Population Density and PM2.5 Pollution: Evidence from China

Kaijun Nong[a*], Pengfei Liu[b], Lei Zhu[c]

School of Economics and Management, Beihang University, 37 Xueyuan Road, Beijing 100191, China

[a*]nkj314383788@163.com, [b]pengfei_liu@uri.edu, [c]leizhu@buaa.edu.cn

Abstract. Air pollution, especially fine particulate matter (PM2.5), is a major global environmental and public health concern. In China's rapid urbanization, high population concentration intensifies pollution sources and worsens air quality. This study uses high-resolution remote sensing data and national census data from 2000, 2010, and 2020, with county-level units as the basis for analyzing the relationship between population density and PM2.5 pollution.

Baseline OLS regression shows a significant elasticity of 0.087 after controlling for province-by-year fixed effects. To address potential endogeneity, slope and seismic risk index are introduced as exogenous instruments. IV results show that, after controlling for province-by-year fixed effects, the pollution elasticity of population density increases to 0.223.

Further heterogeneity analysis reveals that the impact is more pronounced in economically underdeveloped regions, suggesting weaker pollution control capacity and stronger marginal effects of population agglomeration. These findings highlight the importance of considering regional economic disparities when designing air pollution control policies to achieve more equitable and effective environmental outcomes.

Keywords: Air pollution, Population density, Environmental protection.

1 Introduction

Air pollution, especially fine particulate matter (PM2.5), has become a major global concern in both environmental and public health domains. As a key indicator of air quality, PM2.5 not only causes significant harm to the human respiratory and cardiovascular systems, but is also closely associated with the incidence of various chronic diseases and rising mortality rates. Numerous studies have shown that long-term exposure to high concentrations of PM2.5 significantly increases the risk of developing illnesses such as lung cancer, asthma, and heart disease, and may even lead to premature death [1,2].

Units with larger population sizes are more likely to experience higher levels of pollution [3]. As the world's second most populous country, China is currently undergoing

N. B. Abdul Ghani et al. (eds.), *Proceedings of the 2025 6th International Conference on Big Data and Social Sciences (ICBDSS 2025)*, Advances in Computer Science Research 126,
https://doi.org/10.2991/978-94-6239-598-5_24

a critical phase of rapid urbanization alongside accelerating population aging. According to the 2020 National Bulletin on the Development of Aging Services, by the end of 2020, the proportion of the population aged 60 and above had reached 18.70%, with those aged 65 and above accounting for 13.50%. Meanwhile, data from the National Bureau of Statistics show that China's urbanization rate rose steadily from 36.22% in 2000 to 63.9% in 2020. The urbanization process has led to a high concentration of population in cities, significantly increasing urban population density. This trend has further driven the rapid expansion of industrial production, transportation, and construction activities—sectors that are typically accompanied by substantial PM2.5 emissions—thereby exacerbating air pollution problems.

In addition, meteorological conditions such as precipitation and wind speed, as well as residents' lifestyles, also play important roles in the dispersion of particulate matter, further complicating the relationship between population density and pollution exposure [4,5,6]. Against this backdrop, the impact of urban scale and population density on air quality has become a prominent research topic in environmental economics, urban planning, and related disciplines. A growing body of literature has demonstrated a significant positive correlation between population size and particulate pollution [7,8], and further confirmed the positive association between population density and PM2.5 concentration [9,10].

To gain a deeper understanding of how population agglomeration affects environmental pollution, and to provide empirical evidence for the formulation of effective pollution control policies, this study conducts a quantitative analysis of the relationship between population density and PM2.5 pollution. The analysis is based on high-resolution remote sensing data at the micro level, combined with data from the 2000, 2010, and 2020 national population censuses, using county-level administrative units as the basic unit of analysis. The results show that, after controlling for province-by-year fixed effects, the estimated pollution elasticity of population density is 0.223. This implies that a 1% increase in population density in a given region is associated with an approximate 0.223% increase in PM2.5 concentration. Further heterogeneity analysis reveals that this effect is more pronounced in economically underdeveloped areas, suggesting that weak infrastructure and limited environmental governance capacity may amplify the marginal impact of population agglomeration on air pollution.

2 Data

The population data used in this study are sourced from the national population censuses organized by the Chinese government. These datasets comprehensively record multidimensional information such as population size, spatial distribution, and age structure across different regions of China. They are highly authoritative and representative, providing a solid foundation for analyzing the relationship between population and environmental factors at the regional level. The PM2.5 pollution data are obtained from the National Tibetan Plateau Data Center and are formatted as gridded data with a spatial resolution of 1 kilometer. These data effectively capture pollution distribution at the micro scale and are well-suited for spatial matching with variables such as population density.

To enhance the robustness and explanatory power of the regression model, this study introduces multiple control variables aimed at minimizing the interference of potential confounding factors in the analysis of the main variable relationships. Among them, the Normalized Difference Vegetation Index (NDVI), which measures vegetation coverage and growth vitality, is derived from NASA's MODIS dataset, specifically the MOD13A3 product. It is also gridded data with a spatial resolution of approximately 1 kilometer. Nighttime light data, which reflect surface illumination intensity during nighttime and are commonly used as proxies for economic activity and industrial emissions, are sourced from publicly available datasets developed by Zuoqi Chen and collaborators [11], with a resolution of about 0.5 kilometers. Additionally, wind speed data are obtained from the National Centers for Environmental Information (NCEI) under the National Oceanic and Atmospheric Administration (NOAA), while precipitation data are primarily sourced from the ERA5-Land dataset jointly released by the European Union and the European Centre for Medium-Range Weather Forecasts (ECMWF). Within the scope of this study, NDVI serves as an indicator of natural absorption capacity, nighttime light intensity broadly reflects industrial structure and energy consumption, and wind speed and precipitation represent key meteorological conditions. These control variables span both natural and anthropogenic dimensions, enabling a more comprehensive identification of the true relationship between population density and PM2.5 pollution.

Considering that ordinary least squares (OLS) estimation may be affected by endogeneity issues, this study further adopts an instrumental variable (IV) strategy to improve identification accuracy. Specifically, two geological factors—slope and seismic risk index—are selected as exogenous instruments for population density.

Variation in slope within a region not only affects land use and urban expansion but also significantly influences population distribution decisions. Steep terrain typically entails higher construction costs and limited accessibility, leading to lower population density. While some studies suggest that topographic features may influence PM2.5 concentrations, such effects are generally indirect—operating through the blocking role of mountains on wind speed, which in turn affects pollutant dispersion [12]. These mechanisms do not imply a direct impact of slope on PM2.5 emissions or sources. In this study, wind speed is explicitly controlled as a key meteorological variable, which helps mitigate the indirect influence of topography on pollution dispersion. This enhances the validity of the exclusion restriction assumption and supports the use of slope as an instrumental variable for population density.

Slope data are derived from the General Bathymetric Chart of the Oceans (GEBCO) terrestrial topographic model, with a spatial precision of approximately 0.5 kilometers. Using ArcGIS geographic information system software, the study processes digital elevation model (DEM) data by calculating elevation differences between adjacent pixels to obtain local slope values and generate high-resolution slope raster maps.

The seismic risk index is constructed based on the global earthquake event dataset provided by the United States Geological Survey (USGS). To comprehensively capture the potential impact of seismic activity on population distribution, the study first defines the research area based on China's geographic coordinates and then extends it outward by 150 kilometers to form a buffer zone that includes adjacent regions. This buffer design accounts for the indirect influence of nearby seismic events on population distribution within China, thereby improving data completeness and spatial coverage. Within the constructed research area, kernel density analysis is applied to calculate the spatial concentration of earthquake events for each year, using earthquake magnitude as a reference indicator to quantify seismic activity into gridded data. The resulting seismic risk raster map features high spatial resolution and explanatory power, serving as a valid instrumental variable for population density. It satisfies both relevance and exogeneity requirements, providing strong support for subsequent empirical analysis.

3 Method

3.1 Baseline Model

In this study, population density is defined as the ratio of the total population within an administrative region to its land area. Data at this level are highly representative and provide a comprehensive reflection of the socioeconomic conditions and environmental characteristics within each region, thereby enhancing the explanatory power and policy relevance of the empirical analysis.

To ensure accurate identification of the impact of population density on PM2.5 pollution, this study employs a fixed effects regression model. By introducing regional fixed effects, the model effectively controls for relatively invariant characteristics within regions—such as topography, historical urban layout, and infrastructure conditions—that may interfere with pollution dispersion. Additionally, year fixed effects are included to eliminate the influence of time-varying macro policies, economic fluctuations, and other systemic environmental factors, thereby improving the rigor and credibility of the regression results.

The baseline regression model used in this study is specified as follows:

$$\ln\left(\mathrm{PM}_{2.5_{it}}\right) = \omega \ln\left(\mathrm{Density}_{it}\right) + \gamma X_{it} + \alpha_c + \psi_t + \phi_t \eta_c + \epsilon_{it} \tag{1}$$

where i and t represent location and year, respectively; $\mathrm{PM2.5}_{it}$ denotes the PM2.5 concentration in location i during year t; $\mathrm{Density}_{it}$ represents the population density of that location in the same year; ω is the key coefficient of interest, measuring the elasticity of pollution concentration with respect to population density; X_{it} is a vector of control variables including precipitation, wind speed, nighttime light intensity, and NDVI; γ is the vector of coefficients associated with these control variables; α_c denotes regional fixed effects; ψ_t denotes year fixed effects; and $\phi_t \times \eta_c$ represents region-by-year fixed effects.

It is important to note that the set of fixed effects adopted may vary across different model specifications. In particular, when region-by-year fixed effects are introduced, some fixed effect terms may be automatically dropped due to multicollinearity among variables. All regression results are clustered at the county level to enhance the robustness of the estimates. In the baseline regression analysis, this study reports results both with and without fixed effects to more comprehensively identify the role of time-invariant confounding factors across different temporal and spatial dimensions.

3.2 IV Model

In empirical analysis, traditional OLS regression may lead to underestimation of coefficients[13]. To address the challenges brought by potential endogeneity, this study adopts the instrumental variable (IV) strategy. This approach introduces exogenous variables that are highly correlated with population density but do not directly affect pollution levels, thereby mitigating estimation bias caused by endogeneity. Specifically, two key geological factors—slope and seismic risk index—are selected as instruments for population density. These geographic characteristics influence land suitability for construction, thereby spatially constraining the expansion of urban built-up areas and indirectly affecting population density distribution. However, aside from their impact on population density, these geological factors do not directly influence PM2.5 concentrations, thus theoretically satisfying both the relevance and exogeneity requirements of valid instruments.

In terms of model specification, this study employs the standard two-stage least squares (2SLS) estimation method, with the model equations defined as follows:

$$\ln\left(\text{Density}_{it}\right) = \delta Z_{it} + \gamma_z X_{it} + \alpha_c + \psi_t + \phi_t \eta_c + \mu_{it} \tag{2}$$

$$\ln\left(\text{PM}_{2.5_{it}}\right) = \omega \ln\left(\widehat{\text{Density}}_{it}\right) + \gamma X_{it} + \alpha_c + \psi_t + \phi_t \eta_c + \epsilon_{it} \tag{3}$$

In these equations, Z_{it} represents the vector of exogenous instrumental variables; δ is the coefficient estimated in the first stage for the instruments; and ω is the key coefficient estimated in the second stage, measuring the elasticity of PM2.5 pollution with respect to population density. All other variables are defined consistently with the baseline regression model. Similar to the baseline specification, this study reports IV estimation results under various model settings, including specifications with and without fixed effects. All regressions are clustered at the county level to ensure robustness and statistical significance of the estimates.

4 Result

4.1 OLS

The baseline effect of population density on air pollution is examined using ordinary least squares (OLS) regression, with detailed results presented in Table 1. Across different model specifications, the OLS estimates consistently show that the coefficient for population density is positive and statistically significant at the 1% level, indicating a clear upward trend in PM2.5 concentration as population agglomeration increases.

This finding suggests that, regardless of whether various fixed effects—including province-by-year fixed effects—are introduced, or whether additional control variables are included, the positive impact of population density on air pollution remains robust and significant.

Table 1. OLS regression results.

	ln(PM2.5)			
ln(density)	0.151*** (0.002)	0.131*** (0.004)	0.102*** (0.004)	0.0871*** (0.004)
night-light	0.0223*** (0.001)	0.0114*** (0.001)	0.0221*** (0.001)	0.00884*** (0.001)
NDVI	0.413*** (0.028)	-0.0699* (0.036)	-0.727*** (0.031)	-0.233*** (0.036)
rain	92.55*** (2.392)	-97.24*** (4.146)	-21.69*** (5.378)	-33.96*** (5.387)
wind	0.0410*** (0.002)	0.0289*** (0.002)	0.0398*** (0.002)	-0.0165*** (0.002)
constant	3.807*** (0.027)	3.574*** (0.033)	4.052*** (0.038)	3.616*** (0.043)
Fix-effect				
Year		Yes		Yes
Province			Yes	Yes
Province-by-year				Yes
Observations	8133	8133	8133	8133

As shown in Figure 1, under the model specification that includes province-by-year fixed effects, the estimated pollution elasticity of population density is 0.087. This implies that a 1% increase in population density in a given region leads to an approximate 0.087% increase in PM2.5 concentration. This result aligns with the environmental pressures commonly observed in urban planning and economic activities, and provides strong empirical support and quantitative evidence for understanding the long-term impact of population agglomeration on air quality during the urbanization process.

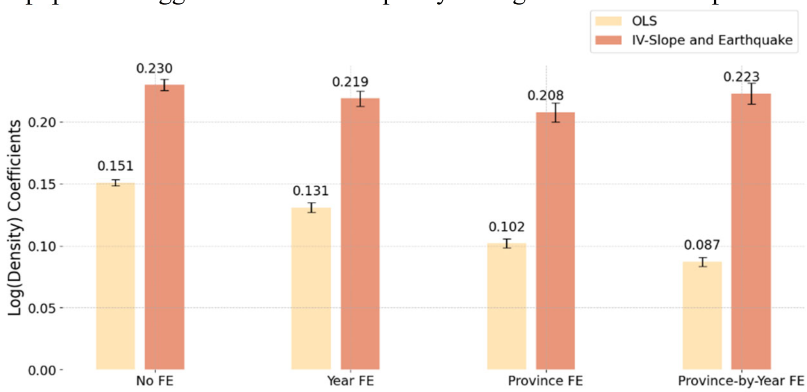

Fig. 1. The regression results of OLS and IV.

4.2 IV

To address potential endogeneity issues and ensure the reliability of the estimation results, this study further adopts an instrumental variable (IV) strategy to overcome empirical identification challenges. Two geological factors—slope and seismic risk index—are introduced as exogenous instruments for population density. Table 2 presents the detailed results of the IV regression. The IV estimates show that the coefficient for population density remains positive and statistically significant at the 1% level, further confirming the positive effect of population density on air pollution.

Table 3 reports the first-stage regression results of the IV strategy. Both slope and seismic risk index pass the weak identification and under-identification tests and exhibit strong statistical significance in the first-stage regression, indicating that they effectively explain the variation in population density as an endogenous variable and satisfy the relevance requirement for valid instruments.

Figure 1 also presents the regression results using slope and seismic risk index as instrumental variables. Compared to the baseline OLS model, the IV-estimated coefficient increases significantly. In particular, after introducing province-by-year fixed effects, the pollution elasticity of population density rises to 0.223, indicating that the OLS model underestimated the true impact. After controlling for endogenous biases, the true impact of population aggregation on air pollution became more significant. Taking the PM2.5 level of a typical city as an example, an elasticity of 0.087 indicates that for every 1% increase in population density, the PM2.5 concentration will rise by approximately 0.087%. If the annual average PM2.5 concentration in a certain city is 50 μg/m^3, the corresponding pollution increment is approximately 0.044 μg/m^3. In contrast, when the elasticity is 0.223, the pollution increment is approximately 0.112 μg/m^3. In high-density urban areas, its cumulative effect may bring significant environmental and health risks. This study modified the internal variables through the IV method, more accurately revealing the strong correlation between population concentration and air pollution. This finding emphasizes the importance of identification strategies in environmental impact assessment and also suggests that high-density urban areas may face more severe pollution pressure, requiring enhanced spatial planning and emission control. It provides convincing empirical evidence for understanding the trade-off between population dynamics and environmental quality in the context of urban expansion, and offers important insights for designing differentiated pollution control and urban planning policies.

Table 2. IV regression results.

	ln(PM2.5)			
ln(density)	0.230***	0.219***	0.208***	0.223***
	(0.004)	(0.006)	(0.008)	(0.008)
night-light	0.0386***	0.0303***	0.0372***	0.031***
	(0.001)	(0.002)	(0.002)	(0.002)
NDVI	0.791***	0.527***	0.682***	0.288***
	(0.031)	(0.055)	(0.052)	(0.073)
rain	95.91***	99.07***	-12.74*	-16.61**

	ln(PM2.5)			
	(2.565)	(5.503)	(6.570)	(7.428)
wind	0.0541***	0.0446***	0.0438***	0.026***
	(0.002)	(0.002)	(0.002)	(0.003)
constant	3.776***			
	(0.026)			
Fix-effect				
Year		Yes		Yes
Province			Yes	Yes
Province-by-year				Yes
Observations	8359	8359	8359	8359

Table 3. IV First-Stage regression results.

	ln(density)			
Slope	-0.245 ***	-0.246 ***	-0.182 ***	-0.185 ***
	(0.006)	(0.005875)	(0.008)	(0.008)
Seismic risk index	0.016 *	0.034 ***	0.046 ***	0.070 ***
	(0.008)	(0.008)	(0.009)	(0.010)
Fix-effect				
Year		Yes		Yes
Province			Yes	Yes
Province-by-year				Yes
Observations	8359	8359	8359	8359

5 Heterogeneity Analysis

The formation of PM2.5 is influenced not only by natural conditions but also closely linked to anthropogenic activities such as urban transportation and industrial production. In urban settings, these activities typically rely on a relatively developed economic foundation. However, economic development varies significantly across regions, reflected in disparities in economic levels, fiscal resources, and infrastructure capacity. To further investigate how economic heterogeneity affects the relationship between population density and pollution, this study conducts subgroup analysis based on regional GDP levels and urbanization rates. GDP data are sourced from the China County Statistical Yearbook, while urbanization rate data are obtained from the Seventh National Population Census.

Table 4 presents the detailed results of the heterogeneity. This study collected GDP data of each county from 2000 to 2020, calculated the average GDP of each county, and then divided the data samples into high and low groups based on the median GDP level and urbanization rate respectively. Interaction terms between population density and these grouping variables are constructed and incorporated into the OLS and IV regression models. After controlling for province-by-year fixed effects, the results show that the estimated coefficient for population density in high-GDP regions is 0.239, which is significantly lower than that in low-GDP regions. Similarly, in the urbanization rate grouping, the pollution elasticity of population density in highly urbanized areas is 0.222, clearly lower than the 0.244 elasticity observed in less urbanized regions. All results are statistically significant at the 1% level, indicating that in economically underdeveloped areas with relatively lagging infrastructure, the capacity for pollution control is weaker, and the increase in population density exerts a stronger influence on PM2.5 concentration.

Table 4. The result of Heterogeneity Analysis

	A.GDP		B.urbanization rate	
	OLS	IV	OLS	IV
	ln(PM2.5)			
High-GDP	0.0988*** (0.00395)	0.239*** (0.0106)		
Low-GDP	0.0944*** (0.00433)	0.268*** (0.0130)		
High-urbanization			0.0940*** (0.004)	0.222*** (0.008)
Low-urbanization			0.103*** (0.004)	0.244*** (0.010)
Provinve-by-year	Yes	Yes	Yes	Yes
Observations	6921	6921	7899	7899

This finding reveals an important linkage between economic development and environmental governance capacity. It underscores the need to account for regional economic heterogeneity when formulating pollution control policies, and advocates for allocating more resources to environmentally vulnerable regions in order to achieve more equitable and efficient improvements in air quality.

6 Conclusion

Air pollution has become a widely discussed environmental issue in recent years and is a key factor affecting future sustainable development. This study focuses on the relationship between population exposure and PM2.5 concentration in China, systematically evaluating the elasticity of pollution exposure with respect to population density under different model specifications.

After controlling for province-by-year fixed effects, the OLS regression results indicate that the estimated elasticity of population density on PM2.5 concentration is 0.087, meaning that a 1% increase in population density leads to an approximate 0.087% increase in PM2.5 concentration. To improve the accuracy of identifying the effect of population density on pollution exposure, this study introduces two geological variables—slope and seismic risk index—as exogenous instruments and applies an IV strategy to mitigate endogeneity in the model. The IV regression results show that the pollution elasticity of population density rises significantly to 0.223 and remains robust at the 1% significance level, implying that a 1% increase in population density results in a 0.223% increase in PM2.5 concentration. This suggests that the OLS model may underestimate the true impact of population density on pollution when endogeneity is not addressed.

To further explore the underlying mechanisms, this study conducts a heterogeneity analysis based on economic development levels. The results reveal that economic factors significantly influence the elasticity of population density on PM2.5 pollution, with the effect being more pronounced in economically underdeveloped regions. This phenomenon highlights the close interaction between economic development and environmental governance capacity, emphasizing the necessity of implementing differentiated policy interventions under varying economic contexts to achieve more targeted and effective pollution control outcomes.

Disclosure of Interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Zang, S.T., Wu, Q.J., Li, X.Y., et al.: Long-term PM2.5 exposure and various health outcomes: An umbrella review of systematic reviews and meta-analyses of observational studies. Sci. Total Environ. 812, 152381 (2022)
2. Ni, R., Su, H., Burnett, R.T., et al.: Long-term exposure to PM2.5 has significant adverse effects on childhood and adult asthma: A global meta-analysis and health impact assessment. One Earth 7(11), 1953–1969 (2024)
3. Borck, R., Schrauth, P.: Urban pollution: A global perspective. J. Environ. Econ. Manag. 126, 103013 (2024)
4. Wang, Q., Kwan, M.P., Zhou, K., et al.: The impacts of urbanization on fine particulate matter (PM2.5) concentrations: Empirical evidence from 135 countries worldwide. Environ. Pollut. 247, 989–998 (2019)
5. Jing, Z., Liu, P., Tuanhui, W., et al.: Effects of meteorological factors and anthropogenic precursors on PM2.5 concentrations in cities in China. Sustainability 12(9), 3550 (2020)
6. Kim, H., Huh, J.Y., Na, G., et al.: Lifestyle practices that reduce seasonal PM2.5 exposure and their impact on COPD. Sci. Rep. 13(1), 11822 (2023)
7. Glaeser, E.L.: Are cities dying? J. Econ. Perspect. 12(2), 139–160 (1998)
8. Borck, R., Schrauth, P.: Population density and urban air quality. Reg. Sci. Urban Econ. 86, 103596 (2021)
9. Ahlfeldt, G.M., Pietrostefani, E.: The economic effects of density: A synthesis. J. Urban Econ. 111, 93–107 (2019)
10. Carozzi, F., Roth, S.: Dirty density: Air quality and the density of American cities. J. Environ. Econ. Manag. 118, 102767 (2023)
11. Chen, Z., Yu, B., Yang, C., et al.: An extended time-series (2000–2023) of global NPP-VIIRS-like nighttime light data. Harvard Dataverse (2020)

12. Wen, Y., Xiao, J., Yang, J., Cai, S., Liang, M., Zhou, P.: Quantitatively disentangling the geographical impacts of topography on PM2.5 pollution in China. Remote Sens. 14(24), 6309 (2022)
13. Heblich, S., Trew, A., Zylberberg, Y.: East-side story: Historical pollution and persistent neighborhood sorting. J. Polit. Econ. 129(5), 1508–1552 (2021)