatial dependence on error items. The other items, AITE, ADTE of SDEM-DID and DSDEM-DID models are the same as those in the SDID and DSDID models.

3.5 Results

2.5.1 Empirical Results

Figure 3.2: CSR and CER scores parallel trends

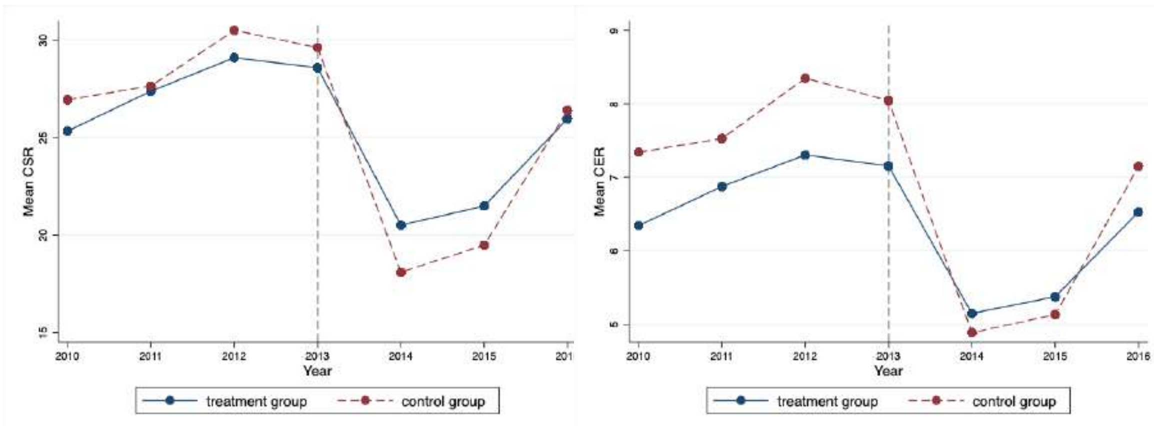


Figure 2.2 illustrates the parallel trends in city-level average Corporate Social Responsibility (CSR) and Carbon Emission Reduction (CER) scores. Before 2013, both the treatment and control groups exhibit similar trends, with the control group consistently maintaining higher scores than the treatment group. Starting from 2013, a noticeable decline occurs in the CSR and CER scores of both groups. One plausible explanation for this decline is the potential implementation of a more stringent air quality standard under the RAPM policy, rendering emissions management by listed firms inadequate to meet these heightened standards. Consequently, CSR and CER scores experience a decrease in 2013. However, from 2014 onward, there is a discernible increase in CSR and CER scores for both groups. Notably, the treatment group's scores remain consistently higher than those of the control group during this period, suggesting that firms within the treatment group may have enhanced their emissions management practices after a one-year preparation period.

Figure 3.3: Event study of CSR (left figure) and CER (right figure) scores

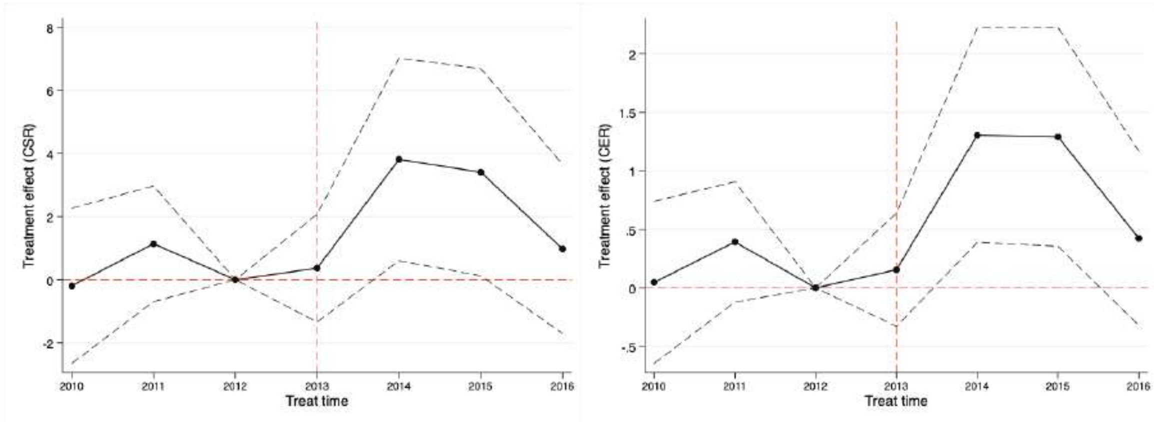


Figure 2.3 presents the event study of average CSR and CER scores at the city level, with the year 2012 serving as the reference period. Before the treatment, the difference in CSR and CER scores between the treatment group and the control group was insignificant compared to the reference period, indicating that the CSR and CER scores of both groups were largely parallel. However, after 2013, a noticeable increasing trend in the difference of CSR and CER scores between the treatment and control groups emerges, compared to the reference period. This suggests that the RAPM policy has a positive effect on the CSR and CER scores of listed companies. It's noteworthy that the treatment effect of the RAPM policy diminishes in 2016.

This paper utilizes two distinct spatial weighted matrices for the spatial analysis. The first matrix is the 800KM distance binary contiguity matrix, where a city's neighbors are defined as those within 800KM. Before standardization, elements in this matrix are set to one if two cities are neighbors and 0 otherwise. The second matrix is the 800KM inverse distance spatial weighted matrix, where elements represent the inverse distance between a city and its neighbors before standardization, and 0 otherwise. The diagonal elements of both matrices are set to zero since a city cannot be a neighbor to itself.

The direct and indirect causal effects, using 800KM[3] contiguity spatial weighted matrix for the spatial econometrics analysis, of the RAPM policy on CSR and

[3] The rationale for using 800 kilometers as a threshold is that, within this distance, every city can be considered to have at least one neighboring city.

CER scores is reported in Tables 2.3 to 2.7 and Tables 2.9 to 2.12, respectively. The estimated results using 800KM inverse distance matrix are reported in Tables 2.13 to 2.20. Standard errors (SE), clustered at the city level, for the estimated coefficients are presented in brackets.

Table 2.3 presents the estimated results of standard DID, SDID, DSDID, SDEM-DID, and DSDEM-DID models using the 800KM binary contiguity spatial weighted matrix. In Table 2.4, the estimated results controlling for the other moderator variables are displayed.

Table 3.3: CSR results of spatial models with contiguity matrix without controlling the other variables (no FDI)

Spatial matrix:	237 cities	800KM contiguity		800KM contiguity	
Dependent variable: CSR	DID	SDID	DSDID	SDEM-DID	DSDEM-DID
No FDI	(1)	(2)	(3)	(4)	(5)
did	1.828*** (0.909)	1.847** (0.927)	1.682 (2.691)	1.434 (0.912)	-0.727 (2.818)
WD	N	-1.028 (0.906)	N	-2.106** (0.668)	N
WttD	N	N	-0.022 (0.264)	N	-0.344 (0.475)
WnttD	N	N	-0.106 (0.869)	N	-1.542*** (0.470)
λ	N	N	N	0.681*** (0.048)	0.678*** (0.048)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

In Table 2.3, employing the standard DID model for analysis reveals that the RAPM policy has a statistically significant (at the 5% level) positive treatment effect on the CSR scores of treated cities, leading to an increase of 1.8 points in their average CSR score. The SDID model's estimation also shows a significant and positive average direct treatment effect (ADTE) of 1.8 (significant at the 5% level), but the average indirect treatment effect (AITE) on the CSR scores of treated cities' neighboring cities is not statistically significant. However, in the DSDID model, both ADTE and AITE of the RAPM policy are found to be insignificant.

After controlling for spatial correlations in error terms, the SDEM-DID model estimation indicates that the ADTE of the RAPM policy on treated cities' average CSR scores is not statistically significant. However, the AITE of the RAPM policy on treated cities' neighboring cities' average CSR scores is negative and significant at the 1% level, with a value of -2.1[4]. This suggests that while the implementation of the RAPM policy may not affect the average CSR scores of treated cities, it leads to a significant decrease of 2.1 points in the average CSR scores of treated cities' neighboring cities.

In the DSDEM-DID model estimation, the ADTE of the RAPM policy on treated cities' average CSR scores remains insignificant. The AITE of RAPM on treated cities' treated neighboring cities is also insignificant, but on treated cities' untreated neighboring cities, the value is -1.5, and it is significant at the 1% level. This implies that the implementation of the RAPM policy does not affect the average CSR scores of treated cities and their neighboring treated cities, but it does lead to a significant decrease of 1.5 points in the average CSR scores of treated cities' untreated neighboring cities.

---

[4] Values of all estimated AITE and SEs need to do a calculation after SDID, DSDID, SDEM-DID and DSDEM-DID models' original estimation. Because the estimated AITE depends on $\overline{WD}$, which is the average proportion of treated neighbors (Delgado & Florax, 2015). For example, the originally estimated AITE of RAPM in SDEM-DID in table 3 is -12.63316 and direct SE is 4.003488, the $\overline{WD}$ is 0.1667408. Hence, the estimated AITE= -12.663316*0.1667408 = -2.106463, the $SE = \sqrt{4.003488^2 * 0.1667408^2} = 0.6675448$.

Table 3.4: CSR results of spatial models with contiguity matrix controlling the other variables (no FDI)

Spatial matrix:	237 cities	800KM contiguity		800KM contiguity	
Dependent variable: CSR	DID	SDID	DSDID	SDEM-DID	DSDEM-DID
No FDI	(1)	(2)	(3)	(4)	(5)
did	1.767* (0.905)	1.770* (0.924)	2.026 (2.747)	1.301 (0.946)	-0.021 (2.895)
WD	N	0.018 (0.839)	N	-2.128*** (0.679)	N
WttD	N	N	-0.025 (0.295)	N	-0.486 (0.482)
WnttD	N	N	0.044 (0.802)	N	-1.509*** (0.482)
logPOP	-4.297 (6.186)	-4.075 (6.136)	-4.106 (6.178)	1.197 (1.275)	1.161 (1.277)
logGDP	4.452 (2.753)	4.472 (2.732)	4.470 (2.736)	0.829 (1.044)	0.868 (1.047)
GDP2Ratio	0.010 (0.130)	0.0148 (0.129)	0.015 (0.129)	0.138** (0.056)	0.136** (0.056)
WlogPOP	N	-2.681 (2.512)	-2.702 (2.477)	-4.762 (3.558)	-4.458 (3.611)
WlogGDP	N	3.010 (2.035)	3.020 (2.037)	4.837 (2.962)	4.647 (2.987)
WGDP2Ratio	N	-0.055 (0.179)	-0.054 (0.179)	0.058 (0.243)	0.064 (0.243)
λ	N	N	N	0.677*** (0.048)	0.676*** (0.048)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

The estimated results, after accounting for other city-level moderator variables, are presented in Table 2.4. The significance and values of the estimated average direct treatment effect (ADTE) and average indirect treatment effect (AITE) of the RAPM policy closely align with the estimations in Table 2.3.

The other variables in the models exhibit insignificant direct or indirect effects on the average CSR scores of treated cities, except for log second industry GDP weights of total GDP (GDP2Ratio) in the SDEM-DID and DSDEM-DID models.

Notably, an increase in a city's GDP2Ratio has a statistically significant (at the 5% level) positive effect on its average CSR scores. Quantitatively, a 10% rise in a city's second industry GDP weights of total GDP results in a statistically significant increase of 0.014 points in its average CSR scores.

Based on the above analysis, it appears that the SDEM-DID and DSDEM-DID models are more suitable for the analysis. This conclusion is drawn considering that the spatial autocorrelation coefficients λ are all statistically significant at the 1% level, indicating the presence of significant spatial autocorrelations in the error terms. The SDEM model proves effective in controlling spatial autocorrelation between error terms, accounting for factors like omitted variables that might exhibit spatial autocorrelation. In contrast, SDID and standard DID models lack this capability. This observation might explain why the AITEs of the RAPM on treated cities' neighboring cities are significant in the SDME-DID model but insignificant in the SDID model. Consequently, for the subsequent analysis, only the estimations of SDEM-DID and DSDEM-DID models will be reported.

Table 3.5: CSR results of spatial models with contiguity matrix (with FDI)

Spatial matrix:	800KM contiguity		800KM contiguity	
Regions: 222 cities	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CSR	(1)	(2)	(3)	(4)
did	1.320 (0.898)	-1.710 (3.198)	1.215 (0.934)	-1.570 (3.350)
WD	-1.957*** (0.720)	N	-1.538 (0.944)	N
WttD	N	-0.155 (0.559)	N	-0.093 (0.579)
WnttD	N	-1.479*** (0.503)	N	-1.248* (0.676)
logPOP	N	N	0.328 (1.451)	0.335 (1.452)
logGDP	N	N	0.459 (1.323)	0.461 (1.323)
GDP2Ratio	N	N	0.182*** (0.061)	0.180*** (0.061)
logFDI	N	N	0.644 (0.650)	0.682 (0.651)
WlogPOP	N	N	-4.312 (6.231)	-4.815 (6.252)
WlogGDP	N	N	-2.959 (5.681)	-1.972 (5.772)
WGDP2Ratio	N	N	-0.011 (0.268)	-0.026 (0.268)
WlogFDI	N	N	2.034 (3.410)	1.546 (3.450)
λ	0.701*** (0.051)	0.698*** (0.051)	0.701*** (0.051)	0.698*** (0.051)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Table 2.5 presents estimations from the SDEM-DID and DSDEM-DID models, utilizing a dataset that includes foreign direct investment (FDI). Due to missing FDI data, 15 cities were excluded from the analysis. ADTEs of the RAPM on treated cities' average CSR consistently appear to be insignificant in the estimations of both models.

For the SDEM-DID model, estimated AITE for the RAPM on treated cities' neighboring cities' average CSR scores are significant at the 1% level (with a value of -1.957) before controlling for other explanatory variables, but AITE become insignificant after incorporating these controls.

However, in the estimations of the DSDEM-DID models, the AITEs of the RAPM on treated cities' untreated neighboring cities' average CSR scores remain significant at 1% (with a value of -1.479) and 10% (with a value of -1.248) levels before and after controlling for other explanatory variables, respectively. All other control variables' estimations remain insignificant, except for GDP2Ratio. Notably, foreign direct investment appears to have an insignificant direct or indirect impact on a city's average CSR scores, suggesting that in China, the influence of government policies and economic conditions on companies may outweigh the impact of investors.

Table 3.6: CSR results of spatial models with contiguity matrix (no FDI) in eastern and central regions

Spatial matrix:	800KM contiguity		800KM contiguity	
Regions: East Central	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CSR	(1)	(2)	(3)	(4)
did	1.453 (0.986)	-3.872 (5.175)	0.957 (1.020)	-5.500 (5.347)
WD	-2.62* (0.915)	N	-3.321*** (1.226)	N
WttD	N	0.038 (0.986)	N	-0.072 (1.006)
WnttD	N	-1.884*** (0.626)	N	-2.485*** (0.848)
logPOP	N	N	0.100 (1.580)	0.212 (1.584)
logGDP	N	N	1.683 (1.152)	1.693 (1.153)
GDP2Ratio	N	N	0.145** (0.069)	0.150** (0.069)
WlogPOP	N	N	3.061 (8.349)	2.625 (8.385)
WlogGDP	N	N	5.516 (6.199)	6.198 (6.256)
WGDP2Ratio	N	N	0.011 (0.340)	-0.076 (0.351)
λ	0.750*** (0.056)	0.752*** (0.056)	0.741*** (0.057)	0.745*** (0.057)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Tables 2.6 and 2.7 present the estimated results using data from the eastern and central regions of China. In Table 2.7, 2 cities were excluded due to missing foreign direct investment (FDI) data.

Consistent with the estimations using the entire dataset, the ADTE of RAPM on treated cities' average CSR scores are all insignificant in all models. The AITEs of RAPM on treated cities' neighboring and untreated neighboring cities' average CSR scores are all significant, with negative signs. However, the extent of AITEs in the eastern and central regions are larger than those for the entire country, likely influenced by the concentration of most listed companies in these regions.

Table 3.7: CSR results of spatial models with contiguity matrix (with FDI) in eastern and central regions

Spatial matrix:	800KM contiguity		800KM contiguity	
Regions: East Central	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CSR	(1)	(2)	(3)	(4)
did	1.461 (0.993)	-3.292 (5.194)	0.899 (1.028)	-5.730 (5.398)
WD	-2.563*** (0.915)	N	-3.532** (1.245)	N
WttD	N	-0.045 (1.004)	N	-0.115 (1.035)
WnttD	N	-1.806*** (0.616)	N	-2.597*** (0.854)
logPOP	N	N	-0.083 (1.630)	0.065 (1.636)
logGDP	N	N	1.916 (1.484)	1.838 (1.487)
GDP2Ratio	N	N	0.150** (0.069)	0.155** (0.069)
logFDI	N	N	-0.271 (0.784)	-0.192 (0.787)
WlogPOP	N	N	-3.732 (9.060)	-4.446 (9.106)
WlogGDP	N	N	15.699* (8.676)	16.865* (8.778)
WGDP2Ratio	N	N	0.300 (0.390)	0.227 (0.397)
WlogFDI	N	N	-7.869 (5.131)	-8.305 (5.156)
λ	0.745*** (0.057)	0.745*** (0.057)	0.734*** (0.059)	0.738*** (0.059)
Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.				

Tables 2.9 through 2.12 (in the Appendix) present the direct and indirect treatment effects of the RAPM policy on average CER scores for treated cities and

their neighboring cities. The signs and significance of these estimated results mirror those of the models examining the direct and indirect treatment effects of the RAPM policy on CSR scores. However, it's noteworthy that the values of the direct and indirect treatment effects on CER scores are smaller than those on CSR scores. This difference may be attributed to the fact that average city-level CER scores tend to be lower than average CSR scores.

Tables 2.13 to 2.20 (in the Appendix) present the estimated ADTEs and AITEs of the RAPM on treated cities' and their neighboring cities' average CSR and CER scores, using the 800KM inverse distance matrix to estimate SDEM-DID and DSDEM-DID models. ADTEs are consistently found to be insignificant, while AITEs on neighboring cities and untreated neighboring cities are significant only when analyzing data for the entire regions and without controlling for other variables. However, when analyzing data specifically for the eastern and central regions, all ADTEs and AITEs are found to be insignificant. This discrepancy in results may be attributed to the influence of different spatial weighted matrices on the estimation process.

2.5.2 Robustness check

As part of the robustness check, placebo studies were conducted, assuming that the RAPM was implemented at the beginning of 2012. Utilizing SDEM-DID and DSDEM-DID models with an 800KM contiguity spatial weighted matrix for estimation, the results are presented in Table 2.8. Both when not controlling for and controlling for other explanatory variables, the ADTE and AITE of the RAPM on treated cities and their neighbors are found to be consistently insignificant. These results align with the main findings of this paper.

Table 3.8: Robustness check

Spatial matrix:	800KM contiguity		800KM contiguity	
Regions: 237 cities	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CSR	(1)	(2)	(3)	(4)
did	1.018	0.248	0.446	0.530
	(1.030)	(2.362)	(1.076)	(2.372)
WD	-0.304	N	1.134	N
	(0.829)		(0.953)	
WttD	N	-0.009	N	0.427
		(0.437)		(0.462)
WnttD	N	-0.337	N	0.714
		(0.652)		(0.729)
λ	0.718***	0.716***	0.672***	0.672
	(0.043)	(0.043)	(0.048)	(0.048)
Dependent variable: CER	(1)	(2)	(3)	(4)
did	0.197	-0.072	0.160	0.285
	(0.296)	(0.672)	(0.307)	(0.673)
WD	-0.084	N	0.431	N
	(0.233)		(0.265)	
WttD	N	0.007	N	0.149
		(0.125)		(0.131)
WnttD	N	-0.102	N	0.290
		(0.180)		(0.201)
λ	0.703***	0.701***	0.648***	0.649***
	(0.044)	(0.044)	(0.050)	(0.050)
Control variables	N	N	Y	Y

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

3.6 Conclusion

This paper examines the impact of the Real-Time Air Pollution Monitoring (RAPM) policy on the Corporate Social Responsibility (CSR) and Corporate Environmental Responsibility (CER) of listed companies in China. The analysis is conducted using city-level data and employs various spatial models.

Spatial models offer the advantage of analyzing both the direct and indirect treatment effects of a policy on a city, unlike the standard DID model, which only examines direct treatment effects (Delgado & Florax, 2015). In this study, spatial models such as SDID, DSDID, SDEM-DID, and DSDEM-DID are employed. The estimated spatial autocorrelation coefficients indicate that the SDEM-DID and DSDEM-DID models are more suitable for the analysis.

Using the SDEM-DID model, 800KM contiguity spatial weighted matrix, and analyzing the entire dataset before controlling for other moderator variables and FDI, the implementation of RAPM does not seem to affect the average CSR or CER scores of treated cities. However, it leads to a decrease of 2.1 and 0.6 (as shown in the second row of column (4) in Table 2.3 and the second row of column (1) in Table 2.9) points in the average CSR and CER scores of treated cities' neighboring cities. When employing the DSDEM-DID model to estimate, ADTEs of RAPM on treated cities remain insignificant, and AITEs on treated cities' treated neighboring cities' average CSR and CER are also insignificant. However, the AITEs of RAPM on treated cities' untreated neighboring cities' average CSR and CER are significant and negative. This aligns with the argument presented by Chagas et al. (2016), suggesting that after the implementation of RAPM, the AITE of the RAPM policy is stronger on the average CSR and CER scores of treated cities' untreated neighboring cities than on the average CSR and CER scores of treated cities' treated neighboring cities. The average CSR and CER scores of treated cities' untreated neighboring cities witness a decrease of 1.5 points (fourth row of column (5) in Table 2.3) and 0.4 points (fourth row of column (2) in Table 2.9), respectively.

The AITE of the RAPM policy on CSR and CER scores of treated cities' neighboring cities is lower in the DSDEM-DID model than in the SDEM-DID model. This difference arises because the AITE of the RAPM in the SDEM-DID model includes RAPM's spatial spillover effects on all neighbors of treated cities, while the RAPM's spatial spillover effects in the DSDEM-DID model are only on treated cities' untreated neighbors.

After controlling for other moderator variables but without considering FDI, the DSDEM-DID model still shows that the implementation of RAPM has insignificant ADTE and AITE on treated cities' and their treated neighboring cities' average CSR and CER scores. However, the AITEs of RAPM on treated cities' untreated neighboring cities' average CSR and CER scores remain significant. After introducing control for FDI, the AITEs of RAPM on treated cities' and treated cities' untreated neighboring cities' average CSR and CER scores become insignificant.

When utilizing data from the eastern and central regions and controlling for all other moderator variables, including FDI, the DSDEM-DID models indicate that the AITEs of RAPM on treated cities' and treated cities' untreated neighboring cities' average CSR and CER scores are significant and negative.

When employing the 800KM inverse distance spatial weighted matrix for analysis, the DSDEM-DID models show that the AITEs of RAPM on treated cities' and treated cities' untreated neighboring cities' average CSR and CER scores are significant only when not controlling for other variables and using the entire dataset for analysis. The AITEs of RAPM become insignificant when focusing on the eastern and central regions.

In summary, the implementation of the real-time air pollution monitoring policy does not have a significant impact on the CSR or CER scores of the treated cities. However, it does lead to a decrease in the CSR or CER scores of the treated cities' neighboring cities, particularly those untreated neighboring cities, whether considered on a national scale or within the eastern and central regions. This phenomenon suggests that companies located in the treated cities may respond to the policy by changing suppliers or relocating manufacturing plants to cities without such stringent environmental monitoring. Consequently, these companies could reduce their environmental protection costs. However, this shift may negatively impact the CSR or CER scores of listed companies in cities that have not implemented the policy.

The implications of this research for policymakers underscore the importance of considering the spatial spillover effects when implementing new regulations. While a policy may bring positive effects to a specific city, it's crucial to recognize that its impact on neighboring cities might be negative. These adverse spillover effects arise due to spatial autocorrelations in economic or demographic factors across different cities. As encapsulated in Tobler's first law – 'everything is related to everything else, but nearby things are more related than distant things' (Tobler (1970), p. 236) – policymakers should be attentive to the interconnectedness of nearby regions and carefully assess the potential ripple effects of their policies beyond the immediate implementation area.

Second, to bolster CSR or CER scores and mitigate negative spillover effects, local governments across China can provide subsidies to companies to enhance their environmental performance.

Thirdly, Chinese local governments could incentivive companies to publicly disclose further CSR and CER information on their websites or in annual reports, making it easier for residents to monitor companies' environmental performances.

3.7 Appendix

Table 3.9: CER results of spatial models with contiguity matrix (no FDI)

Spatial matrix:	800KM contiguity		800KM contiguity	
Regions: 237 cities	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CER	(1)	(2)	(3)	(4)
did	0.416 (0.261)	-0.149 (0.802)	0.508* (0.269)	0.312 (0.823)
WD	-0.596*** (0.183)	N	-0.503*** (0.187)	N
WttD	N	-0.104 (0.135)	N	-0.133 (0.137)
WnttD	N	-0.432*** (0.129)	N	-0.350*** (0.132)
logPOP	N	N	0.655 (0.373)	0.650* (0.374)
logGDP	N	N	-0.267 (0.305)	-0.261 (0.306)
GDP2Ratio	N	N	0.064*** (0.016)	0.064*** (0.016)
WlogPOP	N	N	-1.269 (1.006)	-1.224 (1.021)
WlogGDP	N	N	1.304 (0.838)	1.276 (0.845)
WGDP2Ratio	N	N	0.040 (0.068)	0.040 (0.068)
λ	0.663*** (0.049)	0.660*** (0.050)	0.658*** (0.050)	0.657*** (0.050)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Table 3.10: CER results of spatial models with contiguity matrix (with FDI)

Spatial matrix:	800KM contiguity		800KM contiguity	
Regions: 222 cities	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CER	(1)	(2)	(3)	(4)
did	0.399 (0.255)	-0.364 (0.907)	0.481* (0.264)	-0.086 (0.948)
WD	-0.557*** (0.200)	N	-0.377 (0.263)	N
WttD	N	-0.059 (0.159)	N	-0.040 0.164)
WnttD	N	-0.413*** (0.139)	N	-0.295 (0.188)
logPOP	N	N	0.397 (0.426)	0.398 (0.426)
logGDP	N	N	-0.332 (0.384)	-0.332 (0.384)
GDP2Ratio	N	N	0.079*** (0.018)	0.079*** (0.018)
logFDI	N	N	0.172 (0.186)	0.180 (0.186)
WlogPOP	N	N	-1.154 (1.801)	-1.257 (1.807)
WlogGDP	N	N	-0.197 (1.588)	0.004 (1.614)
WGDP2Ratio	N	N	0.009 (0.075)	0.005 (0.075)
WlogFDI	N	N	0.103 (0.966)	0.004 (0.977)
λ	0.696*** (0.052)	0.692*** (0.052)	0.687*** (0.053)	0.685*** (0.053)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Table 3.11: CER results of spatial models with contiguity matrix (no FDI) in eastern and central regions

Spatial matrix:	800KM contiguity		800KM contiguity	
Regions: East Central	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CER	(1)	(2)	(3)	(4)
did	0.385 (0.279)	-1.052 (1.455)	0.366 (0.287)	-1.313 (1.499)
WD	-0.653*** (0.247)	N	-0.753** (0.335)	N
WttD	N	0.033 (0.279)	N	0.024 (0.283)
WnttD	N	-0.469*** (0.166)	N	-0.569** (0.229)
logPOP	N	N	0.476 (0.448)	0.504 (0.450)
logGDP	N	N	-0.083 (0.328)	-0.080 (0.328)
GDP2Ratio	N	N	0.075*** (0.020)	0.077*** (0.020)
WlogPOP	N	N	1.764 (2.332)	1.639 (2.340)
WlogGDP	N	N	1.283 (1.700)	1.453 (1.712)
WGDP2Ratio	N	N	-0.018 (0.092)	-0.040 (0.094)
λ	0.732*** (0.060)	0.732*** (0.060)	0.720*** (0.061)	0.723*** (0.061)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Table 3.12: CER results of spatial models with contiguity matrix (with FDI) in eastern and central regions

Spatial matrix:	800KM contiguity		800KM contiguity	
Regions: East Central	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CER	(1)	(2)	(3)	(4)
did	0.388 (0.281)	-0.912 (1.462)	0.356 (0.289)	-1.393 (1.514)
WD	-0.646*** (0.248)	N	-0.797** (0.339)	N
WttD	N	0.010 (0.284)	N	0.022 (0.291)
WnttD	N	-0.455*** (0.165)	N	-0.593** (0.230)
logPOP	N	N	0.419 (0.463)	0.457 (0.465)
logGDP	N	N	-0.067 (0.420)	-0.089 (0.421)
GDP2Ratio	N	N	0.077*** (0.020)	0.078*** (0.020)
logFDI	N	N	-0.034 (0.221)	-0.012 (0.222)
WlogPOP	N	N	0.218 (2.530)	0.024 (2.541)
WlogGDP	N	N	3.755 (2.379)	4.045 (2.401)
WGDP2Ratio	N	N	0.054 (0.106)	0.036 (0.108)
WlogFDI	N	N	-1.966 (1.426)	-2.077 (1.431)
λ	0.727*** (0.061)	0.727*** (0.061)	0.713*** (0.063)	0.716*** (0.062)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Table 3.13: CSR results of spatial models with inverse distance matrix (no FDI)

Spatial matrix:	800KM inverse distance		800KM inverse distance	
Regions: 237 cities	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CSR	(1)	(2)	(3)	(4)
did	1.355 (0.994)	-1.382 (2.239)	0.947 (1.023)	-0.990 (2.290)
WD	-1.239** (0.616)	N	-0.351 (0.754)	N
WttD	N	-0.108 (0.359)	N	0.093 (0.377)
WnttD	N	-1.119** (0.441)	N	-0.520 (0.556)
logPOP	N	N	0.776 (1.506)	0.832 (1.507)
logGDP	N	N	1.453 (1.160)	1.443 (1.160)
GDP2Ratio	N	N	0.122** (0.058)	0.122** (0.058)
WlogPOP	N	N	2.817 (3.655)	2.213 (3.692)
WlogGDP	N	N	-4.757 (3.425)	-4.132 (3.464)
WGDP2Ratio	N	N	0.280 (0.175)	0.254 (0.176)
λ	0.663*** (0.045)	0.650*** (0.046)	0.646*** (0.046)	0.639*** (0.047)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Table 3.14: CSR results of spatial models with inverse distance matrix (with FDI)

Spatial matrix:	800KM inverse distance		800KM inverse distance	
Regions: 222 cities	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CSR	(1)	(2)	(3)	(4)
did	1.237 (0.975)	-1.811 (2.448)	0.939 (1.006)	-1.088 (2.562)
WD	-1.129* (0.638)	N	-0.085 (0.782)	N
WttD	N	-0.014 (0.405)	N	0.215 (0.422)
WnttD	N	-1.062** (0.453)	N	-0.355 (0.582)
logPOP	N	N	-0.344 (1.592)	-0.290 (1.594)
logGDP	N	N	1.115 (1.391)	1.096 (1.391)
GDP2Ratio	N	N	0.157** (0.062)	0.158** (0.062)
logFDI	N	N	0.727 (0.689)	0.754 (0.690)
WlogPOP	N	N	0.440 (4.999)	-0.066 (5.019)
WlogGDP	N	N	-6.257 (4.725)	-5.178 (4.852)
WGDP2Ratio	N	N	0.273 (0.221)	0.247 (0.221)
WlogFDI	N	N	0.364 (2.793)	-0.082 (2.825)
λ	0.683*** (0.047)	0.668*** (0.049)	0.668*** (0.048)	0.660*** (0.049)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Table 3.15: CSR results of spatial models with inverse distance matrix (no FDI) in eastern and central regions

Spatial matrix:	800KM inverse distance		800KM inverse distance	
Regions: east central	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CSR	(1)	(2)	(3)	(4)
did	1.155 (1.151)	-2.466 (3.325)	0.502 (1.173)	-1.431 (3.397)
WD	-0.975 (0.884)	N	0.311 (1.090)	N
WttD	N	0.131 (0.613)	N	0.400 (0.644)
WnttD	N	-0.933 (0.569)	N	-0.074 (0.729)
logPOP	N	N	-1.044 (1.669)	-0.980 (1.675)
logGDP	N	N	2.349* (1.266)	2.339* (1.267)
GDP2Ratio	N	N	0.114 (0.071)	0.115 (0.071)
WlogPOP	N	N	3.189 (6.572)	3.250 (6.556)
WlogGDP	N	N	-5.512 (4.946)	-5.188 (4.962)
WGDP2Ratio	N	N	0.601** (0.300)	0.555* (0.307)
λ	0.744*** (0.052)	0.733*** (0.053)	0.738*** (0.053)	0.734*** (0.054)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Table 3.16: CSR results of spatial models with inverse distance matrix (with FDI) in eastern and central regions

Spatial matrix:	800KM inverse distance		800KM inverse distance	
Regions: east central	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CSR	(1)	(2)	(3)	(4)
did	1.076 (1.166)	-2.137 (3.364)	0.462 (1.188)	-1.678 (3.421)
WD	-0.897 (0.905)	N	0.256 (1.119)	N
WttD	N	0.108 (0.633)	N	0.415 (0.668)
WnttD	N	-0.831 (0.574)	N	-0.120 (0.727)
logPOP	N	N	-1.147 (1.714)	-1.072 (1.719)
logGDP	N	N	2.214 (1.537)	2.185 (1.539)
GDP2Ratio	N	N	0.116* (0.070)	0.118* (0.070)
logFDI	N	N	0.203 (0.842)	0.225 (0.843)
WlogPOP	N	N	-0.451 (7.001)	-0.395 (6.984)
WlogGDP	N	N	-0.423 (6.954)	0.151 (6.972)
WGDP2Ratio	N	N	0.659** (0.320)	0.617* (0.324)
WlogFDI	N	N	-4.926 (4.161)	-5.136 (4.154)
λ	0.744*** (0.053)	0.733*** (0.054)	0.731*** (0.054)	0.726*** (0.055)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Table 3.17: CER results of spatial models with inverse distance matrix (no FDI)

Spatial matrix:	800KM inverse distance		800KM inverse distance	
Regions: 237 cities	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CER	(1)	(2)	(3)	(4)
did	0.393 (0.284)	-0.381 (0.635)	0.387 (0.291)	-0.020 (0.651)
WD	-0.370** (0.170)	N	-0.028 (0.211)	N
WttD	N	-0.035 (0.102)	N	0.038 (0.107)
WnttD	N	-0.324*** (0.121)	N	-0.080 (0.156)
logPOP	N	N	0.521 (0.443)	0.534 (0.443)
logGDP	N	N	-0.062 (0.341)	-0.065 (0.341)
GDP2Ratio	N	N	0.058*** (0.017)	0.058*** (0.017)
WlogPOP	N	N	0.772 (1.052)	0.636 (1.066)
WlogGDP	N	N	-1.394 (0.975)	-1.252 (0.991)
WGDP2Ratio	N	N	0.079 (0.049)	0.073 (0.050)
λ	0.646*** (0.047)	0.632*** (0.048)	0.629*** (0.047)	0.623*** (0.048)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Table 3.18: CER results of spatial models with inverse distance matrix (with FDI)

Spatial matrix:	800KM inverse distance		800KM inverse distance	
Regions: 222 cities	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CER	(1)	(2)	(3)	(4)
did	0.372 (0.278)	-0.470 (0.693)	0.389 (0.285)	-0.037 (0.726)
WD	-0.336* (0.782)	N	0.034 (0.220)	N
WttD	N	-0.012 (0.115)	N	0.066 (0.119)
WnttD	N	-0.305** (0.126)	N	-0.043 (0.164)
logPOP	N	N	0.149 (0.467)	0.160 (0.467)
logGDP	N	N	-0.111 (0.403)	-0.115 (0.404)
GDP2Ratio	N	N	0.070*** (0.018)	0.070*** (0.018)
logFDI	N	N	0.193 (0.197)	0.199 (0.197)
WlogPOP	N	N	0.297 (1.449)	0.187 (1.456)
WlogGDP	N	N	-1.361 (1.331)	-1.127 (1.372)
WGDP2Ratio	N	N	0.094 (0.062)	0.088 (0.063)
WlogFDI	N	N	-0.241 (0.791)	-0.334 (0.800)
λ	0.673*** (0.049)	0.657*** (0.050)	0.655*** (0.050)	0.649** (0.051)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Table 3.19: CER results of spatial models with inverse distance matrix (no FDI) in easter and central regions

Spatial matrix:	800KM inverse distance		800KM inverse distance	
Regions: east central	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CER	(1)	(2)	(3)	(4)
did	0.275 (0.326)	-0.795 (0.936)	0.181 (0.330)	-0.311 (0.957)
WD	-0.239 (0.248)	N	0.267 (0.312)	N
WttD	N	0.061 (0.173)	N	0.183 (0.182)
WnttD	N	-0.247 (0.158)	N	0.086 (0.208)
logPOP	N	N	0.105 (0.474)	0.122 (0.475)
logGDP	N	N	0.142 (0.360)	0.140 (0.360)
GDP2Ratio	N	N	0.066*** (0.020)	0.066*** (0.020)
WlogPOP	N	N	1.349 (1.865)	1.372 (1.859)
WlogGDP	N	N	-1.859 (1.404)	-1.768 (1.410)
WGDP2Ratio	N	N	0.170* (0.087)	0.157* (0.089)
λ	0.738*** (0.055)	0.724*** (0.057)	0.737*** (0.055)	0.732*** (0.057)
Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.				

Table 3.20: CER results of spatial models with inverse distance matrix (with FDI) in eastern and central regions

Spatial matrix:	800KM inverse distance		800KM inverse distance	
Regions: east central	SDEM-DID	DSDEM-DID	SDEM-DID	DSDEM-DID
Dependent variable: CER	(1)	(2)	(3)	(4)
did	0.256 (0.330)	-0.702 (0.948)	0.167 (0.334)	-0.385 (0.965)
WD	-0.224 (0.254)	N	0.265 (0.321)	N
WttD	N	0.052 (0.179)	N	0.194 (0.189)
WnttD	N	-0.222 (0.159)	N	0.078 (0.208)
logPOP	N	N	0.052 (0.487)	0.071 (0.489)
logGDP	N	N	0.095 (0.436)	0.087 (0.436)
GDP2Ratio	N	N	0.066*** (0.020)	0.067*** (0.020)
logFDI	N	N	0.075 (0.2378)	0.080 (0.238)
WlogPOP	N	N	0.542 (1.987)	0.564 (1.981)
WlogGDP	N	N	-0.642 (1.975)	-0.480 (1.981)
WGDP2Ratio	N	N	0.184** (0.092)	0.172* (0.093)
WlogFDI	N	N	-1.264 (1.172)	-1.320 (1.170)
λ	0.737*** (0.056)	0.724*** (0.057)	0.731*** (0.057)	0.725*** (0.058)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Chapter 4

The Effect of Minimum Wage on Employment: A Spatial Analysis from a Panel of Chinese Cities

Abstract

A significant portion of the existing literature suggests that the effects of the minimum wage on employment are indeterminate due to the ambiguous theories surrounding employment and the minimum wage, endogeneity resulting from omitted variables, spatial heterogeneity, and spatial dependencies between variables. In China, varying levels of economic development across different regions of the country contribute to differential effects of the minimum wage on employment. This paper employs panel data from 2004 to 2011, encompassing 263 Chinese cities, to examine the direct and indirect minimum wage elasticity of the employment rate. The spatial Durbin model is utilized to control for endogeneity caused by the spatial dependence of unobserved factors and spatial heterogeneity. The findings, based on nationwide data, reveal that the direct and indirect minimum wage elasticities of the employment rate are significantly positive and negative, measuring at 0.164 and -0.201, respectively. When holding other variables constant, a 10% increase in employment in one city results in a 1.64% increase in the employment rate within that city, while simultaneously causing a 2.01% decrease in the employment rate of neighboring cities. In the relatively developed eastern and northeastern regions, raising the minimum wage shows insignificant positive direct effects and negative indirect effects on the employment rate. Conversely, in the less developed central and western regions, raising the minimum wage of a city directly increases its employment rate, but there are insignificant negative or positive indirect effects on the employment rate of that city's neighboring cities.

4.1 Introduction

In China, the effect of the minimum wage on employment has been a subject of extensive debate in various studies. Drawing on Chinese provincial panel data, several studies indicate that an increase in the minimum wage leads to a reduction in the employment of low-skilled workers working in private enterprises (Shi, 2010; G. Wang & Yao, 2014; W. Sun, Wang, & Zhang, 2015). However, contrasting perspectives suggest that the minimum wage can stimulate the employment of rural migrant workers and the manufacturing industry in China, particularly in conditions characterized by a non-discriminatory buyer monopsony labor market (Luo, 2007; Shi, 2011; Z. Han & An, 2007).

Moreover, the impact of the minimum wage on employment in China varies across regions. Specifically, in the more developed eastern region with higher GDP, the minimum wage exhibits a negative effect on employment, irrespective of enterprise ownership, industry type, or worker gender (Mou & Wang, 2015; Fang & Lin, 2015; Ni, Wang, & Yao, 2011). On the contrary, in the less developed central and western regions, an increase in the minimum wage is associated with improved employment outcomes, regardless of firm ownership, industry type, or workers' gender (Mou & Wang, 2015; Ni et al., 2011).

Figure 4.1: Annually average GDP and minimum wage

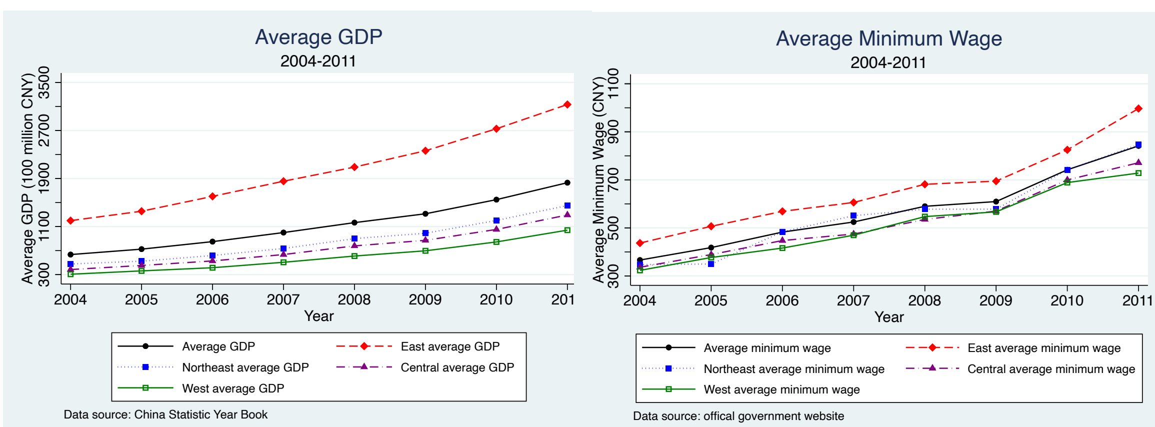


Figure 3.1 illustrates the annual average minimum wage and GDP in 263 cities from 2004 to 2011. The data clearly depicts a steady increase in the average minimum wage during this period. There is a notable plateau in the years 2008 to 2009, likely influenced by the global financial crisis. The eastern region, being the most developed in China, exhibits both the highest GDP and a superior minimum

wage standard compared to other regions. In contrast, the minimum wage in the less developed northeast region is nearly equivalent to the overall average minimum wage standard of all cities. Compared to the prosperous eastern and the moderately developed northeast regions, the central and western regions of China have the lowest minimum wage standards, corresponding to their status as the least developed regions. The annual average GDP of the northeast region is lower than the average of all cities but higher than the average GDP of the central and western regions; the western region displays the lowest annual average GDP. In general, economically developed provinces have established higher minimum wage standards than less developed ones. The provinces with the highest minimum wage levels are situated in the eastern part of China, aligning with their high GDP. Conversely, the provinces in the west of China exhibit lower minimum wage levels and lower GDP.

Since January 2004, a minimum wage regulation has governed all provincial cities in mainland China. This regulation mandates that minimum wage standards in each province should be adjusted at least every two years to align with local living standards. The minimum wage applies to both full-time and part-time employees across various types of organizations, including state-owned enterprises, private enterprises, private non-enterprise units, and employees in self-employed businesses. The minimum wage standards of different cities and provinces are frequently adjusted, and each province has considerable autonomy in formulating its own minimum wage standards (Fang & Lin, 2015). Furthermore, owing to the unbalanced economic development and variations in living standards among cities or provinces, the gap in minimum wage standards between different areas has gradually widened. This divergence offers rich data at the provincial and city levels for studying the impact of the minimum wage on employment.

Two challenges arise when studying the impact of the minimum wage on employment. The first challenge is related to empirical evidence, which may be affected by endogeneity due to the spatial heterogeneity between unobserved factors and both employment and the minimum wage (Totty, 2017). Consequently, studies investigating the impact of the minimum wage often utilize two-way fixed effects (Totty, 2017). However, it is acknowledged that two-way fixed effects may not entirely address concerns about unobserved heterogeneity (Allegretto, Dube, & Reich, 2011).

Allegretto et al. (2011) contend that spatial heterogeneity in employment trends introduces a bias in the effect of the minimum wage on employment, leading to an overstatement of precision. They observed that the significantly negative effect of the minimum wage on employment, evident when employing two-way fixed effects alone, became insignificantly positive after controlling for spatial heterogeneity.

Second, the competitiveness among different cities may contribute to spatial dependence in minimum wage levels. Different cities establish their minimum wage standards based on local employment patterns, demographics, education levels, industrial competition, and unobservable factors like local living conditions (Huang & Chand, 2015; Totty, 2017). However, neighboring cities have incentives to participate in minimum wage competition to attract a plentiful labor force (Li, Kanbur, & Lin, 2019). Furthermore, the effects of minimum wage increases in different cities are spatially interdependent; a change in the minimum wage in one city may influence its own employment and employment in neighboring cities (Majchrowska & Strawiński, 2021). Spatial dependence between different cities may also exist for employment (Kalenkoski & Lacombe, 2013; Pavlyuk et al., 2011). Therefore, investigating the spatial correlation between minimum wages and employment in different cities is crucial. Various factors, such as ambiguity in the theory of employment and the minimum wage, endogeneity due to omitted variables, spatial heterogeneity, and spatial dependencies between variables, could lead to biased estimates.

This paper employs city-level data and the Spatial Durbin Model (SDM) to study the effect of the minimum wage on the employment rate, utilizing city-level panel data covering 263 Chinese cities from 2004 to 2011. All variables are transformed into logarithmic form, allowing for the analysis of minimum wage elasticities of the employment rate. Given the variations in minimum wage standards and implementation dates across cities (Fang & Lin, 2015), city-level data is essential for analyzing the effect of the minimum wage on employment.

The primary contribution of this paper lies in the effective utilization of the Spatial Durbin Model (SDM) model, incorporating a spatial weighted matrix[1], and controlling for city and time fixed effects in the analysis. This approach proves to be effective in addressing heterogeneity in spatial dependence and mitigating endo-

[1] The spatial weighted matrix reflects the degree of spatial spillovers between neighboring cities.

geneity issues (Zangger, 2019). The SDM model not only tackles endogeneity arising from spatial dependence in omitted variables but also addresses spatial dependence inherent in exogenous independent variables (Fingleton & Le Gallo, 2010; Pace & LeSage, 2008; Zangger, 2019).

In contrast to existing research, which often draws conclusions regarding the impact of raising the minimum wage on local employment as either negative, positive, or insignificant (G. Wang & Yao, 2014; W. Sun et al., 2015; X.-Y. Xiao & Xiang, 2009; Z. Han & An, 2007), this paper, utilizing the SDM model with different spatial weighted matrices, contributes the finding that an increase in a city's minimum wage not only stimulates the local employment rate but also decreases the employment rate in neighboring cities.

The estimated results from the SDM model, utilizing nationwide data with 10 and 15 neighbors spatial weighted matrices[2], indicate that a 10% increase in the minimum wage results in a 1.64% and 1.97% increase in the employment rate within the city. However, it also leads to a 2.01% and 3.05% decrease in the employment rate in neighboring cities.

Since different regions in China exhibit varying levels of economic development, employment patterns, and demographic characteristics, this paper initially analyzes the data for all 263 cities. Subsequently, the study narrows down its focus to examine the impact of the minimum wage on the employment rate in four specific regions of China: (1) the economically developed eastern region, (2) the less developed northeastern region, (3) the underdeveloped central region, and (4) the least developed western region.

In the eastern and northeastern regions, characterized by relatively developed economies, an increase in the minimum wage shows insignificant positive direct effects and negative indirect effects on the employment rate. Conversely, in the central and western regions, which have relatively underdeveloped economies, an increase in the minimum wage leads to an increase in the employment rate in that city, with insignificant indirect effects on the employment rate in neighboring cities.

[2] The 10 and 15 neighbors' spatial weighted matrices include all 10 and 15 neighboring cities of each city, respectively.

For robustness checks, various inverse distance spatial weighted matrices[3] were employed. The results consistently indicate significantly positive direct effects and negative indirect effects of the minimum wage on the employment rate, reinforcing the conclusions drawn in this paper.

The remainder of the paper is organized as follows. In section two, the relevant literature is reviewed. In section three, the methodology is explained and the data base for this study is described. In section four, the results are presented and discussed. In section five, conclusions are presented.

4.2 Related Literature

A substantial body of research has delved into the impact of increasing the minimum wage on employment levels, yet there is considerable disagreement. The consequences of raising the minimum wage may manifest as positive, negative, or insignificant, with additional complexities arising from heterogeneous regional effects and spatial spillover effects on employment. These three dimensions are explored in the following analysis.

Effects of raising the minimum wage

Based on provincial panel data from 1996 to 2007, Shi (2010) found that short-term fluctuations in the minimum wage in China have no effect on the employment of low-skilled workers, but in the long run, these workers are negatively affected. Brown, Gilroy, and Kohen (1981) argued that increases in the minimum wage have a negative effect on the employment levels of the elderly and teenagers. G. Wang and Yao (2014) analyzed panel data from 30 provinces in China from 2000 to 2010 and found a negative effect of the minimum wage on the employment of low-skilled workers. W. Sun et al. (2015) analyzed provincial panel data from 2003 to 2009 and found that increases in the minimum wage have a negative effect on employment levels, but only for workers in private and individual enterprises.

In contrast, Luo (2007) used Shanghai time series data from 1993 to 2005 to argue that the minimum wage has a positive effect on low-wage employment. The

[3]Inverse distance matrices are based on the reciprocal of the distances between cities' central points to construct the spatial matrix.

author demonstrated that this conclusion is based on a non-discriminatory buyer monopsony labor market in China. X.-Y. Xiao and Xiang (2009) used panel data from six developed cities from 1995 to 2006 in China to study the effects of the minimum wage on wage levels and employment. They argued that the minimum wage compresses the wage distribution, which means that the minimum wage increases the wages of low-wage workers. And the minimum wage has a positive effect on employment in China, especially for low-wage workers. Shi (2011) showed that the minimum wage has different effects on employment levels in different occupations. Based on data from 27 provinces in China between 1996 and 2007, the result was that an increase in the previous year's minimum wage has a significant positive effect on construction employment, but an increase in the current year's minimum wage has a significant negative effect on manufacturing employment.

Z. Han and An (2007) studied the effect of the minimum wage on employment using data from a first-tier Chinese city (Shenzhen); the results showed that an increase in the minimum wage has an insignificant effect on the employment rate. Zavodny (2000) used state and individual level data to examine the effect of the minimum wage on teen employment in the U.S. The results showed that the minimum wage has an insignificant effect on teen employment. Kim, Lim, et al. (2018) employed panel data from 25 OECD countries, covering the period from 2000 to 2014, to study the effect of minimum wage increases on employment levels. They assumed that firms could hire both skilled and unskilled workers. The empirical results concluded that a 1 percent increase in the minimum wage reduced the rates of employment and unemployment by 0.7 percent and 0.64 percent, respectively. They argued that minimum wage increases have limited effects on employment.

Regional effects

The minimum wage also exhibits heterogeneous regional effects on employment. Mou and Wang (2015) utilized panel data from 27 provinces in China spanning 1997 to 2010 and employed the generalized method of moments (GMM) model to examine the impact of the minimum wage on employment, irrespective of the ownership of the firm, the type of industry, or the gender of the workers. They argued that the minimum wage has an insignificant effect on employment at the provincial level, a negative effect in central and eastern China, and a positive effect in western China.

J. Wang and Gunderson (2011) employed provincial data to investigate the effect of the minimum wage on the employment levels of rural migrant workers in China. They found that an increase in the minimum wage increased the employment level of rural migrant workers in state-owned enterprises in the prosperous eastern region. However, the effect was negative on the employment levels of rural migrant workers in central and western regions.

Fang and Lin (2015) demonstrated that in the eastern and central regions of China, an increase in the minimum wage has elevated the unemployment levels of young adults and low-skilled workers. In contrast, in the western region, the minimum wage has a positive (but insignificant) effect on the employment of young adults and low-skilled workers. They explained that the eastern, central, and western regions differ in terms of economic levels and industry competition, leading to varied effects of the minimum wage on employment levels. Ni et al. (2011) examined the effect of the minimum wage on employment in China using regional panel data from 2000 to 2005. They showed that the minimum wage has an insignificant effect on employment across China, regardless of the ownership of the firm, the type of industry, or the gender of workers. In the eastern region, the minimum wage has a significant negative effect on employment. In contrast, the minimum wage has a significant positive effect on employment in the central and western regions.

Vom Berge and Frings (2020) found that the minimum wage has heterogeneous regional effects on construction employment in Germany between 1997 and 2002. The minimum wage significantly reduced employment in eastern Germany but had no effect on employment in western Germany, likely because the minimum wage standard and economic levels in western Germany are relatively higher than in the eastern part. The effect of the minimum wage on employment also correlates with the economic level of different regions in a country.

Spatial dependence

The spatial dependence of the minimum wage effect on employment between different regions should be recognized. Biased and inconsistent estimates result from treating spatial dependence as a trivial factor in econometric analyses (Dolton, Bondibene, & Stops, 2015). Huang and Chand (2015) applied spatial econometric methodology to analyze the spatial interdependence of wage standards across different provinces in China. The results show the existence of significant provincial wage spatial spillover effects among Chinese provinces. Since the 2000s, regional wage inequality has begun to decline, possibly because provinces with high wage standards have positive spatial spillover effects on their neighboring provinces with lower wage standards. Li et al. (2019) conducted a spatial autoregressive (SAR) model to study city-level competition in minimum wage standards in China from 2004 to 2012. The minimum wage standards of different cities have significant positive spatial spillover effects; that is, when one city adjusts its minimum wage standard, it affects the adjustment of minimum wage standards in its neighboring cities.

Based on U.S. census data, Kalenkoski and Lacombe (2008) used a SAR model to demonstrate that a 10% increase in the minimum wage would result in a 3.2% decrease in youth employment, which is 28% higher than an OLS estimation. They also argued that adopting a spatial econometric model is necessary when analyzing the effect of the minimum wage on employment because omitted geographic factors can bias regression results. Dolton et al. (2015) concluded that in the UK, the minimum wage has a negative effect on employment for all workers and significant spatial spillovers to neighboring regions. They also emphasized that minimum wage adjustments and employment are not spatially independent across regions due to transport links. Studies on minimum wage or employment should consider testing for spatial dependence of data across regions.

Kalenkoski and Lacombe (2013) utilized U.S. state-level panel data from 1990 to 2004 and spatial econometric models to investigate how changes in the minimum wage affected teen employment. They employed SAR and SDM models for the analysis, revealing that a 10% increase in the real effective minimum wage leads to a 2.11% decrease in teen employment. This decrease encompasses both direct and indirect effects of the minimum wages. They also argued that teen employment is spatially dependent across states. Pavlyuk et al. (2011) studied how different economic factors affect employment rates in districts of Latvia using an SDM model and cross-sectional data for the years 2005, 2006, 2007, and 2008. The authors found a strong spatial dependence of employment rates between different districts.

Majchrowska and Strawiński (2021) utilized the SDM model to investigate spatial dependencies in the relationship between non-agricultural employment and the relative minimum wage[4]. They employed a panel dataset covering 380 Polish labor markets from 2006 to 2018. The study revealed significant spatial dependence between a location and its neighbors regarding employment and the relative minimum wage. For the years 2006 to 2012, the relative minimum wage had an insignificant effect on employment. However, for the years 2013 to 2018, the relative minimum wage exhibited significant negative direct and positive indirect effects on local employment. In other words, an increase in the relative minimum wage in a city would reduce local employment but increase employment in that city's neighboring cities. Thus, a city's employment would be affected not only by the increase in the relative minimum wage in that city but also by the increase in the relative minimum wage in neighboring cities. The relative minimum wage had different effects in different time periods because the minimum wage increased more during the period 2013 to 2018 than it did in the period 2006 to 2012.

4.3 Data

The panel dataset comprises city-level data for the developed eastern[5], the less developed northeastern[6], central[7], and the least developed western[8] regions of China from 2004 to 2011. The dataset encompasses a total of 263 cities.

The city-level panel data consist of demographic and macroeconomic variables. Demographic variables include employment, educational level, rural population, and the number of hospital beds. Macroeconomic variables comprise the minimum wage level, GDP, and agricultural GDP. All data were compiled from the China Statistical Yearbook, except for the minimum wage, which was extracted from the official government website of each city.

The dependent variable in this analysis is the logarithm of the employment rate, computed as the number of people employed in urban state-owned enterprises and urban private enterprises. This figure is subsequently divided by the total

[4]Relative minimum wage is the minimum wage relative to the average wage.

[5]The developed eastern region includes the provinces of Shanghai, Beijing, Tianjin, Shandong, Guangdong, Jiangsu, Hebei, Zhejiang, and Fujian.

[6]The less developed northeast region includes Liaoning, Jilin, and Heilongjiang provinces.

[7]The less developed central region includes Anhui, Shanxi, Jiangxi, Henan, Hubei, and Hunan.

[8]The least developed western region includes Yunnan, Sichuan, Ningxia, Guangxi, Gansu, Guizhou, Chongqing, and Shannxi.

population in each city. Both categories of enterprises fall under the purview of the 2004 minimum wage regulation.

The primary explanatory variable is the logarithm of the minimum wage. The monthly minimum wage is recorded at the city level, with different cities within a province having distinct minimum wage standards. Following the approach outlined in W. Sun et al. (2015), the highest monthly minimum wage across various cities within a province is selected for analysis.

The embodiment of the household registration system is the Hukou[9], which also has a significant effect on employment in China. Liu (2005) pointed out that the Hukou system poses a barrier to rural residents seeking urban employment. Even if individuals with a rural Hukou manage to acquire an urban Hukou later in life, they still face reduced chances of securing employment in the state or private sectors, where employers typically offer healthcare benefits. Therefore, the logarithmic rural population[10] ration can be used as the Hukou index. The calculation of the rural population ratio is the ratio of the rural population in a city to the total population of that city.

The education ratio is calculated as the population with a college degree or higher divided by the total population[11] of each city. This metric serves as an indicator to gauge the impact of education on employment. Yahong and Khan (2021) demonstrated that individuals with a college degree or above have an increased likelihood of being employed. Subsequently, the independent variable in this analysis is the logarithm of the education rate.

People with good health significantly enhance their employability in the labor market. Introducing a model that incorporates a city's medical resources can help mitigate potential biases stemming from factors influencing labor demand or supply (Yahong & Khan, 2021). Consequently, this paper employs the logarithm of hospital beds per capita in a city as an indicator of its medical resources, a measure utilized

[9] In China, Hukou is a household registration system that officially designates a person as a permanent resident of an urban or rural area. It includes identifying information such as name, parents, spouse, and date of birth.

[10] Rural population ratio is calculated by dividing the population with a rural Hukou but living in an urban area by the total population of each city.

[11] I use the total population to calculate the education ratio because I couldn't obtain specific population data of college age and older at the city level from the China Statistical Yearbook.

by Song et al. (2020) and Tian and Pan (2021). The calculation involves dividing the number of hospital beds in a city by its total population.

To further account for factors impacting labor demand, the macroeconomic variables log GDP per capita and log agricultural GDP per capita are included. Given that all variables are logarithmic, the interpretation of estimates revolves around the elasticity of the employment rate. Table 3.1 provides summary statistics for these variables, with a total of 2,104 observations.

Table 4.1: Summary statistics

Variable	Observation	Mean	Std. Dev.	Min	Max
Log employment rate	2104	-0.8399621	0.2646834	-1.444656	0.3812283
log minimum wage	2104	2.734319	0.1422476	2.230449	3.120574
Log rural population rate	2104	-0.2461352	0.3871328	-3.971705	-0.0349644
Log education rate	2104	-2.150298	0.4927485	-4.181715	-0.8960228
Log GDP per capita	2104	4.248253	0.3474784	3.33452	5.614727
Log agriculture GDP per capita	2104	3.336115	0.2099332	2.392669	4.02467
Log hospital beds per capita	2104	-2.220343	0.8320607	-3.064348	0.0649017

4.4 Empirical Strategy and Results

The model can be written as follows:

$$LEMPR_{it} = \alpha + LMW_{it}\beta + X_{it}\gamma + c_i + \zeta_t + \epsilon_{it} \qquad (1)$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

where $LEMPR_{it}$ represents the logarithmic employment rate in city i on day t. LMW_{it} denotes the logarithmic minimum wage in city i on day t, and X_{it} encompasses vectors of city demographic and macroeconomic variables. These variables include the logarithmic rural population rate, the logarithmic education rate, the logarithmic hospital beds per capita, the logarithmic GDP per capita, and the logarithmic agricultural GDP per capita. Additionally, c_i and ζ_t represent space- and time-specific effects, accounting for time- and space-invariant city characteristics, respectively. Finally, ϵ_{it} stands for the error term. The focal point of interest lies in the estimated coefficient β, which reflects the minimum wage rate elasticity of the employment rate.

Urban minimum wages and employment are likely to exhibit spatial correlation (Li et al., 2019; Majchrowska & Strawiński, 2021; Huang & Chand, 2015; Kalenkoski & Lacombe, 2013; Pavlyuk et al., 2011). Controlling for spatial interaction effects in the model can provide a better understanding of the spatial structure, thereby improving the explanatory power of the econometric model (Tosetti et al., 2018). Moreover, the use of spatial econometrics along with spatial weighted matrix regression in the analysis is an effective way to control for the spatial dependence of omitted variables that could cause endogeneity problems (Zangger, 2019). In spatial econometrics, three main spatial econometric models are commonly used: the spatial autoregressive model, the spatial error model, and the spatial Durbin model.

The spatial autoregressive model (SAR):

$$Y = \rho WY + \alpha l_n + X\beta + \epsilon \qquad (2)$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

Where Y denotes an $N*1$ vector as the dependent variable for every region (observation) in the sample $(i = 1, 2, ..., N)$. l_n is an $N*1$ vector of ones associated with constant term parameter α, X denotes an $N*K$ matrix as the independent variables for every region (observation) in the sample $(i = 1, 2, ..., N)$ that is associated with the parameter β, and ϵ is an $N*1$ vector of disturbance terms. W represents and $N*N$ spatial weighted matrix, WY is a spatially lagged dependent variable and denotes the endogenous interaction effects between the dependent variables. A parameter ρ measures the strength of spatial dependence between dependent variables of different regions; it is a $K*1$ vector.

The spatial error model (SEM):

$$Y = \alpha l_n + X\beta + \mu \qquad (3)$$

$$\mu = \lambda W\mu + \epsilon$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

Where $Y, \alpha l_n, X\beta$, and ϵ contain the same variables as the SAR model. $W\mu$ represents spatially autocorrelated error terms, signifying the interaction effects between the disturbance terms of different observations. The scalar parameter λ gauges the strength of dependence between cities; it is a $K * 1$ vector.

The spatial Durbin model (SDM):

$$Y = \rho WY + \alpha l_n + X\beta + WX\theta + \epsilon \qquad (4)$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

Where WX is a spatially lagged independent variable, representing the exogenous interaction effects between independent variables. The parameters ρ and θ quantify the strength of spatial dependence between regions and are $K * 1$ vectors. All other variables' contents are the same as in the SAR model.

If no spatial dependence exists between variables, the equation can be written as:

$$Y = \alpha l_n + X\beta + \epsilon \qquad (5)$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

This equation represents a standard linear regression model (OLS).

The binary spatial contiguity and spatial inverse distance matrices[12] are commonly employed in spatial econometric model regressions. W represents a first-order spatial weighted matrix and is time-invariant. In this matrix, the rows correspond to observation i, and the columns correspond to observation j, reflecting the impact of neighboring observation j on observation i (J. LeSage & Pace, 2009). Consider four cities labeled C1, C2, C3, and C4. C1 is a neighbor of C2 and C3. C2 is a neighbor of C1 and C3. C3 is a neighbor of C1, C2, and C4. C4 is a neighbor of C3.

First, we generate a first-order binary contiguity matrix C:

[12] Similar to the spatial weight matrices used in the analyses in chapter 1.

$$C = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

In matrix C, elements equal to 1 indicate that two regions share a common border, while 0 indicates that they do not. For instance, a value of 1 in row 1 and columns 2 and 3 signifies that C1 shares a border with both C2 and C3. All other elements in row 1 have a value of zero. Elements in rows 2, 3, and 4 carry the same implications as row 1, signifying the shared borders between the corresponding cities. Notably, diagonal elements in the matrix are set to zero, signifying that regions are not considered neighbors of themselves.

The matrix is typically normalized to a spatially weighted matrix with row sums of 1 to create a spatial lag or linear combination of values from neighboring observations. This spatial weighted matrix is:

$$W = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

WY and WX can be understood as the weighted average of the surrounding dependent variables and the weighted average of the surrounding independent variables. $W\mu$ is a weighted average of the surrounding error terms.

The spatial inverse distance matrix is constructed based on the distance between regions. The elements of the spatial inverse distance matrix are:

$$W_{ij} = \tfrac{1}{d_{ij}},\ i \neq j;\ 0, i = j$$

The construction of a spatial inverse distance matrix is the same as that of the binary contiguity spatial weighted matrix. The diagonal elements of the spatial inverse distance matrix are also zero. The off-diagonal elements represent the inverse distance between cities, where d_{ij} is the distance between city i and city j. A spatial inverse matrix ensures that cities are considered neighbors based on their distance, capturing spatial spillover effects from cities without shared borders to other cities.

Many other spatial weighted matrices exist in spatial econometric research, such as the k-nearest adjacent matrix and the threshold inverse distance matrix. Additionally, some matrices are constructed based on economic standards or the coordinates of observations.

Spatial econometric models encompass three spatial interaction effects and two effects (J. LeSage & Pace, 2009). The three spatial interaction effects are endogenous interaction effects, exogenous effects, and interaction effects between error terms. Endogenous and exogenous interaction effects gauge whether the dependent and explanatory variables, respectively, and error terms of neighboring cities are spatially dependent. Two effects are direct and indirect effect. The direct effect signifies the impact of a change in an explanatory variable on the dependent variable in the same unit. In contrast, the indirect effect measures the spatial spillover effect of variables in a city to its neighboring cities, which is typically constrained to zero in standard econometric models. Indirect effects are further categorized into global and local indirect effects. Global indirect effects represent the impact of a change in a city's explanatory variable that spills over to all other areas, including those not directly connected by a spatial weighted matrix. Additionally, J. LeSage and Pace (2009) argue that the global indirect effect is transmitted to the area of origin of the effect. On the other hand, local indirect effects specifically measure the spatial spillover effects among areas connected according to a weighted spatial matrix.

The SDM model is utilized to introduce direct and indirect effects due to its inclusion of both endogenous and exogenous interaction effects. The reduced form of the SDM model (4) can be expressed through the following calculation:

$$Y = (I - \rho W)^{-1}(\alpha l_n + X\beta + WX\theta + \epsilon)$$

Taking the partial derivatives of $E(Y)$ of the above equation with respect to the K_{th} explanatory variables of X in the city (observation) 1 up to N, we obtain:

$$\begin{bmatrix} \frac{\partial E(Y)}{\partial X_{1k}} & ... & \frac{\partial E(Y)}{\partial X_{nk}} \end{bmatrix} = (I-\rho W)^{-1} \begin{bmatrix} \beta_k + W_{11}\theta_k & W_{12}\theta_k & \cdots & W_{1n}\theta_k \\ W_{21}\theta_k & \beta_k + W_{22}\theta_k & \cdots & W_{2n}\theta_k \\ \vdots & \vdots & \cdots & \vdots \\ W_{n1}\theta_k & W_{n2}\theta_k & \cdots & \beta_k + W_{nn}\theta_k \end{bmatrix}$$

Since the diagonal elements of the spatial weighted matrix W are zero, the $W_{11}, W_{22}, ..., W_{nn}$ are all equal to 0. The above partial derivatives matrix can be re-written as:

$$\begin{bmatrix} \frac{\partial E(Y)}{\partial X_{1k}} & ... & \frac{\partial E(Y)}{\partial X_{nk}} \end{bmatrix} = (I-\rho W)^{-1} \begin{bmatrix} \beta_k & W_{12}\theta_k & \cdots & W_{1n}\theta_k \\ W_{21}\theta_k & \beta_k & \cdots & W_{2n}\theta_k \\ \vdots & \vdots & \cdots & \vdots \\ W_{n1}\theta_k & W_{n2}\theta_k & \cdots & \beta_k \end{bmatrix} = (I-\rho W)^{-1}(\beta_k I_n + \theta_k W)$$

where W is the spatial weighted matrix, and I_n is an identity matrix. The expression of the spatial multiplier matrix is an infinite series as follows:

$$(I-\rho W)^{-1} = I + \rho W + \rho^2 W^2 + \rho^3 W^3 + \cdots$$

W represents the first-order contiguity neighbors. The matrix W^2 shows the second-order contiguity neighbors, which are the neighbors to the first-order neighbors. For example, if C1 has only one contiguity neighbor, C2. C2 has two contiguity neighbors C1 and C3. Then the first-order neighbor of C1 is C2, and the second-order neighbors of C1 are C1 and C3. The matrix W^3 shows the third-order contiguity neighbors, and so on.

Since the diagonal elements of the first matrix term I on the right-hand side of the above equation are 1, the non-diagonal elements are zero, and this term represents a direct effect of a change in X. However, the diagonal elements of the second matrix term ρW on the right-hand side of the above equation are assumed to be zero, and this term represents an indirect effect of a change in the explanatory variables. All other terms on the right-hand side of the above equation represent higher-order indirect effects.

J. LeSage and Pace (2009) defined the direct effect and the indirect effect measured by the diagonal elements and the non-diagonal elements of the matrix $(I - \rho W)^{-1}(\beta_k I_n + \theta_k W)$. They also suggest calculating the average sums of diagonal and non-diagonal elements of the matrix $(I - \rho W)^{-1}(\beta_k I_n + \theta_k W)$ to obtain the average direct and indirect effects, respectively.

Based on the analysis of direct and indirect effects above, the direct and indirect effects of the other models can be derived. SEM and OLS models only have a direct effect, which can be obtained through β_k. The SAR model includes endogenous interaction effects WY but no exogenous interaction effects WX. The direct effect and indirect effect can be obtained directly by $(I - \rho W)^{-1}\beta_k I_n$. The SDM model includes both endogenous and exogenous interaction effects, and the direct effect and indirect effect can be obtained directly by $(I - \rho W)^{-1}(\beta_k I_n + \theta_k W)$.

Although the SEM model accounts for spatial dependence in the disturbance process, it does not provide an indirect effect because $E(\epsilon) = 0$. Based on the definition of the local and global indirect effects, the SAR model estimates global indirect effects, and the SDM model estimates both global and local indirect effects.

The SAR and SEM models are trendy in spatial econometric analysis. However, SAR model include a consistent ratio between direct and indirect effects for all variables and the inability to discern whether a significant indirect effect is due to omitted spatially correlated explanatory variables or spatially correlated dependent variables (Halleck Vega & Elhorst, 2015), the indirect effects of the SAR model are only global. SEM model is it lacks the ability to account for spatial spillover effects, because it solely controls for spatial dependence in error terms (Halleck Vega & Elhorst, 2015).

J. LeSage and Pace (2009) recommended the widespread use of the SDM model in spatial analysis because it controls for both exogenous and endogenous spatial effects and has no prior restrictions on the ratio between direct and indirect effects. Moreover, the SDM model can control the endogeneity caused by the spatial dependence inherent in omitted variables and exogenous independent variables (J. P. LeSage & Pace, 2008). The process of how the SDM model controls this endogeneity is shown as follows:

$$Y = X\beta + O \qquad (6)$$

$$O = \rho WO + X\gamma + \epsilon \qquad (7)$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

Assume that Y is influenced by an exogenous variable X and an omitted variable O, and that O is not only spatially dependent but also correlated with X. Rearranging equation (7) yields the following:

$$O = (I - \rho W)^{-1}(X\gamma + \epsilon) \qquad (8)$$

ubstituting equation (8) into (6) and rearranging yields the SDM model (9):

$$Y = \rho WY + X(\beta + \gamma) - \rho WX\beta + \epsilon \qquad (9)$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

Spatial econometrics has evolved from estimating cross-sectional data to panel data. J. P. Elhorst (2017) concluded that spatial panel econometric models have two advantages. First, they can control for space-specific effects. Because applying spatial models does not imply that all unobserved space-specific and time-invariant variables can be controlled for. For instance, cities near the ocean will have different economic and political backgrounds than inland cities. Second, spatial panel econometrics can control for time-specific effects caused by space-invariant variables, such as the adjustment of government policies that can significantly affect the functioning of the economy.

The extension of the spatial cross-sectional model to the spatial panel model for N observations over T periods is obtained by adding a subscript t, and by introducing space-specific effects (r_i) and time-specific effects (ζ_t) that control for unobserved time-invariant and space-invariant variables, as follows:

The panel SAR model is:

$$Y_{it} = \rho WY_{it} + \alpha l_n + X_{it}\beta + r_i + \zeta_t + \epsilon_{it}$$

$$\epsilon_{it} \sim N(0, \sigma^2 I_n)$$

The panel SEM model is:

$$Y_{it} = \alpha l_n + X_{it}\beta + r_i + \zeta_t + \mu_{it}$$

$$\mu_{it} = \lambda W \mu_{it} + \epsilon$$

$$\epsilon_{it} \sim N(0, \sigma^2 I_n)$$

The panel SDM model is:

$$Y_{it} = \rho W Y_{it} + \alpha l_n + X_{it}\beta + W X_{it}\theta + r_i + \zeta_t + \epsilon_{it}$$

$$\epsilon_{it} \sim N(0, \sigma^2 I_n)$$

where i is the city of observation ($i = 1, 2, 3, ..., N$) and t is the time period ($t = 1, 2, 3, ..., T$).

The spatial panel models used in this research can be written as:

The panel SAR model:

$$LEMPR_{it} = \alpha + \rho W LEMPR_{it} + LMW_{it}\beta + X_{it}\gamma + c_i + \zeta_t + \epsilon_{it}$$

$$\epsilon_{it} \sim N(0, \sigma^2 I_n)$$

The panel SEM model:

$$LEMPR_{it} = \alpha + LMW_{it}\beta + X_{it}\gamma + c_i + \zeta_t + \mu_{it}$$

$$\mu_{it} = \lambda W \mu_{it} + \epsilon$$

$$\epsilon_{it} \sim N(0, \sigma^2 I_n)$$

The panel SDM model:

$$LEMPR_{it} = \alpha + \rho W LEMPR_{it} + LMW_{it}\beta + X_{it}\gamma + W LMW_{it}\theta_1 + W X_{it}\theta_2 + c_i + \zeta_t + \epsilon_{it}$$

$$\epsilon_{it} \sim N(0, \sigma^2 I_n)$$

W is the spatial weighted matrix. θ_1 and θ_2 represent the effects of the neighborhood minimum wage and other control variables on the employment rate. The primary coefficients of interest are β and θ_1. The space- and time-specific factors are treated as fixed effects. Maximum Likelihood (ML) estimators are extended to include fixed effects. The response parameters of the fixed effects model can be estimated by first concentrating out the fixed effects, a process known as demeaning. The resulting equation can then be estimated using ML estimation (J. P. Elhorst, 2014).

Before conducting the spatial econometrics analysis, the global and local Moran's I tests[13] (Anselin, 1995) are employed to test the spatial autocorrelation of the log employment rate and log minimum wage.

Secondly, classical and robust panel Lagrange multiplier (LM) tests[14] can be employed to test whether spatial econometric models are more appropriate for describing the data than a model without spatial interaction effects.

Thirdly, Lagrange multiplier tests (Burridge, 1981) are employed to determine whether the SDM model can be simplified to SAR or SEM models.

Fourthly, a Hausman test for spatial models (J. P. Elhorst, 2017) can be applied to determine whether spatial and/or temporal factors can be treated as fixed or random effects.

Different spatial weight matrices are employed for the analysis. These matrices encompass binary contiguity, 10, and 15 neighbors' spatial weighted matrices, along with inverse distance matrices, including 400KM, 500KM, and an entire country inverse distance matrix. The binary contiguity spatial weighted matrix is formed by considering cities' neighbors that share borders with a specific city. The 10 and 15 neighbors' spatial weighted matrices include all 10 and 15 neighboring cities of each city, respectively. Those spatial weight matrices are employed to compare various estimates.

The 400KM, 500KM, and all-range inverse distance matrices are employed for robustness checks. These matrices are constructed based on the reciprocal of the

[13]All results of panel global and local Morans' I tests are generated in Stata using Maurizio Pisati's tools for spatial analysis.

[14]Anselin, Gallo, and Jayet (2008) specified the classic LM test for panel data, and J. Elhorst (2010) developed the robust LM test for panel data.

distances between cities' central points to create the spatial matrix. In the 400KM and 500KM inverse distance matrices, a city's neighbors are limited to not exceed these distance thresholds. In the all-range inverse distance matrix, every city is considered to have neighbors because every city has a distance from another city. The inverse distance spatial weight matrices are utilized for robustness checks.

4.5 Results

3.5.1 Empirical Results

Table 4.2: Global Moran's I test

Minimum wage global spatial autocorrelation test						
matrix	Binary contiguity	10 neighbours	15 neighbours	400KM inverse distance	500KM inverse distance	Entire country inverse distance
Variable	Moran's I	Moran's I	Moran's I	Moran's I	Moran's I	Moran's I
Log	0.585***	0.561***	0.542***	0.528***	0.501***	0.204***
minimum wage	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log	0.029***	0.031***	0.025***	0.019***	0.019***	0.005**
employment rate	(0.045)	(0.000)	(0.000)	(0.006)	(0.002)	(0.038)
Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.						

Table 3.2, Figure 3.2, and 3.3 (Figure 3.2 and 3.3 is in the Appendix) present the results of the global and local Moran's I tests using different spatial weighted matrices for log minimum wage and log employment rate. Both variables exhibit significant global and local spatial autocorrelations. However, the log minimum wage demonstrates a stronger global and local spatial autocorrelation compared to the log employment rate.

Table 4.3: The (robust) Lagrangian multiplier (LM) test results between spatial models and the OLS model

Dependent variable: log employment rate						
Independent variable:	Coefficient	t-value	p-value	95% conf. interval		
log minimum wage	-0.187***	-6.130	0.000	-0.2463088 -0.1269003		
Log rural population rate	-0.129***	-14.330	0.000	-0.1463544 -0.1111267		
Log education rate	0.100***	11.920	0.000	0.0816508 0.1138091		
Log GDP per capita	0.509***	33.360	0.000	0.4795003 0.5394047		
Log agriculture GDP per capita	-0.136***	-8.180	0.000	-0.1690033 -0.1036314		
Log hospital beds per capita	0.015***	3.180	0.002	0.0055942 0.0236673		
Diagnostics checking, null hypothesis: no spatial autocorrelation						
Spatial weights matrixes	Binary contiguity	10 neighbors	15 neighbors	400KM inverse distance	500KM inverse distance	Entire country inverse distance
Spatial error:						
Lagrange multiplier	0.041 (0.840)	1.151 (0.283)	1.016 (0.313)	2.658 (0.103)	1.998 (0.158)	1.471 (0.225)
Robust Lagrange multiplier	1.385 (0.239)	0.063 (0.802)	0.232 (0.630)	0.303 (0.582)	1.004 (0.316)	3.040* (0.081)
Spatial lag:						
Lagrange multiplier	1.142 (0.285)	4.047** (0.044)	5.063** (0.024)	10.428*** (0.001)	12.251*** (0.000)	13.392*** (0.000)
Robust Lagrange multiplier	2.487 (0.115)	2.958* (0.085)	4.279** (0.039)	8.073*** (0.004)	11.258*** (0.001)	14.962*** (0.000)
Notes: ***indicates significance at the 1% confidence level, **at 5% and *at 10%.						

Table 3.3 displays the results of the (robust) Lagrangian multiplier (LM) tests for the diagnostic check between the SAR model, the SEM model, and the OLS model without controlling for spatial dependencies of the variables. These tests are conducted using different spatial weighted matrices. When utilizing the contiguity spatial weighted matrix, the classical and robust LM test results indicate that general OLS is more suitable for the analysis. However, the test results consistently support the SAR model when employing other spatial weighted matrices, suggesting that the SAR model provides a better fit for the data compared to SEM or general OLS models.

Table 4.4: Lagrange multiplier (LM) test for comparing the SAR, SEM, and SDM models

Spatial weights matrixes	Binary contiguity	10 neighbors	15 neighbors	400KM inverse distance	500KM inverse distance	Entire country inverse distance
Spatial error model						
Null hypothesis: SDM model can be simplified to the SEM model, $H_0 : \theta + \lambda\beta = 0$						
Lagrange multiplier	37.22***	20.27***	29.92***	26.33***	25.11***	27.23***
	(0.000)	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)
Spatial autoregressive model						
Nnull hypothesis: SDM model can be simplified to the SAR model, $H_0 : \theta = 0$						
Lagrange multiplier	32.99***	17.88***	26.79***	24.63***	24.76***	24.33***
	(0.000)	(0.007)	(0.000)	(0.000)	(0.000)	(0.001)
Notes: ***indicates significance at the 1% confidence level, **at 5% and *at 10%.						

It is important to note that (robust) LM tests may not account for exogenous interaction effects, such as spatial dependence between independent variables or error terms (J. P. Elhorst, 2017). This may explain why the OLS model performed better when the spatially weighted matrix of contiguity was used for the test. Therefore, this paper also conducts a Lagrange multiplier (LM) test to compare the SAR, SEM, and SDM models while controlling for space and time fixed effects. The results are presented in Table 3.4. Across all results under different space-weighted matrices, it is evident that the SDM model cannot be simplified to SEM or SAR models.

Table 4.5: Hausam test

ull hypothesis: the preferred model is random effects						
Spatial weights matrices	Binary contiguity	10 neighbors	15 neighbors	400KM inverse distance	500KM inverse distance	Entire country inverse distance
SDM model Hausman test	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic
Chi-square	-178.40	-177.37	-86.47	-257.21	-254.40	-3478.99
Notes: Chi-square value less than zero indicates rejecting the null hypothesis.						

The results of the Hausam tests with different spatial weighted matrices in Table 3.5 show that the random effects model must be rejected. Therefore, the fixed effects SDM model is more appropriate to study the effects of the minimum wage on employment.

Table 4.6: Entire data analysis, controlling city- and time- fixed effects

Dependent variable: log employment rate				
Spatial weighted matrix:	Pooled OLS	SDM Binary contiguity	SDM 10 neighbors	SDM 15 neighbors
Log minimum wage	0.047	0.078	0.165***	0.198***
	(0.346)	(0.187)	(0.005)	(0.004)
Log rural population rate	-0.009	-0.002	-0.004	-0.003
	(0.557)	(0.871)	(0.782)	(0.835)
Log education rate	0.001	-0.006	-0.007	-0.007
	(0.971)	(0.708)	(0.684)	(0.661)
Log GDP per capita	0.064**	0.059*	0.061**	0.064**
	(0.026)	(0.056)	(0.037)	(0.023)
Log agriculture GDP per capita	-0.218***	-0.051	-0.127*	-0.108
	(0.001)	(0.470)	(0.085)	(0.126)
Log hospital beds per capita	0.011**	0.013	0.011	0.0106
	(0.022)	(0.117)	(0.157)	(0.148)
W*Log minimum wage rate	N.A.	-0.024	-0.193**	-0.277***
		(0.786)	(0.021)	(0.009)
W*Log rural population rate	N.A.	-0.024	-0.022	-0.042
		(0.194)	(0.620)	(0.442)
W*Log education rate	N.A.	0.074**	0.099	0.100
		(0.027)	(0.119)	(0.203)
W*Log GDP per capita	N.A.	0.009	-0.002	-0.029
		(0.853)	(0.968)	(0.751)
W*Log agriculture GDP per capita	N.A.	-0.315***	-0.121	-0.182
		(0.001)	(0.245)	(0.106)
W*Log hospital beds per capita	N.A.	-0.005	-0.004	-0.005
		(0.622)	(0.692)	(0.588)
ρ	N.A.	0.172***	0.234***	0.259***
		(0.000)	(0.000)	(0.000)
log minimum wage rate				
Direct effect	N.A.	0.080	0.164***	0.197***
		(0.177)	(0.006)	(0.005)
Indirect effect	N.A.	-0.014	-0.201**	-0.305**
		(0.885)	(0.041)	(0.016)
Total effect	N.A.	0.066	-0.037	-0.109
		(0.449)	(0.688)	(0.301)
Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.				

Table 3.6 presents the results of the regressions using the entire country data for the analysis, including the pooled OLS and the SDM model with different spatial weighted matrices (while controlling for city and time fixed effects). Without considering the spatial dependence of the variables, the minimum wage elasticity of the employment rate appears to be positive but insignificant in the pooled OLS model. When employing the SDM model with a binary contiguity spatial weighted matrix to analyze the data, the log minimum wage also shows an insignificant direct and indirect effects on the log employment rate.

The minimum wage exhibits significant direct and indirect effects on the employment rate when utilizing alternative spatial weighted matrices. Taking the results of the SDM model with a 10-neighbors contiguity spatial weighted matrix as an example, the parameter estimates for log minimum wage and W*log minimum wage are significant at the 10% and 5% levels, respectively. However, interpreting these coefficients directly can be challenging in spatial econometrics (J. P. Elhorst, 2014). Therefore, it is essential to examine the direct and indirect effects of the log minimum wage. The differences between the parameter estimates and their direct effects stem from feedback effects[15] (J. P. Elhorst, 2014).

Estimated results of SDM model with the 10-neighbors contiguity spatial weighted matrix, the minimum wage demonstrates a significantly positive direct effect and a significantly negative indirect effect on the employment rate at the 1% and 5% levels, respectively. In practical terms, this implies that a 10% increase in the minimum wage of a city corresponds to a 1.64% increase in the employment rate of that city but a 2.01% decrease in the employment rate of its neighboring cities. While the direct effect is positive, the indirect effect is negative, resulting in a total minimum wage elasticity of the employment rate that is negative (though insignificant). Notably, using a 15-neighbors contiguity spatial weighted matrix in the SDM model leads to higher magnitudes for both direct and indirect effects of the log minimum wage on the log employment rate compared to the results obtained with a 10-neighbors contiguity spatial weighted matrix.

The findings presented in Tables 3.6 underscore the potential bias or inaccuracy

[15]Feedback effects involve the impact of a city that travels through neighboring cities and returns to the original city.

in pooled OLS regression estimates when the spatial dependencies of variables are neglected. Spatial econometrics becomes essential for adequately controlling these spatial dependencies. Additionally, the choice of different spatial weighted matrices introduces variability in the estimated results of the SDM model due to the diverse feedback effects associated with a city's various neighbors.

Notably, the indirect effect of the minimum wage becomes more prominent when a city has a greater number of neighboring cities. This suggests that when a city increases its minimum wage, the impact is not limited to workers migrating from immediately adjacent cities but may extend to more distant neighboring cities. This observation might explain why the indirect elasticity of the employment rate concerning the minimum wage is insignificant when a binary contiguity matrix is employed for the analysis.

Tables 3.8 to 3.11, presented in the Appendix, elucidate the impact of the minimum wage on the employment rate in specific regions of China. Table 3.8 focuses on the eastern region, characterized by a relatively developed economy. In this context, the minimum wage exhibits both insignificant direct and indirect effects on the employment rate.

In Table 3.9, estimations are provided for China's northeastern region, which features a less developed economy compared to the eastern region. The direct minimum wage elasticity of the employment rate is 0.44, reaching significance at the 5% level. However, the indirect minimum wage elasticity of the employment rate is negative and statistically insignificant.

Tables 3.10 presents the estimated results for China's underdeveloped central region. After incorporating city and time fixed effects, the pooled OLS estimation suggests a minimum wage elasticity of the employment rate at 0.059, reaching significance at the 10% level. This indicates inelasticity and low significance, implying that the minimum wage exerts almost no effect on the employment rate. It is noteworthy that OLS estimates might be biased as they do not account for the spatial dependencies of variables. When employing the SDM model with a binary contiguity spatial weighted matrix to analyze the data for the central region, the minimum wage exhibits both insignificant direct and indirect effects on the employment rate. However, utilizing other spatial weighted matrices for analysis reveals varied outcomes. The direct minimum wage elasticity of the employment rate is positive and significant at a relatively low level, while the indirect minimum wage elasticity of employment is negative but insignificant.

The results presented in Tables 3.11 focus on the impact of the minimum wage on the employment rate in the least developed western region of China. After accounting for city and time fixed effects, the pooled OLS analysis indicates that the minimum wage has an insignificant effect on the employment rate. However, employing the SDM model with binary contiguity, 10 neighbors, and 15 neighbors spatial weighted matrices reveals varying outcomes, with the direct minimum wage elasticity of the employment rate being positive and significant at different levels.

Based on the comprehensive analysis using the entire country's data and regressing the SDM model with 10 and 15 neighbors spatial weighted matrices, the findings consistently show significantly positive direct minimum wage elasticity and significantly negative indirect minimum wage elasticity for the employment rate. In simpler terms, as the minimum wage increase in a city increases, its own employment rate tends to increase, but this positive effect is accompanied by a negative impact on the employment rates of neighboring cities.

In the developed eastern region of China, the minimum wage exhibits insignificant direct or indirect effects on the employment rate. Conversely, in the relatively developed northeastern region, the direct minimum wage elasticity of the employment rate is positive and significant at the 5% level, while the indirect minimum wage elasticity is negative but statistically insignificant when analyzed using the SDM model with the binary contiguity matrix. The use of other spatially weighted matrices results in insignificant impacts of the minimum wage on the employment rate. For both the underdeveloped central and least developed western regions, the direct minimum wage elasticity of the employment rate is significantly positive, though the significance varies with changes in spatial weighted matrices. The indirect minimum wage elasticity remains insignificantly negative.

The minimum wage has a nationwide impact, as labor migration is not restricted within regions but can occur across the entire country. Therefore, in the analysis using national data, the minimum wage exhibits significant indirect effects on the employment rate.

3.5.2 Robustness check

Table 4.7: Robustness check using entire data

Dependent variable: log employment rate			
Spatial weighted matrix:	400KM inverse distance	500KM inverse distance	Entire country inverse distance
Log minimum wage	0.166***	0.161***	0.165***
	(0.008)	(0.008)	(0.004)
Log rural population rate	-0.005	-0.005	-0.006
	(0.711)	(0.702)	(0.649)
Log education rate	-0.008	-0.007	-0.005
	(0.641)	(0.667)	(0.755)
Log GDP per capita	0.058**	0.057*	0.057*
	(0.047)	(0.056)	(0.056)
Log agriculture GDP per capita	-0.091	-0.096	-0.103
	(0.201)	(0.174)	(0.14)
Log hospital beds per capita	0.017**	0.018**	0.013*
	(0.025)	(0.019)	(0.082)
W*Log minimum wage rate	-0.224**	-0.243**	-0.728**
	(0.031)	(0.027)	(0.025)
W*Log rural population rate	-0.023	-0.015	-0.084
	(0.690)	(0.833)	(0.616)
W*Log education rate	0.083	0.097	0.060
	(0.162)	(0.200)	(0.833)
W*Log GDP per capita	0.006	-0.006	0.119
	(0.954)	(0.960)	(0.742)
W*Log agriculture GDP per capita	-0.214*	-0.198	-0.612*
	(0.063)	(0.121)	(0.093)
W*Log hospital beds per capita	-0.014	-0.016	-0.018
	(0.170)	(0.123)	(0.496)
ρ	0.324***	0.374***	0.624***
	(0.000)	(0.000)	(0.000)
log minimum wage rate			
Direct effect	0.164**	0.160***	0.161***
	(0.010)	(0.009)	(0.005)
Indirect effect	-0.253*	-0.294*	-1.797
	(0.063)	(0.063)	(0.106)
Total effect	-0.089	-0.134	-1.636
	(0.476)	(0.367)	(0.136)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Different inverse distance spatial weight matrices, including 400KM, 500KM, and for the entire country, are employed for robustness checks. The results in Table 3.7 display the SDM model estimations with different inverse distance matrices using the entire country's data. When using the 400KM and 500KM inverse distance spatial weighted matrices for the analysis, the direct and indirect effects of the minimum wage on the employment rate are significantly positive and negative, respectively. This is consistent with the analysis using the 10 and 15 neighbors spatial weighted matrices.

However, when using the entire country inverse distance spatial weighted matrix for the analysis, the direct minimum wage elasticity of the employment rate remains significantly positive, but the indirect minimum wage elasticity of the employment rate becomes insignificant.

The magnitude of spatial spillover effects of minimum wage decreases as a city has more neighbors or as the distance between a city and its neighbors increases. This approach aligns with Tobler's first law, as articulated by Sui (2004), stating that "everything is related to everything else, but nearby things are more related than distant things" (p. 236). This implies that workers are more likely to move from their current cities to neighboring cities, rather than moving extensively across the entire country.

4.6 Conclusion

In this paper, a panel dataset encompassing city-level data from China's eastern, northeastern, central, and western regions spanning the years 2004 to 2011 was utilized to investigate the influence of the minimum wage on the employment rate. Notably, all variables were log-transformed to facilitate the computation of minimum wage elasticity concerning the employment rate.

Methodologically, this paper employs a spatial Durbin model (SDM) to mitigate endogeneity resulting from the omission of variables with spatial dependence or spatial heterogeneity. Additionally, different spatial weighted matrices are utilized to capture the extent of spatial spillovers between neighboring cities. City- and time-fixed effects are controlled for addressing unobserved variables in cities or over time.

The paper yields three key results. Firstly, the global and local spatial Moran's I tests reveal spatial autocorrelation in both log minimum wage and log employment rate, substantiating the rationale for adopting spatial models. The robustness of the results is enhanced compared to conventional models, particularly considering the spatial autocorrelation of log minimum wage and log employment rate, a consideration addressed by the SDM model.

Secondly, employing the entire dataset to estimate the SDM model with 10 and 15 neighbors' spatial weighted matrices while controlling for city and time fixed effects, both the direct and indirect minimum wage elasticity of the employment rate are found to be significantly positive and negative, respectively. In essence, when the minimum wage in a city increases, the employment rate in that city rises, while the employment rate in neighboring cities decreases. Taking the results from the SDM model with the 10 neighbors' spatial weighted matrix as an illustration, the direct and indirect minimum wage elasticity of the employment rate are positive and negative values, 0.164 (significant at the 1% level) and -0.201 (significant at the 5% level), respectively. This implies that a 10% increase in the minimum wage in a city would lead to a 1.64% increase in its employment rate but a 2.01% decrease in the employment rate of its neighboring cities, holding the other variables constant.

Thirdly, in the economically more developed eastern and northeastern regions, an increase in the minimum wage has an insignificant direct and indirect effects on the employment rate. Conversely, in the less developed central and western regions, an increase in the minimum wage directly boosts the employment rate of that city but has an insignificant indirect effect on the employment rate of neighboring cities.

The increase in a city's minimum wage has the potential to attract a larger labor force from other cities, as noted by Li et al. (2019). This may explain the significant positive direct and negative indirect effects on the employment rate observed when analyzing entire country data. The prospect of a higher minimum wage often prompts population migration across different regions, leading to an increase in the employment rate in cities where the minimum wage has risen, but a decrease in cities where it has fallen.

However, when analyzing regional data, the study imposes constraints by confining populations to regions with similar economic levels. Consequently, the indirect effect of an improved minimum wage on the employment rate becomes insignificant because populations cannot migrate across regions. Conversely, the direct effects of an enhanced minimum wage on the employment rate are significantly positive in underdeveloped central and western regions, aligning with the findings of Ni et al. (2011).

Based on the results, here are three policy suggestions for the Chinese central government. Firstly, the central government should increase investment in infrastructure construction in the less developed central and western regions, attract more enterprises to create job opportunities, strengthen basic education, and narrow the development gap between these regions and the more prosperous eastern regions. This initiative aims to retain more young labor forces in the central and western regions and further diminish the economic development disparities across the country.

Secondly, the central government should facilitate labor mobility by addressing the constraints imposed by the household registration system (Hukou system), which impedes migration. For instance, rural Hukou populations may lack access to healthcare, or their children may face difficulties registering in local schools. Therefore, comprehensive reform of the Hukou system is necessary to enable migrants to integrate into urban life seamlessly.

Thirdly, the central government should acknowledge the regional disparities in employment responses to minimum wage adjustments and contemplate implementing a flexible minimum wage structure. This structure should consider the cost of living and economic development levels of each region. By doing so, the government can promote equitable labor practices and support sustainable economic growth throughout the nation.

4.7 Appendix

Figure 4.2: Log minimum wage local Moran's I test

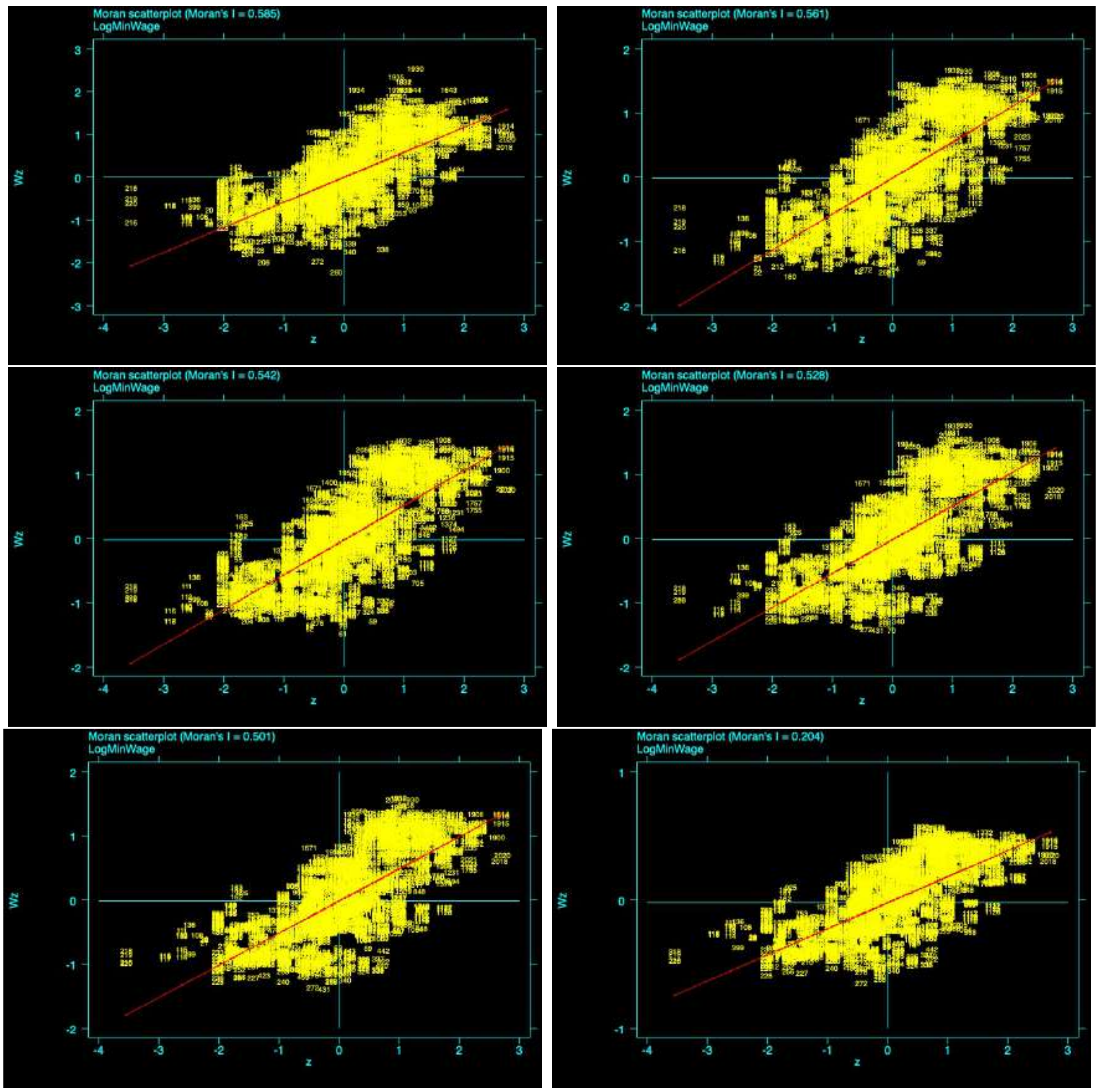


Figure 4.3: Log employment rate local Moran's I test

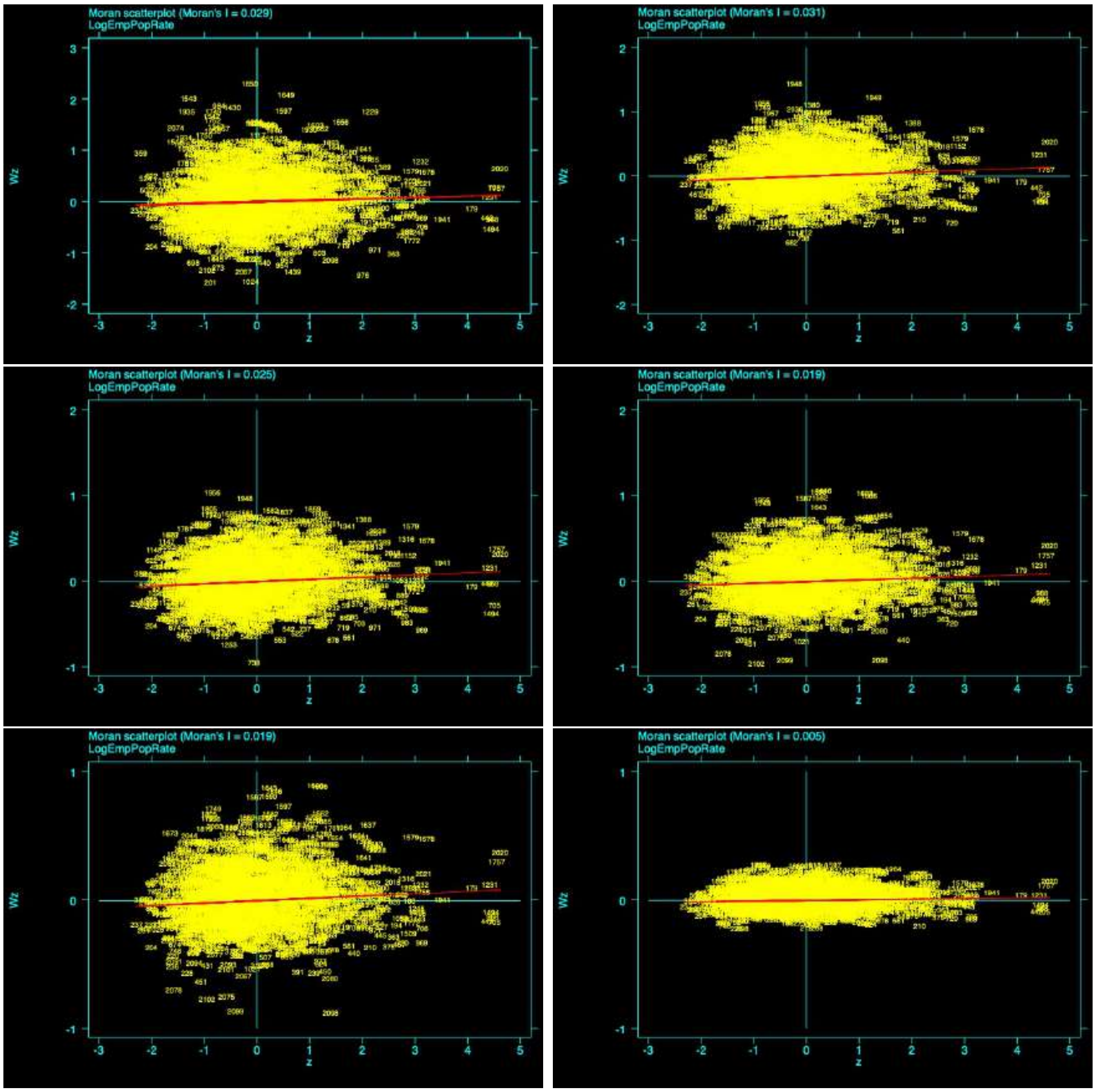


Table 4.8: East region analysis, controlling city- and time- fixed effects

Dependent variable: log employment rate				
Spatial weighted matrix:	Pooled OLS	Binary contiguity	10 neighbors	15 neighbors
Log minimum wage	0.046 (0.394)	-0.112 (0.509)	-0.005 (0.979)	0.007 (0.968)
Log rural population rate	0.038 (0.251)	0.021 (0.457)	0.021 (0.492)	0.028 (0.361)
Log education rate	0.033 (0.361)	0.019 (0.577)	0.025 (0.463)	0.024 (0.437)
Log GDP per capita	0.230*** (0.005)	0.185** (0.012)	0.200*** (0.004)	0.218*** (0.001)
Log agriculture GDP per capita	0.073 (0.521)	0.161 (0.143)	0.059 (0.605)	0.077 (0.478)
Log hospital beds per capita	-0.188** (0.046)	-0.206** (0.016)	-0.212** (0.010)	-0.235*** (0.003)
W*Log minimum wage rate	N.A.	0.063 (0.775)	-0.074 (0.733)	-0.187 (0.415)
W*Log rural population rate	N.A.	-0.031 (0.366)	-0.184 (0.221)	-0.403** (0.026)
W*Log education rate	N.A.	0.060 (0.472)	0.095 (0.276)	0.072 (0.529)
W*Log GDP per capita	N.A.	0.153 (0.188)	0.021 (0.918)	-0.241 (0.461)
W*Log agriculture GDP per capita	N.A.	-0.087 (0.630)	0.008 (0.975)	-0.243 (0.504)
W*Log hospital beds per capita	N.A.	-0.010 (0.942)	0.225 (0.379)	0.649* (0.092)
ρ	N.A.	0.177** (0.028)	0.340*** (0.000)	0.395*** (0.000)
log minimum wage rate				
Direct effect	N.A.	-0.103 (0.538)	-0.001 (0.996)	0.007 (0.968)
Indirect effect	N.A.	0.045 (0.844)	-0.109 (0.677)	-0.300 (0.337)
Total effect	N.A.	-0.058 (0.738)	-0.110 (0.625)	-0.293 (0.319)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Table 4.9: Northeast region analysis, controlling city- and time- fixed effects

Dependent variable: log employment rate				
Spatial weighted matrix:	Pooled OLS	Binary contiguity	10 neighbors	15 neighbors
Log minimum wage	0.024*	0.441**	0.235	0.200
	(0.065)	(0.013)	(0.123)	(0.129)
Log rural population rate	-0.226***	-0.165**	-0.205***	-0.141*
	(0.004)	(0.012)	(0.002)	(0.079)
Log education rate	-0.018	-0.027	-0.023	-0.020
	(0.659)	(0.507)	(0.549)	(0.565)
Log GDP per capita	0.519***	0.439***	0.502***	0.335***
	(0.001)	(0.003)	(0.001)	(0.009)
Log agriculture GDP per capita	-0.118	-0.031	-0.157	-0.158
	(0.383)	(0.791)	(0.230)	(0.165)
Log hospital beds per capita	-0.008	-0.003	-0.020	-0.150***
	(0.653)	(0.871)	(0.326)	(0.000)
W*Log minimum wage rate	N.A.	-0.270	0.109	0.128
		(0.365)	(0.727)	(0.735)
W*Log rural population rate	N.A.	-1.114***	0.309	1.492
		(0.002)	(0.669)	(0.139)
W*Log education rate	N.A.	0.045	-0.187	0.222
		(0.508)	(0.133)	(0.126)
W*Log GDP per capita	N.A.	-0.162	-0.343	-3.288***
		(0.578)	(0.505)	(0.002)
W*Log agriculture GDP per capita	N.A.	-0.392	-1.425**	-1.013
		(0.147)	(0.022)	(0.329)
W*Log hospital beds per capita	N.A.	0.085	0.325	-1.600***
		(0.334)	(0.194)	(0.009)
ρ	N.A.	0.113	-0.290	-0.738***
		(0.106)	(0.111)	(0.000)
log minimum wage rate				
Direct effect	N.A.	0.440**	0.241	0.205
		(0.013)	(0.136)	(0.152)
Indirect effect	N.A.	-0.262	0.022	-0.026
		(0.405)	(0.937)	(0.924)
Total effect	N.A.	0.179	0.263	0.180
		(0.483)	(0.197)	(0.341)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Table 4.10: Central region analysis, controlling city- and time- fixed effects

Dependent variable: log employment rate				
Spatial weighted matrix:	Pooled OLS	Binary contiguity	10 neighbors	15 neighbors
Log minimum wage	0.059*	0.105	0.249*	0.240*
	(0.081)	(0.424)	(0.080)	(0.062)
Log rural population rate	0.006	0.007	0.006	0.001
	(0.345)	(0.245)	(0.398)	(0.881)
Log education rate	-0.019	-0.033	-0.031	-0.027
	(0.696)	(0.520)	(0.498)	(0.581)
Log GDP per capita	-0.046	0.005	0.0002	-0.027
	(0.890)	(0.988)	(0.999)	(0.937)
Log agriculture GDP per capita	-0.167	-0.165	-0.218	-0.194
	(0.283)	(0.288)	(0.155)	(0.214)
Log hospital beds per capita	0.292	0.278	0.288	0.302
	(0.159)	(0.153)	(0.131)	(0.105)
W*Log minimum wage rate	N.A.	0.038	-0.227	-0.282
		(0.811)	(0.162)	(0.165)
W*Log rural population rate	N.A.	-0.012	-0.030	0.006
		(0.563)	(0.595)	(0.920)
W*Log education rate	N.A.	0.003	0.153	0.020
		(0.969)	(0.397)	(0.947)
W*Log GDP per capita	N.A.	-0.158	-0.092	-0.128
		(0.455)	(0.704)	(0.683)
W*Log agriculture GDP per capita	N.A.	0.072	0.482*	0.486
		(0.699)	(0.079)	(0.104)
W*Log hospital beds per capita	N.A.	0.235	0.611	0.916*
		(0.365)	(0.142)	(0.063)
ρ	N.A.	0.005	-0.250**	-0.450**
		(0.935)	(0.014)	(0.010)
log minimum wage rate				
Direct effect	N.A.	0.110	0.260*	0.253*
		(0.414)	(0.082)	(0.064)
Indirect effect	N.A.	0.036	-0.245	-0.288
		(0.819)	(0.108)	(0.093)
Total effect	N.A.	0.147	0.015	-0.035
		(0.301)	(0.887)	(0.770)

Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.

Table 4.11: West region analysis, controlling city- and time- fixed effects

Dependent variable: log employment rate				
Spatial weighted matrix:	Pooled OLS	Binary contiguity	10 neighbors	15 neighbors
Log minimum wage	0.012	0.100	0.177**	0.221*
	(0.538)	(0.235)	(0.022)	(0.071)
Log rural population rate	-0.063***	-0.059***	-0.055***	-0.041
	(0.000)	(0.000)	(0.000)	(0.008)
Log education rate	-0.007	-0.009	-0.005	-0.008
	(0.753)	(0.712)	(0.806)	(0.742)
Log GDP per capita	0.016	0.021	0.014	0.048
	(0.834)	(0.803)	(0.845)	(0.515)
Log agriculture GDP per capita	-0.153	-0.093	-0.088	-0.081
	(0.243)	(0.520)	(0.554)	(0.565)
Log hospital beds per capita	0.179**	0.149	0.158	0.141
	(0.049)	(0.141)	(0.127)	(0.119)
W*Log minimum wage rate	N.A.	-0.061	-0.231	-0.363
		(0.627)	(0.138)	(0.139)
W*Log rural population rate	N.A.	-0.091	-0.025	0.195
		(0.327)	(0.889)	(0.410)
W*Log education rate	N.A.	0.075*	0.203***	0.179
		(0.085)	(0.001)	(0.027)
W*Log GDP per capita	N.A.	-0.133	-0.155	-0.707
		(0.485)	(0.452)	(0.173)
W*Log agriculture GDP per capita	N.A.	-0.060	0.072	-0.117
		(0.776)	(0.797)	(0.728)
W*Log hospital beds per capita	N.A.	0.124	0.350	0.151
		(0.496)	(0.187)	(0.711)
ρ	N.A.	0.067	-0.050	-0.093
		(0.218)	(0.541)	(0.324)
log minimum wage rate				
Direct effect	N.A.	0.102**	0.181**	0.227*
		(0.230)	(0.023)	(0.073)
Indirect effect	N.A.	-0.056	-0.230	-0.359
		(0.670)	(0.129)	(0.125)
Total effect	N.A.	0.046	-0.048	-0.131
		(0.715)	(0.729)	(0.463)
Notes: ***indicates significance at the 1% confidence level, **at 5%, and *at 10%.				

Chapter 5

Conclusion

Chapter 1 highlights four primary findings. First, the global and local spatial Moran's I tests reveal significant spatial autocorrelation in $PM_{2.5}$, underscoring the justification for utilizing spatial models. Notably, the paper's results demonstrate a higher level of robustness compared to prior studies, attributed to its effective handling of spatial dependence among $PM_{2.5}$ levels in various cities.

The second finding underscores a significant indirect effect of $PM_{2.5}$ on sleeplessness. Specifically, a one-unit increase in $PM_{2.5}$ within a city results in an increase in sleeplessness among Weibo users in neighboring cities by three individuals. Notably, this effect is observed before accounting for province-by-season fixed effects and meteorological factors. This contrasts with prior studies (Heyes & Zhu, 2019; Strøm-Tejsen et al., 2016; Lawrence et al., 2018; Billings et al., 2019), which predominantly focused on the direct effects of $PM_{2.5}$ on sleep, so this paper provides additional empirical evidence on the spatial spillover effects of $PM_{2.5}$ on individual sleeplessness. Importantly, in this study, the direct effect of $PM_{2.5}$ on sleeplessness within a city is found to be statistically insignificant.

The third finding is the estimation using winter data and controlling meteorological factors reveals that a statistically significant positive indirect effect of $PM_{2.5}$ on the sleeplessness of Weibo users. Additionally, an insignificant direct effect estimates of $PM_{2.5}$ on the sleeplessness of Weibo users is observed. Quantitatively, a one-unit increase in $PM_{2.5}$ in a city leads to a two-person increase in the sleeplessness of Weibo users in that city's neighboring cities. However, the sleeplessness of Weibo users in that city itself does not appear to be affected.

The fourth finding suggests that the estimates from the Spatial Durbin Model (SDM) is more accurate than those from the Spatial Error Model (SEM) and the Spatial Autoregressive Model (SAR).

By the way, I also estimated air quality index (AQI) the other air pollutants' (PM_{10}, SO_2, CO, and O_3) effect on sleeplessness of Weibo users in winter. The results of the SDM model indicate that both AQI and PM_{10} have a significant positive indirect effect on the sleeplessness of Weibo users. However, the magnitudes of the effects of AQI and PM_{10} are smaller than that of $PM_{2.5}$.

In Chapter 2, the estimated results from standard DID and spatial DID models indicate that the implementation of the RAPM policy has significant (at the 10% level) and positive ADTE on the average CSR scores of treated cities. However, it shows insignificant AITE on the average CSR scores of neighboring cities of the treated cities. When employing the DSDID model for analysis, both the estimated ADTE and AITE are found to be insignificant. This suggests that the spatial dependence of the error term has unobserved disruptive effects on the estimations.

When using the SDEM-DID model, 800KM contiguity spatial weighted matrix, to analyze the entire dataset before controlling for other moderator variables and FDI, the implementation of RAPM does not seem to affect the average CSR or CER scores of treated cities. However, it leads to a decrease of 2.1 and 0.6 points in the average CSR and CER scores of treated cities' neighboring cities. When employing the DSDEM-DID model to estimate, ADTE of RAPM on treated cities remain insignificant, and AITE on treated cities' treated neighboring cities' average CSR and CER are also insignificant. However, the AITE of RAPM on treated cities' untreated neighboring cities' average CSR and CER are significant and negative. This aligns with the argument presented by Chagas et al. (2016), suggesting that after the implementation of RAPM, the AITE of the RAPM policy is stronger on the average CSR and CER scores of treated cities' untreated neighboring cities than on the average CSR and CER scores of treated cities' treated neighboring cities. The average CSR and CER scores of treated cities' untreated neighboring cities witness a decrease of 1.5 points and 0.4 points, respectively.

The AITE of the RAPM policy on CSR and CER scores of treated cities' neighboring cities is lower in the DSDEM-DID model than in the SDEM-DID model.

This difference arises because the AITE of the RAPM in the SDEM-DID model includes RAPM's spatial spillover effects on all neighbors of treated cities, while the RAPM's spatial spillover effects in the DSDEM-DID model are only on treated cities' untreated neighbors.

After controlling for other moderator variables but without considering FDI, the DSDEM-DID model still shows that the implementation of RAPM has insignificant ADTE and AITE on treated cities' and their treated neighboring cities' average CSR and CER scores. However, the AITEs of RAPM on treated cities' untreated neighboring cities' average CSR and CER scores remain significant. After introducing control for FDI, the AITEs of RAPM on treated cities' and treated cities' untreated neighboring cities' average CSR and CER scores become insignificant.

When utilizing data from the eastern and central regions and controlling for all other moderator variables, including FDI, the DSDEM-DID models indicate that the AITEs of RAPM on treated cities' and treated cities' untreated neighboring cities' average CSR and CER scores are significant and negative.

When employing the 800KM inverse distance spatial weighted matrix for analysis, the DSDEM-DID models show that the AITEs of RAPM on treated cities' and treated cities' untreated neighboring cities' average CSR and CER scores are significant only when not controlling for other variables and using the entire dataset for analysis. The AITEs of RAPM become insignificant when focusing on the eastern and central regions.

In summary, the implementation of the real-time air pollution monitoring policy does not have a significant impact on the CSR or CER scores of the treated cities. However, it does lead to a decrease in the CSR or CER scores of the treated cities' neighboring cities, particularly those untreated neighboring cities, whether considered on a national scale or within the eastern and central regions. This phenomenon suggests that companies located in the treated cities may respond to the policy by changing suppliers or relocating manufacturing plants to cities without such stringent environmental monitoring. Consequently, these companies could reduce their environmental protection costs. However, this shift may negatively impact the CSR or CER scores of listed companies in cities that have not implemented the policy.

Chapter 3 yields three key results. Firstly, the global and local spatial Moran's

I tests reveal spatial autocorrelation in both log minimum wage and log employment rate, substantiating the rationale for adopting spatial models. The robustness of the results is enhanced compared to conventional models, particularly considering the spatial autocorrelation of log minimum wage and log employment rate, a consideration addressed by the SDM model.

Secondly, employing the entire dataset to estimate the SDM model with 10 and 15 neighbors' spatial weighted matrices while controlling for city and time fixed effects, both the direct and indirect minimum wage elasticity of the employment rate are found to be significantly positive and negative, respectively. In essence, when the minimum wage in a city increases, the employment rate in that city rises, while the employment rate in neighboring cities decreases. Taking the results from the SDM model with the 10 neighbors' spatial weighted matrix as an illustration, the direct and indirect minimum wage elasticity of the employment rate are positive and negative values, 0.164 (significant at the 1% level) and -0.201 (significant at the 5% level), respectively. This implies that a 10% increase in the minimum wage in a city would lead to a 1.64% increase in its employment rate but a 2.01% decrease in the employment rate of its neighboring cities, holding the other variables constant.

Thirdly, in the economically more developed eastern and northeastern regions, an increase in the minimum wage has an insignificant direct and indirect effects on the employment rate. Conversely, in the less developed central and western regions, an increase in the minimum wage directly boosts the employment rate of that city but has an insignificant indirect effect on the employment rate of neighboring cities.

The increase in a city's minimum wage has the potential to attract a larger labor force from other cities, as noted by Li et al. (2019). This may explain the significant positive direct and negative indirect effects on the employment rate observed when analyzing entire country data. The prospect of a higher minimum wage often prompts population migration across different regions, leading to an increase in the employment rate in cities where the minimum wage has risen, but a decrease in cities where it has fallen.

However, when analyzing regional data, the study imposes constraints by confining populations to regions with similar economic levels. Consequently, the indirect effect of an improved minimum wage on the employment rate becomes insignificant because populations cannot migrate across regions. Conversely, the direct effects of an enhanced minimum wage on the employment rate are significantly positive in underdeveloped central and western regions, aligning with the findings of Ni et al. (2011).

References

Akinseye, O. A., Williams, S. K., Seixas, A., Pandi-Perumal, S. R., Vallon, J., Zizi, F., & Jean-Louis, G. (2015). Sleep as a mediator in the pathway linking environmental factors to hypertension: a review of the literature. *International Journal of Hypertension*, *2015*(926414).

Allegretto, S. A., Dube, A., & Reich, M. (2011). Do minimum wages really reduce teen employment? accounting for heterogeneity and selectivity in state panel data. *Industrial Relations: A Journal of Economy and Society*, *50*(2), 205–240.

Almunawar, M. N., & Low, K. C. P. (2013). Corporate social responsibility and sustainable development: Trends in asia. In *Corporate social responsibility in asia: Practice and experience* (pp. 173–196). Springer.

Amez, S., Vujic, S., Abrath, M., & Baert, S. (2021). *Deteriorated sleep quality does not explain the negative impact of smartphone use on academic performance* (Discussion Paper No. 14547). IZA Discussion Paper.

Anselin, L. (1995). Local indicators of spatial association—lisa. *Geographical Analysis*, *27*(2), 93–115.

Anselin, L. (2003). Spatial externalities. *International Regional Science Review*, *26*(2), 147–152.

Anselin, L., Gallo, J. L., & Jayet, H. (2008). Spatial panel econometrics. In *The econometrics of panel data: Fundamentals and recent developments in theory and practice* (pp. 625–660). Springer.

Anselin, L., & Rey, S. J. (1997). Introduction to the special issue on spatial econometrics. *International Regional Science Review*, *20*(1-2), 1–7.

Ashenfelter, O. C., & Card, D. (1984). *Using the longitudinal structure of earnings to estimate the effect of training programs.* National Bureau of Economic Research Cambridge, Mass., USA.

Bardaka, E., Delgado, M. S., & Florax, R. J. (2018). Causal identification of transit-induced gentrification and spatial spillover effects: The case of the denver light rail. *Journal of Transport Geography*, *71*, 15–31.

Billings, M. E., Gold, D., Szpiro, A., Aaron, C. P., Jorgensen, N., Gassett, A., ... Redline, S. R. (2019). The association of ambient air pollution with sleep apnea: the multi-ethnic study of atherosclerosis. *Annals of the American Thoracic Society*, *16*(3), 363–370.

Bockstael, N. E. (1996). Modeling economics and ecology: the importance of a spatial perspective. *American Journal of Agricultural Economics*, *78*(5),

1168–1180.
Brown, C., Gilroy, C., & Kohen, A. (1981). *Time-series evidence of the effect of the minimum wage on youth employment and unemployment* (Tech. Rep. No. w0790). National Bureau of Economic Research.
Burridge, P. (1981). Testing for a common factor in a spatial autoregression model. *Environment and Planning A*, *13*(7), 795–800.
Buysse, D. J., Reynolds III, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Research*, *28*(2), 193–213.
Cai, H., Nan, Y., Zhao, Y., Jiao, W., & Pan, K. (2020). Impacts of winter heating on the atmospheric pollution of northern china's prefectural cities: evidence from a regression discontinuity design. *Ecological Indicators*, *118*, 106709.
Calderón-Garcidueñas, L., Leray, E., Heydarpour, P., Torres-Jardón, R., & Reis, J. (2016). Air pollution, a rising environmental risk factor for cognition, neuroinflammation and neurodegeneration: The clinical impact on children and beyond. *Revue Neurologique*, *172*(1), 69–80.
Cao, B., Chen, Y., & McIntyre, R. S. (2021). Comprehensive review of the current literature on impact of ambient air pollution and sleep quality. *Sleep Medicine*, *79*, 211–219.
Chagas, A. L., Azzoni, C. R., & Almeida, A. N. (2016). A spatial difference-in-differences analysis of the impact of sugarcane production on respiratory diseases. *Regional Science and Urban Economics*, *59*, 24–36.
Chai, K.-C., Xie, D.-C., Yeh, C.-P., Lan, H.-R., & Cui, Z.-X. (2022). Chinese national civilized city and corporate social responsibility: will civilized city promote corporate social responsibility? *Applied Economics Letters*, *29*(7), 593–596.
Chaix, B., Merlo, J., Subramanian, S., Lynch, J., & Chauvin, P. (2005). Comparison of a spatial perspective with the multilevel analytical approach in neighborhood studies: the case of mental and behavioral disorders due to psychoactive substance use in malmö, sweden, 2001. *American Journal of Epidemiology*, *162*(2), 171–182.
Chan, C. K., & Yao, X. (2008). Air pollution in mega cities in china. *Atmospheric Environment*, *42*(1), 1–42.
Chen, S., Guo, C., & Huang, X. (2018). Air pollution, student health, and school absences: Evidence from china. *Journal of Environmental Economics and Management*, *92*, 465–497.
Chen, X., Shao, S., Tian, Z., Xie, Z., & Yin, P. (2017). Impacts of air pollution and its spatial spillover effect on public health based on china's big data sample. *Journal of Cleaner Production*, *142*, 915–925.
Chen, X., & Ye, J. (2019). When the wind blows: Spatial spillover effects of urban air pollution in china. *Journal of Environmental Planning and Management*, *62*(8), 1359–1376.
Chen, Y., Ebenstein, A., Greenstone, M., & Li, H. (2013). 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), 12936–

12941.

Chen, Y.-C., Hung, M., & Wang, Y. (2018). The effect of mandatory csr disclosure on firm profitability and social externalities: Evidence from china. *Journal of accounting and economics*, *65*(1), 169–190.

Cheng, W.-J., Liang, S.-J., Huang, C.-S., Lin, C.-L., Pien, L.-C., & Hang, L.-W. (2019). Air pollutants are associated with obstructive sleep apnea severity in non-rapid eye movement sleep. *Journal of Clinical Sleep Medicine*, *15*(6), 831–837.

Chiu, C., Lin, D., & Silverman, A. (2012). *China's social-media boom.* McKinsey & Company.

Cole, M. A., Elliott, R. J., Okubo, T., & Zhou, Y. (2013). The carbon dioxide emissions of firms: A spatial analysis. *Journal of Environmental Economics and Management*, *65*(2), 290–309.

Costa-i Font, J., Fleche, S., & Pagan, R. (2021). *The welfare effects of time reallocation: Evidence from daylight saving time* (Working Paper No. 9195). CESifo Working Paper.

Delgado, M. S., & Florax, R. J. (2015). Difference-in-differences techniques for spatial data: Local autocorrelation and spatial interaction. *Economics Letters*, *137*, 123–126.

Dickinson, D. L., McEvoy, D. M., & Bruner, D. M. (2022). The impact of sleep restriction on interpersonal conflict resolution and the narcotic effect. *Journal of Economic Behavior & Organization*, *194*, 71–90.

Doherty, M. J., Youn, C. E., Haltiner, A. M., & Watson, N. F. (2010). Do weather-related ambient atmospheric-pressure changes influence sleep disordered breathing? *Journal of Clinical Sleep Medicine*, *6*(2), 152–156.

Dolton, P., Bondibene, C. R., & Stops, M. (2015). Identifying the employment effect of invoking and changing the minimum wage: A spatial analysis of the uk. *Labour Economics*, *37*, 54–76.

Elhorst, J. (2010). *Spatial panel data models. fischer mm, a getis, editor, handbook of applied spatial analysis.* New York: Springer.

Elhorst, J. P. (2014). Spatial econometrics: from cross-sectional data to spatial panels. In (Vol. 479, pp. 37–88). Springer.

Elhorst, J. P. (2017). Spatial panel data analysis. *Encyclopedia of GIS*, *2*, 2050–2058.

Fan, M., He, G., & Zhou, M. (2020). The winter choke: Coal-fired heating, air pollution, and mortality in china. *Journal of Health Economics*, *71*, 102316.

Fang, T., & Lin, C. (2015). Minimum wages and employment in china. *IZA Journal of Labor Policy*, *4*, 1–30.

Feng, Y., Cheng, J., Shen, J., & Sun, H. (2019). Spatial effects of air pollution on public health in china. *Environmental and Resource Economics*, *73*, 229–250.

Fetscherin, M., Voss, H., & Gugler, P. (2010). 30 years of foreign direct investment to china: An interdisciplinary literature review. *International business review*, *19*(3), 235–246.

Fingleton, B., & Le Gallo, J. (2010). Endogeneity in a spatial context: properties of estimators. In (pp. 59–73). Springer.

Fowler, C. S., Frey, N., Folch, D. C., Nagle, N., & Spielman, S. (2020). Who are the people in my neighborhood?: The "contextual fallacy" of measuring individual context with census geographies. *Geographical Analysis*, *52*(2), 155–168.

Fu, S., Viard, V. B., & Zhang, P. (2021). Air pollution and manufacturing firm productivity: Nationwide estimates for china. *Economic Journal*, *131*(640).

Gibson, M., & Shrader, J. (2018). Time use and labor productivity: The returns to sleep. *Review of Economics and Statistics*, *100*(5), 783–798.

Gilles, S., Akdis, C., Lauener, R., Schmid-Grendelmeier, P., Bieber, T., Schäppi, G., & Traidl-Hoffmann, C. (2018). The role of environmental factors in allergy: A critical reappraisal. *Experimental Dermatology*, *27*(11), 1193–1200.

Hafner, M., Stepanek, M., Taylor, J., Troxel, W. M., & Van Stolk, C. (2017). Why sleep matters—the economic costs of insufficient sleep: a cross-country comparative analysis. *Rand Health Quarterly*, *6*(4).

Halleck Vega, S., & Elhorst, J. P. (2015). The slx model. *Journal of Regional Science*, *55*(3), 339–363.

Han, Y., Li, H., Xiao, Y., Li, A., & Zhu, T. (2021). Influential path of social risk factors toward suicidal behavior—evidence from chinese sina weibo users 2013–2018. *International Journal of Environmental Research and Public Health*, *18*(5), 2604.

Han, Z., & An, N. (2007). Minimum wage and local employment: A spatial panel approach. *Jinan Journal*, *29*(1), 38-44.

Hartescu, I., Morgan, K., & Stevinson, C. D. (2015). Increased physical activity improves sleep and mood outcomes in inactive people with insomnia: a randomized controlled trial. *Journal of sleep research*, *24*(5), 526–534.

He, X., Luo, Z., & Zhang, J. (2022). The impact of air pollution on movie theater admissions. *Journal of Environmental Economics and Management*, *112*, 102626.

Heyes, A., & Zhu, M. (2019). Air pollution as a cause of sleeplessness: Social media evidence from a panel of chinese cities. *Journal of Environmental Economics and Management*, *98*, 102247.

Huang, Q., & Chand, S. (2015). Spatial spillovers of regional wages: Evidence from chinese provinces. *China Economic Review*, *32*, 97–109.

Jin, Y., Andersson, H., & Zhang, S. (2016). Air pollution control policies in china: a retrospective and prospects. *International journal of environmental research and public health*, *13*(12), 1219.

Kalenkoski, C. M., & Lacombe, D. J. (2008). Effects of minimum wages on youth employment: the importance of accounting for spatial correlation. *Journal of Labor Research*, *29*, 303–317.

Kalenkoski, C. M., & Lacombe, D. J. (2013). Minimum wages and teen employment: A spatial panel approach. *Papers in Regional Science*, *92*(2), 407–417.

Kim, C.-U., Lim, G., et al. (2018). Minimum wage and unemployment: An empirical study on oecd countries. *Journal of Reviews on Global Economics*, *7*, 1–9.

Kong, D., Cheng, X., & Jiang, X. (2021). Effects of political promotion on local firms' social responsibility in china. *Economic Modelling*, *95*, 418–429.

Lawrence, W. R., Yang, M., Zhang, C., Liu, R.-Q., Lin, S., Wang, S.-Q., ... Zeng, X.-W. (2018). Association between long-term exposure to air pollution and sleep disorder in chinese children: the seven northeastern cities study. *Sleep*, *41*(9), 1–10.

LeSage, J., & Pace, R. K. (2009). Introduction to spatial econometrics. In (p. 33-41). Chapman and Hall/CRC.

LeSage, J. P., & Pace, R. K. (2008). Spatial econometric modeling of origin-destination flows. *Journal of Regional Science*, *48*(5), 941–967.

Li, Y., Kanbur, R., & Lin, C. (2019). Minimum wage competition between local governments in china. *The Journal of Development Studies*, *55*(12), 2479–2494.

Liu, Z. (2005). Institution and inequality: the hukou system in china. *Journal of comparative economics*, *33*(1), 133–157.

Luo, X. (2007). Analysis of the employment effect of minimum wage under the monopsony of labor market in china. *Finance and Trade Research*, *4*(1).

Ma, Y.-R., Ji, Q., & Fan, Y. (2016). Spatial linkage analysis of the impact of regional economic activities on pm2. 5 pollution in china. *Journal of Cleaner Production*, *139*, 1157–1167.

Majchrowska, A., & Strawiński, P. (2021). Minimum wage and local employment: A spatial panel approach. *Regional Science Policy & Practice*, *13*(5), 1581–1602.

Mou, L., & Wang, G. (2015). The study of minimum wage standard's effect on employment in china. *Journal of Wuhan University*, *15*(5), 936–941.

Ni, J., Wang, G., & Yao, X. (2011). Impact of minimum wages on employment: evidence from china. *Chinese Economy*, *44*(1), 18–38.

Nuvolone, D., Petri, D., & Voller, F. (2018). The effects of ozone on human health. *Environmental Science and Pollution Research*, *25*, 8074–8088.

Obradovich, N., Migliorini, R., Mednick, S. C., & Fowler, J. H. (2017). Nighttime temperature and human sleep loss in a changing climate. *Science Advances*, *3*(5), e1601555.

Pace, R. K., & LeSage, J. P. (2008). Biases of ols and spatial lag models in the presence of an omitted variable and spatially dependent variables. *Available at SSRN 1133438*.

Patella, V., Florio, G., Magliacane, D., Giuliano, A., Crivellaro, M. A., Di Bartolomeo, D., ... (SIAAIC), C. I. (2018). Urban air pollution and climate change:"the decalogue: Allergy safe tree" for allergic and respiratory diseases care. *Clinical and Molecular Allergy*, *16*(1), 1–11.

Pavlyuk, D., et al. (2011). Spatial analysis of regional employment rates in latvia. *Sustainable spatial development*, *2*, 56–62.

Pinkse, J., & Slade, M. E. (2010). The future of spatial econometrics. *Journal of Regional Science*, *50*(1), 103–117.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, 34–58.

Rubin, D. B. (1991). Formal modes of statistical inference for causal effects. *Quality control and applied statistics*, *36*(3), 185–188.

Sánchez, T., Gozal, D., Smith, D. L., Foncea, C., Betancur, C., & Brockmann, P. E. (2019). Association between air pollution and sleep disordered breathing in children. *Pediatric Pulmonology*, *54*(5), 544–550.

Shi, J. (2010). Empirical study on the relationship between minimum wage and employment. *Journal of Guangxi University of Finance and Economics*, *23*(1), 93–97.

Shi, J. (2011). Employment effect of minimum wage in china–research based on construction and manufacturing. In *Advances in applied economics, business and development: International symposium, isaebd 2011, dalian, china, august 6-7, 2011, proceedings, part ii* (pp. 78–82).

Smith, A. C. (2016). Spring forward at your own risk: Daylight saving time and fatal vehicle crashes. *American Economic Journal: Applied Economics*, *8*(2), 65–91.

Song, C., Wang, Y., Yang, X., Yang, Y., Tang, Z., Wang, X., & Pan, J. (2020). Spatial and temporal impacts of socioeconomic and environmental factors on healthcare resources: a county-level bayesian local spatiotemporal regression modeling study of hospital beds in southwest china. *International Journal of Environmental Research and Public Health*, *17*(16), 5890.

Strøm-Tejsen, P., Zukowska, D., Wargocki, P., & Wyon, D. P. (2016). The effects of bedroom air quality on sleep and next-day performance. *Indoor Air*, *26*(5), 679–686.

Sui, D. Z. (2004). Tobler's first law of geography: A big idea for a small world? *Annals of the Association of American Geographers*, *94*(2), 269–277.

Sun, W., Wang, X., & Zhang, X. (2015). Minimum wage effects on employment and working time of chinese workers——evidence based on chns. *IZA Journal of Labor & Development*, *4*(1), 1–22.

Sun, Y., Shi, L., Bao, Y., Sun, Y., Shi, J., & Lu, L. (2018). The bidirectional relationship between sleep duration and depression in community-dwelling middle-aged and elderly individuals: evidence from a longitudinal study. *Sleep Medicine*, *52*, 221–229.

Swandari, F., & Sadikin, A. (2016). The effect of ownership structure, profitability, leverage, and firm size on corporate social responsibility (csr). *Binus Business Review*, *7*(3), 315–320.

Tang, D., Li, L., & Yang, Y. (2016). Spatial econometric model analysis of foreign direct investment and haze pollution in china. *Polish Journal of Environmental Studies*, *25*(1).

Tian, F., & Pan, J. (2021). Hospital bed supply and inequality as determinants of maternal mortality in china between 2004 and 2016. *International journal for equity in health*, *20*(1), 1–15.

Tobler, W. R. (1970). *Spectral analysis of spatial series.* Library Photographic Service, U. of California.

Tosetti, E., Santos, R., Moscone, F., & Arbia, G. (2018). The spatial dimension of health systems. In *Oxford research encyclopedia of economics and finance.*

Totty, E. (2017). The effect of minimum wages on employment: A factor model approach. *Economic Inquiry*, *55*(4), 1712–1737.

Tran, N. H., Fu, L., & Boehe, D. M. (2023). How does urban air pollution affect corporate environmental performance? *Journal of Cleaner Production*, *383*, 135443.

Vaske, J. J. (2011). Advantages and disadvantages of internet surveys: Introduction to the special issue. *Human Dimensions of Wildlife*, *16*(3), 149–153.

Vom Berge, P., & Frings, H. (2020). High-impact minimum wages and heterogeneous regions. *Empirical Economics*, *59*(2), 701–729.

Wang, G., & Yao, X. (2014). Minimum wage and employment based on dynamic equilibrium. *Journal of Zhejiang University of Science and Technology*, *4*, 258–264.

Wang, J., & Gunderson, M. (2011). Minimum wage impacts in china: Estimates from a prespecified research design, 2000–2007. *Contemporary Economic Policy*, *29*(3), 392–406.

Wang, W., Lin, Y., Yang, H., Ling, W., Liu, L., Zhang, W., ... Jiang, G. (2022). Internal exposure and distribution of airborne fine particles in the human body: Methodology, current understandings, and research needs. *Environmental Science & Technology*, *56*(11), 6857–6869.

Williams, A. M., & Phaneuf, D. J. (2019). The morbidity costs of air pollution: Evidence from spending on chronic respiratory conditions. *Environmental and Resource Economics*, *74*, 571–603.

Xiao, Q., Ma, Z., Li, S., & Liu, Y. (2015). The impact of winter heating on air pollution in china. *PloS One*, *10*(1), e0117311.

Xiao, X.-Y., & Xiang, B.-L. (2009). The impact of minimum wage policy on wages and employment in china. In *2009 international conference on information management, innovation management and industrial engineering* (Vol. 2, pp. 102–105).

Xu, H. (2014). Comparing spatial and multilevel regression models for binary outcomes in neighborhood studies. *Sociological Methodology*, *44*(1), 229–272.

Yahong, W., & Khan, S. (2021). A cross-sectional analysis of employment returns to education and health status in china: moderating role of gender. *Frontiers in Psychology*, *12*, 638599.

Yang, R., & Zhong, C. (2022). Analysis on spatio-temporal evolution and influencing factors of air quality index (aqi) in china. *Toxics*, *10*(12), 712.

Zangger, C. (2019). Making a place for space: Using spatial econometrics to model neighborhood effects. *Journal of Urban Affairs*, *41*(8), 1055–1080.

Zanobetti, A., Redline, S., Schwartz, J., Rosen, D., Patel, S., O'Connor, G. T., ... Gold, D. R. (2010). Associations of pm10 with sleep and sleep-disordered breathing in adults from seven us urban areas. *American Journal of Respiratory and Critical Care Medicine*, *182*(6), 819–825.

Zavodny, M. (2000). The effect of the minimum wage on employment and hours. *Labour Economics*, *7*(6), 729–750.

Zhang, C., Liu, Q., Ge, G., Hao, Y., & Hao, H. (2021). The impact of government intervention on corporate environmental performance: Evidence from china's national civilized city award. *Finance Research Letters*, *39*, 101624.

Zhang, G. (2023). Regulatory-driven corporate greenwashing: Evidence from "low-carbon city" pilot policy in china. *Pacific-Basin Finance Journal*, *78*, 101951.

Zhang, Y., Wang, P., & Kwon, J. (2021). Csr in china: Does being close to the central or local government matter? *Sustainability*, *13*(16), 8770.

Zhang, Y., & Zhao, Z. (2022). Environmental regulations and corporate social responsibility: Evidence from china's real-time air quality monitoring policy. *Finance Research Letters*, *48*, 102973.

Zheng, B., Yu, C., Lin, L., Du, H., Lv, J., Guo, Y., ... others (2017). Associations of domain-specific physical activities with insomnia symptoms among 0.5 million chinese adults. *Journal of sleep research*, *26*(3), 330–337.

Zheng, G., Li, K., & Wang, Y. (2019). The effects of high-temperature weather on human sleep quality and appetite. *International Journal of Environmental Research and Public Health*, *16*(2), 270.

Zhou, P., Ma, J., Li, X., Zhao, Y., Yu, K., Su, R., ... Wang, G. (2023). The long-term and short-term effects of ambient air pollutants on sleep characteristics in the chinese population: big data analysis from real world by sleep records of consumer wearable devices. *BMC medicine*, *21*(1), 1–16.