Energy Policy 135 (2019) 110998

Contents lists available at ScienceDirect

Energy Policy

journal homepage: www.elsevier.com/locate/enpol

# Urban public transport and air quality: Empirical study of China cities

Chuanwang Sun[a,b], Wenyue Zhang[b,*], Xingming Fang[c,**], Xiang Gao[d], Meilian Xu[e]

[a] School of Economics & Management, Nanchang University, Nanchang, 330031, China
[b] China Center for Energy Economics Research, School of Economics, Xiamen University, Xiamen, 361005, China
[c] Institute of National Economics, School of Economics, Southwestern University of Finance and Economics, Chengdu, 611130, China
[d] School of Business, Shanghai University of International Business and Economics, Shanghai, 201620, China
[e] Department of International Economics and Business, School of Economics, Xiamen University, Xiamen, 361005, China

ARTICLE INFO

*Keywords:*
Air quality index
Urban public transport
Urban built-up area
2SLS
Air pollution

ABSTRACT

To analyze the impact of the increase of public transport on the urban air quality will contribute to the sustainable development of urbanization. But many existing studies have not paid attention to the potential endogeneity of estimation, which comes from the fact that the deterioration of air quality would in turn affect the policies of public transport investment. This paper attempts to control this endogeneity by introducing an instrument variable of the urban built-up area into the empirical models. Using city-level data from China, our study adopts 2SLS method and conducts a series of robustness tests to ensure the estimation results more convincing and robust. The results show that the urban air quality could be improved if the city provides more buses for public transport. Moreover, after controlling the endogeneity, the marginal improving effect of increasing the public transport on urban air quality could be larger from 0.082 to 0.678. This finding indicates that the endogeneity bias is likely to cause the underestimation of the improving effect, and may result in some errors of the policy decisions of urban investment.

## 1. Introduction

Air pollution has become an urgent problem in many countries who are undergoing a rapid development of industrialization and urbanization (Ghose et al., 2004; Richardson, 2005; Lefèvre, 2009), and is particularly severe in China (Xu and Masui, 2009). The air pollution jeopardizes public health and aggravates living surroundings (Colvile et al., 2001; Querol et al., 2001). As shown in Fig. 1, the national average air quality index (AQI) showed a downward trend before 2013, meaning the improvement of air quality. However, after changing the evaluation criteria in 2013,[1] AQI has gone from a spiral decline to a continuous increase of not less than 5% for three consecutive years. In particular, in the first year after the change of AQI index standard, its growth rate even reached 45.14%. What's more, according to "Chinese Environmental Situation Communique in 2016", urban AQI in 254 cities exceeded the standard with a high rate of 75.1%. More seriously, the average number of days with mild or even heavy pollution accounted for 25.8% in the whole year.

A dramatic increase in the on-road traffic volume induces vehicular pollution and tremendous air pollution costs (Zegas, 1998). As pointed out by Ghose et al. (2004) and Mao et al. (2017), the transport sector is the largest contributor to man-made pollutant emissions in the urban environment. The proportion of emissions from motor vehicles to total emissions is growing rapidly in urban areas (Frost et al., 1996; Colvile et al., 2001; Pan et al., 2016). For example, in China's mega cities like Beijing and Guangzhou, about 80% of CO emissions and 40% of NOx emissions come from automobiles (Giovanis, 2018). Fig. 2 shows the NOx emissions over the past five years nationwide, and the proportion of NOx emissions from vehicle exhaust to total has increased year by year. Therefore, it is very important to face traffic-related air pollution.

* Corresponding author.
** Corresponding author.
*E-mail addresses:* cw_sun@foxmail.com (C. Sun), zhangwenyue1996@qq.com (W. Zhang), xingmingf@163.com (X. Fang), gaoxiang19901005@126.com (X. Gao), mlxu@xmu.edu.cn (M. Xu).

[1] The newly revised Air Quality Standard (GB3095-2012) in 2013 stipulates that Air Pollution Index (API) is replaced by Air Quality Index (AQI), and three items of PM2.5, $O_3$ and CO are added to the evaluation of pollutants on the basis of API, which makes evaluation results more objective.

https://doi.org/10.1016/j.enpol.2019.110998
Received 3 May 2019; Received in revised form 21 August 2019; Accepted 14 September 2019
Available online 20 September 2019
0301-4215/ 

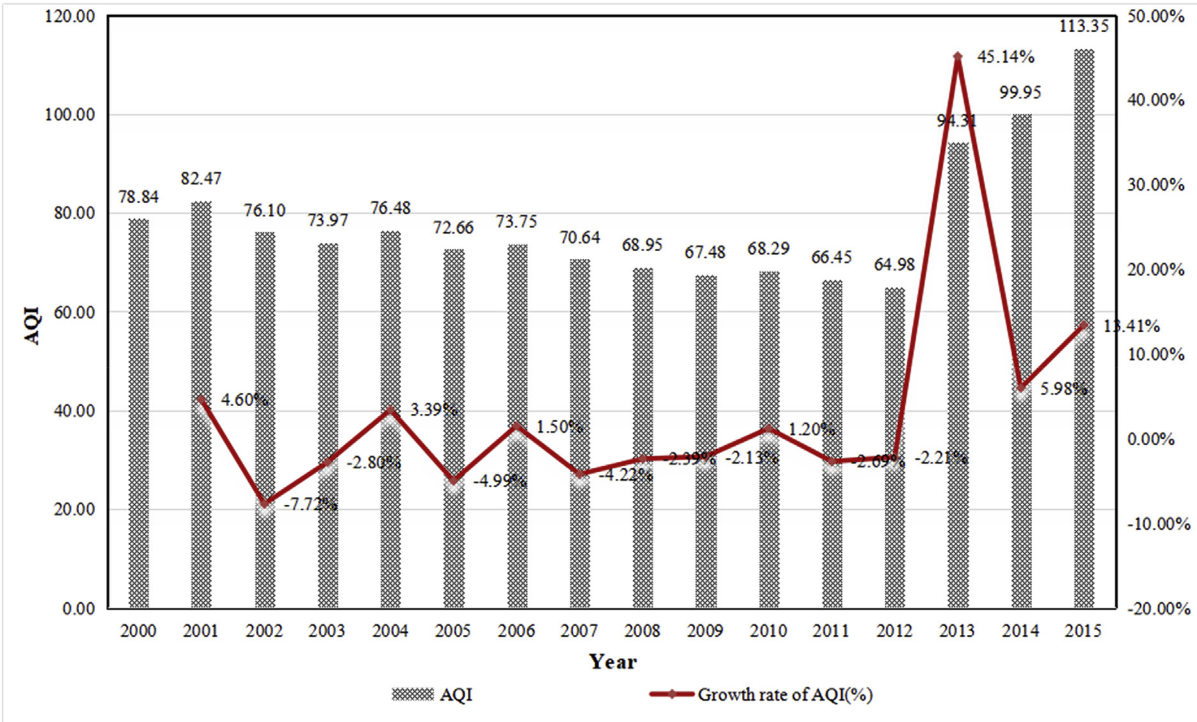


**Fig. 1. AQI in China from 2000 to 2015**
*Source:* The data of AQI are from Environmental Protection Department (EPD) and Growth rate of AQI is calculated by the authors.

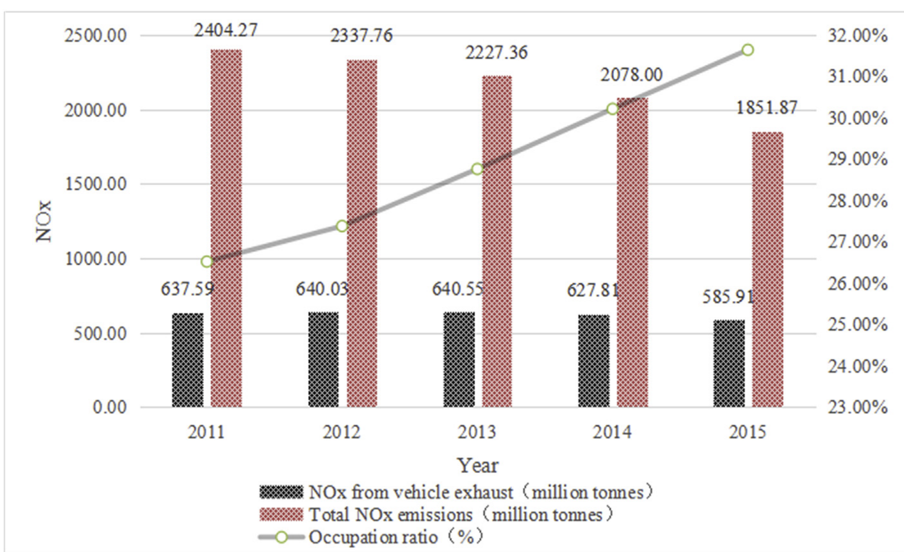


**Fig. 2. NOx emissions in China from 2011 to 2015**
*Source:* The data of NOx are from China Environment Production Database and Occupation ratio of NOx from vehicle exhaust to total is calculated by the authors.

Since the report of the 18th National Congress of Chinese government proposed to build a “beautiful China”, improving air quality has become the focus of Chinese government. The state VI emission standard[2] for passenger car air pollution and the new Environmental Protection Law have been implemented successively in China.[3] In addition, the China's government attempt to control transport-related emissions by establishing and expanding public transport infrastructure, which is considered as a common strategy to ease road congestion and contribute to the sustainable development of urbanization (Lozano et al., 2014; Bel and Holst, 2018).

Regarding the impacts of public transport on air quality, the researchers have done some empirical studies on introducing new transit forms like metro or Bus Rapid Transit (BRT). Chen and Whalley (2012) discover that the opening of the metro reduces CO emission by 5%–15%

[2] The state VI emission standard is the fourth national emission standard for motor vehicle pollutants. The main pollutants emitted by automobiles are HC (hydrocarbons), NOx (nitrogen oxides), CO (carbon monoxide), PM (particulate matter) and so on, which are applied by better catalytic converter active layer, secondary air injection and exhaust recirculation system with cooling device.

[3] The revised environmental protection law came into force on January 1, 2015.

based on Taipei city daily data. An analogous case is provided by Messa, who also points out that strengthening the construction of rail transit has significant and robust pollution control effects. Zheng and Kahn (2013) further verifies that rail transit is an alternative to the original high energy consumption and high pollution travel mode, and produces a certain scale effect to achieve the effect of emission reduction and pollution control. Another research of Bel and Holst (2018) implies that after using BRT, the concentration of all pollutants in Mexico City decreases significantly except for $SO_2$. Kumar et al. (2011) also points out that BRT is more efficient than cars and motorbikes. When travelers choose to ride BRT, the emissions of CO and HC are reduced by 4% and 9% compared with those of cars and motors (Kumar et al., 2011). Furthermore, Reddy et al. (2000) hold the view that bus's emission per passenger-km is much lower than those of other vehicles.

For most cities in the developing countries which are suffering serious air pollution, the impact of bus transportation on air quality should be examined systematically and rigorously. However, in the previous studies it seems that less attention has been paid to this issue in China. Therefore, as a supplementary study of the existing literature, which usually focus on the impact of traffic policy and the effect of the new constructed public transportation on urban air quality, this paper analyzes the impact of the increase of public bus transport on the urban air quality. The differences from the previous research are as follows. Firstly, the existing studies mostly aim to research the impacts on air pollution in specific cities, which are unable to provide the enough evidence on the national public traffic development during the whole country's long run urbanization process. In this paper, we use the panel data of 63 cities in China from 2004 to 2015 to explore the effect of public transportation development. Secondly, many existing studies seldom discuss the potential endogeneity of estimation, which comes from the fact that the deterioration of air quality would in turn affect the policies of public transport investment. This paper attempts to control this endogeneity by introducing the urban built-up area as an instrument variable into the empirical models. In addition, a number of robustness checks are conducted to support the persuasive analyses. Our empirical results could also provide some new policy implement on other developing countries.

The rest of the paper is organized as follows. Section 2 presents the methodology. Section 3 provides data followed by empirical results and further discussion in section 4. Section 5 is conclusions and policy suggestions.

## 2. The methodology

### 2.1. Fixed effect model

To analyze the impact of public transit on urban air pollution which is based on panel data, we firstly establish the relationship between urban air quality and public transportation which is represented by the number of buses. The sample cities have individual heterogeneity in terms of development level, scale level, and geospatial space (Luo et al., 2018; Sun et al., 2018). Therefore, this paper will use the fixed effect model.

Our dependent variable is in-consistent with the study of Sun et al. (2018), who use AQI as an environmental variable to study the substitution and improvement effects of urban rail construction on atmospheric pollutants. Except for the explanatory variable of the number of buses, we further control other factors that affect air quality, such as greening degree of urban built area, bus carrying capacity and climatic factors. Considering the different conditions in each city, there may be missing variables that do not change with time (Sahai, 2010). We conduct the following individual fixed effect model for empirical research:

$$logAQI_{i,t} = \beta_1 logBN_{i,t} + \beta_2 X_{i,t} + \beta_3 D_{i,t} + \mu_i + \varepsilon_{i,t}$$
$$i = 1,2,3...63\ ;\ t = 2004,2005,2006,...,2015 \quad (1)$$

where $i$ indexes city and $t$ indexes year. $AQI$, as the dependent variable, denotes the air quality level. $BN$ reflects the number of bus vehicle. And we take the logarithm of $AQI(logAQI)$ and $BN(logBN)$. $X$ indicates a vector of control variables, consisting of per capita green area, the average annual passenger volume of each bus and climate factors such as temperature, humidity and wind speed. $D$ represents time dummy variables.[4] $\mu_i$ is the unobserved random variable which represents the intercept term of the individual heterogeneity. $\varepsilon_{i,t}$ is the i.d.d disturbance term. $\beta_1$, $\beta_2$ and $\beta_3$ are coefficients.

### 2.2. The endogeneity of the model

Static panel data model often has endogeneity, which mainly comes from interaction between explanatory variables and dependent variables (Anselin and Lozano-Gracia, 2009). If the evidence shows that the positive impact of bus transportation on air pollution control is reliable, it is reasonable to believe that the serious air pollution could reversely affect the policy maker's decisions of the bus facility development. In other words, government regulation of air pollution and the allocation of public transport may have obvious correlations. Therefore, the bus transportation is likely not an absolute exogenous variable, and the reverse causality could generate the endogeneity between these both variables of bus transportation and air pollution.

Specifically, the changes in urban public transport may affect the atmospheric pollution, which is the main issue of our study. In addition, the air pollution level could possibly affect the government decision of investments in transport infrastructure to release the traffic congestion and emission (Sun et al., 2018). This inverse causality may generate the endogeneity and lead to the estimation bias (Zaman and Moemen, 2017). Thus, whether we can identify a variable which has bearings on public transport but not relative to missing variables, is the key to distinguishing the correlation between urban public transport and air quality.

### 2.3. The selection of instrument variable

In order to solve the endogenous problem, this paper introduces a suitable instrumental variable (IV) in the empirical model. Appropriate instrumental variables must be highly correlated with endogenous variables and ensure the condition of exogeneity (Lu et al., 2017; Zugravusoilita et al., 2008). We consider that the urban built-up area with heavy transport pressure, is an important indicator that affects the public transport provided by government in general. The urban built-up area largely determines the allocation and cost of urban public transportation, so there is a high correlation between urban built-up area and public transport.[5] In this context, the area of urban built-up areas will directly affect the amount of urban public transport, and thus the correlation between the instrument variables and the endogenous variable is guaranteed. Furthermore, taking the exogeneity of the instrumental variable into account, the size of urban built-up areas ($logUrbanarea$) is linked with vehicle stock which induces changes in urban air quality, indicating that the instrument variable may be not simply connected with the endogenous variable. For the sake of the exclusiveness of instrumental variables, we incorporate the indicator of per capita green space area ($logGRE$) as a control variable in the measurement model to better control and observe the impact of urban greening degree on the dependent variable.

[4] This time dummy variable takes 2013 as the time node. The dummy value is 0 before 2013 while the virtual value after 2013 (including 2013) is 1.

[5] This paper is based on the results of F-test, identification test and weak instrumental variable test to determine whether the instrumental variable satisfy the correlation with the endogenous variable.

**Table 1**
The interpretation of variables.

| Variable | Variable name |
|---|---|
| *logAQI* | Log of air quality index |
| *logBN* | Log of the number of bus vehicle |
| ln*Car* | Natural logarithm of vehicle ownership |
| ln*GDP* | Natural logarithm of per capita Gross Domestic Product |
| (ln*GDP*)2 | The square of ln*GDP* |
| ln*Second* | Natural logarithm of industrial structure index (%) |
| ln*PD* | Natural logarithm of population density |
| *logPassenger* | Log of the average annual capacity of each bus |
| *logGRE* | Log of Per capita green area |
| *Temp* | Temperature (°C) |
| *Wind* | Wind speed (m/s) |
| *Humi* | Humidity (%) |
| *logUrbanarea* | Log of urban built-up area |
| *logInduswater* | Log of industrial waste water emissions |
| *logBP* | Log of the number of buses per ten thousand people |
| *logIndusgas* | Log of industrial emissions |
| *logRoad* | Log of per capita road area |

Note: The reciprocal first to fourth variables (logInduswater, logBP, logIndusgas and logRoad) will be applied to the robustness test, which will be described in the analysis of robustness test results.

### 2.4. Fixed effect model with 2SLS method

Our model aims at conquering the endogeneity of the explanatory variable. We establish the fixed effect model using the two stages least square method (2SLS) to estimate the parameters (Giovanis, 2018). In the first stage, we regress the endogenous explanatory variable the number of buses (denoted as $logBN_{i,\quad t}$) on the instruments (denoted as $Z_{i,\quad t}$) and other exogenous regressors and save the predicted value $log\hat{B}N_{i,\quad t}$. That is, the sample projection or the predicted value of $logBN_{i,\quad t}$ on $Z_{i,\quad t}$ and other control variables is

$$\log \hat{B}Ni,\ t = \hat{\delta}Z_{i,t} + \hat{\theta}X_{i,t}$$
$$i = 1,2,3...63\ ;\ t = 2004,2005,2006,...,2015 \tag{2}$$

where ˆ represents the fitting value of the variable or the estimated value of the variable coefficient, the $Xi,\ t$ vector is the same control variable as in the ordinary individual fixed effect model, and $Zi,\ t$ is the vector formed by the instrumental variable. $Di,\ t$ is the dummy variable whose value is 0 before 2013 while the virtual value after 2013 (including 2013) is 1. $\theta$ and $\delta$ are the coefficients of the control variables and the instrumental variables. The description of all variables is detailed in Table 1.

In the second stage, we regress the dependent variable (denoted as $logAQI_{i,t}$) onto the predicted value $log\hat{B}N_{i,\quad t}$ and all the regressors, which can be represented as follows:

$$logAQI_{i,t} = \beta^{*}\log\hat{B}Ni,\ t + \beta_2 X_{i,t} + \beta_3 D_{i,t} + \hat{\mu}i,\ t + \varepsilon_{i,t}$$
$$i = 1,2,3...63\ ;\ t = 2004,2005,2006,...,2015 \tag{3}$$

By the regression of two stages, we can get the consistent estimate of $\beta^{*}$. Parameter $\beta^{*}$ can mitigate the endogenous bias and reflect the causality between the explanatory variable number of bus vehicles and the dependent variable air quality index.

### 2.5. The dynamic panel data model

*AQI* is the index of the air quality for a particular period of time. However, it is not only in relation to the current air pollution, but also affected by air pollution of the past, which demonstrates air pollution may have a time lag effect. This paper uses the dynamic panel model of Arellano-Bond estimation proposed by Arellano and Bond (1991) to identify the hysteresis effect of air quality and examine the robustness of *logBN*. Therefore, it is appropriate to let *AQI* ($\log AQI_{t-1}$) lagged by one year separately to test the time-lagged effects. This dynamic model is used for model setting error test in robustness test.

By establishing the dynamic panel data model, we use Generalized Method of Moments (GMM) to make the difference, and eliminate the individual effect and all the non time-varying interpretations. The orthogonality between the error term and the lagged value of the explanatory variable is utilized in the moment condition (Sun et al., 2018). Generally, dynamic panel data model with difference GMM has the following advantages. (1) It includes lagged terms of explanatory variables or dependent variables, which is suitable for analyzing individual dynamic behavior (Fattahi, 2015). (2) It can estimate the time-lag effects and solve endogenous problems (Chen and Whalley, 2012; Souza and Gomes, 2015). In order to eliminate the endogeneity of the instrumental variable, the Sargan statistic is used as an observation of the over-identification test. We set the lagged terms ($logAQI_{t-1}$) of *logAQI* as explanatory variables to estimate the time-lag effects and establish the following model:

$$logAQI_{i,t} = \alpha_1 logAQI_{i,t-1} + \beta_1 logBN_{i,t} + \beta_2 X_{i,t} + \beta_3 D_{i,t} + \mu_i + \varepsilon_{i,t}$$
$$i = 1,2,3...63\ ;\ t = 2004,2005,2006,...,2015 \tag{4}$$

## 3. Data

In this paper, the 63 sample cities[6] are from 29 provinces/cities/autonomous regions in China (excluding Hong Kong, Macao, Taiwan, Tibet Autonomous Region and Qinghai Province). Considering the availability of city-level data, annual data are used in our study.

### 3.1. Data of dependent variable

Different from some studies which used the Air Pollution Index (API), we choose the annual mean of the air quality index as the dependent variable.[7] The newly revised Air Quality Standard (GB3095-2012) in 2013 stipulates that API is replaced by *AQI*, and it was no longer published after 2013. In addition, three items of PM2.5, $O_3$ and CO are added to the evaluation of pollutants on the basis of API. Furthermore, the dynamic changes of the data will have a huge impact on the statistical results. As previously argued, taking *AQI* as an explanatory variable can effectively compensate for the deficiency of API.

### 3.2. Data of explanatory variable

The core explanatory variable in this paper is "urban public transport", and the number of urban buses is selected as the proxy variable. The number of buses not only represents the development of urban basic public transport, but also reflects the satisfaction of local governments' public services (Gómez-Perales et al., 2007). Badami and Haider (2007) indicate that increased buses can meet the rapid growth of mass transit demand as well as reduce the individual motor vehicle activity at low cost.

### 3.3. Data of control variables

Referring to the existing research, we find that other variables also

[6] The newly revised Ambient Air Quality Standard (GB3095-2012) stipulates that the API Index will be replaced by the AQI index in 2013.Since January 1, 2013, a total of 74 key environmental protection cities nationwide have begun to announce AQI. Among the 74 cities that issued air quality index, 11 cities such as Lasa and Xining had a large number of missing data on explanatory variables, so only 63 samples were retained in the final sample cities. According to the National Economic and Social Survey Bureau's 2015 National Economic and Social Development Statistical Communique, the population of sample cities accounted for 26.56% of the total population in the country, which has a strong representation.

[7] The AQI published by the environmental protection department is daily data, but the empirical analysis uses annual data as the research object, so AQI will be treated annually.

affecting urban air quality such as urban greening rate, traffic sharing rate and natural climate factors (Wehner and Wiedensohler, 2003; Luo et al., 2018). In order to control the impact of these factors, we use "per capita green area" (*GRE*), "annual passenger volume per bus" (*Passenger*), "temperature" and "wind speed" respectively. *GRE* represents the level of urban greening, reflecting the overall environmental status of the region (Saito et al., 1991). If only the number of buses increase, various of problems such as high empty rate or less passenger volume may occur, which will lead to the decline in the capacity of traffic pressure sharing and cause air pollution. Thus, *Passenger* is selected as the control variable in this paper. *Passenger* is equal to the total passenger volume divided by the number of buses. In addition, meteorological factors, such as wind direction, wind speed, temperature and relative humidity, are the main reasons for the formation, accumulation and diffusion of gaseous pollutants. Accordingly, the annual average temperature (*Temp*), humidity (*Humi*) and mean wind velocity (*Wind*) are adopted to control the effects of different natural climatic factors on air quality (Leightnerab, 2008).

### 3.4. Data source

This paper uses the unbalanced panel data of 63 cities from 2004 to 2015, with a total sample size of 756. The AQI comes from the monitoring data of Environmental Protection Department (EPD). The number of buses, the number of buses per thousand people, the per capita green area and the area of the urban built-up area and total passenger volume[8] are all derived from the China Economy Information NET (CEINET). Natural climate data such as daily temperature, wind speed and humidity come from China Meteorological Science Data Sharing Service Network. Tables 1 and 2 present the interpretation and summary statistics for variables respectively.

## 4. Empirical results

In this section, we first present results based on individual-specific effects model, then explore the impact of increases in the number of buses on urban air quality examined by the 2SLS method and finally conduct placebo tests and robustness checks to strengthen the reliability of our findings.

### 4.1. Regression analysis of individual fixed effect model

Table 3 presents the empirical results based on the individual fixed effect model which is used to estimate individual heterogeneity (Sahai, 2010). First, as expected, *logBN* is found to be negatively correlated with *logAQI* after controlling the urban effect. The empirical analysis shows that the increases in the number of bus will lead to beneficial environmental effects. According to the results of column (4) with complete control variables, for every 1% increase in bus amount, the level of air quality index will drop by 0.082%. In other words, increasing the urban public transport will improve urban air quality. As suggested by Zheng and Kahn (2013), the more residents travel by public transport, the less traffic congestion and the better air quality would be. A recent study in Barcelona also indicates that NOx and BC shows higher levels during the period of bus drivers' strikes (an increase of 4.1% and 7.7%), which highlights the essentiality of buses in mitigating the concentration of urban air pollution (Basagaña et al., 2018). But the increasing number of bus delivery may have negative effects inversely. For example, traffic congestion is common in rush hour and the lack of bus lanes in some cities can cause inconvenience to buses (Gómez-Perales et al., 2007). In terms of time and efficiency, residents generally do not choose to take bus but other means of vehicles, which reduces the load rate and traffic sharing rate of certain bus route. While the final result proves that the positive influence is greater than the negative one. Furthermore, as shown in column (1)–(4), the coefficients of *logBN* gradually increase with the addition of control variables, meaning that there is a high correlation between the number of buses and the control variables, and the independence of buses to air pollution.

The results of the other control variables in Table 3 also conform to the expectation. The per capita green area has a significant negative impact on the air quality index. The increase in the annual passenger transport volume of each bus may reflect the increased traffic sharing rate and frequency of the public use of the bus, which improves the urban air quality.

**Table 2**
Summary statistics for variables.

| Variable | Observation number | Mean | Standard error | Minimum value | Maximum value |
|---|---|---|---|---|---|
| *AQI* | 756 | 68.55 | 14.58 | 30.24 | 128.96 |
| *BN* | 756 | 3674.76 | 4698.72 | 90 | 31716 |
| *Passenger* | 756 | 17.82 | 14.35 | 0.97 | 244.28 |
| *GRE* | 756 | 39.87 | 20.38 | 9.39 | 164.58 |
| *Temp* | 756 | 15.83 | 4.62 | 2.18 | 25.58 |
| *Wind* | 756 | 2.34 | 1.28 | 1.05 | 33.89 |
| *Humi* | 756 | 67.72 | 9.51 | 34.07 | 86.09 |
| *Urbanarea* | 756 | 272.66 | 256.17 | 34 | 1563 |
| *Induswater* | 756 | 13172.86 | 14529.58 | 4.08 | 86496 |
| *BP* | 756 | 11.83 | 11.56 | 0.94 | 115 |
| *Indusgas* | 756 | 2558.74 | 2414.48 | 9 | 15161 |
| *Road* | 756 | 13.19 | 6.92 | 0.31 | 64 |

**Table 3**
Regression results of individual fixed effect model.

| Dependent variable:log*AQI* | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| log*BN* | −0.103*** | −0.0752*** | −0.0851*** | −0.0822*** |
| | (-6.96) | (-4.84) | (-5.13) | (-4.88) |
| log*GRE* | – | −0.0785*** | −0.0720*** | −0.0690*** |
| | | (-4.96) | (-4.43) | (-4.17) |
| log*Passenger* | – | – | −0.0228 | −0.0224 |
| | | | (-1.70) | (-1.66) |
| *Temp* | – | – | – | 0.0041 |
| | | | | (0.92) |
| *Wind* | – | – | – | −0.0027 |
| | | | | (-0.87) |
| *Humi* | – | – | – | −0.000197 |
| | | | | (-0.22) |
| *Constant* | 4.945*** | 5.025*** | 5.140*** | 5.059*** |
| | (42.46) | (43.45) | (38.38) | (27.02) |
| *Time dummy variable* | YES | YES | YES | YES |
| *Individual fixed effect* | YES | YES | YES | YES |
| *N* | 756 | 756 | 756 | 756 |
| $R^2$ | 0.2062 | 0.2335 | 0.2367 | 0.2385 |
| *F-test* | 47.00 | 46.30 | 46.22 | 25.32 |
| | [0.00] | [0.00] | [0.00] | [0.00] |
| *Hausman Test* | 37.73 | 27.63 | 31.15 | 80.97 |
| | [0.00] | [0.00] | [0.00] | [0.00] |

Note: t statistics in parentheses * p < 0.1; **p < 0.05; ***p < 0.01.

### 4.2. Regression analysis of the instrumental variable

Considering that the regression results of the fixed effect model may have endogenous bias (Sun et al., 2018), the number of buses and urban air quality may be mutually causal. We adopt the urban built-up area (*logUrbanarea*) as the instrumental variable of urban public transportation, and use two-stage least squares (2SLS) method based on the fixed effect model. In order to visually perceive the correlation between the instrumental variable selected in this paper and the core

[8] The annual passenger volume of each bus is equal to the total passenger volume divided by the number of buses.

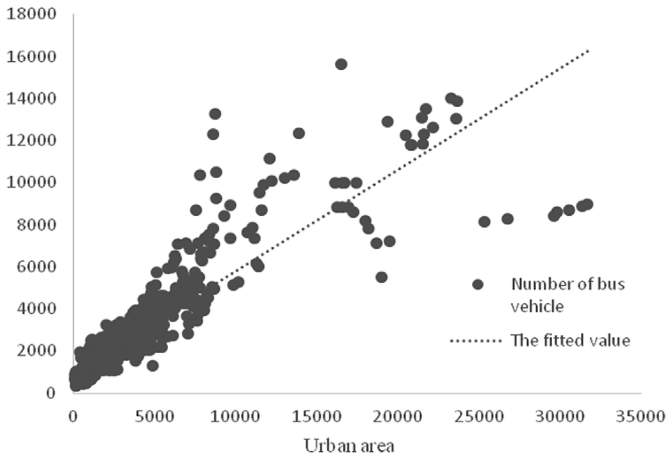


**Fig. 3. Scatter plot of urban built-up area**.

explanatory variable (*logBN*), we perform scatter fitting on the sample data. The scatter plot is shown in Fig. 3 below.

Fig. 3 illustrates that there is a significant positive correlation between the urban built-up area and the number of bus vehicle. This confirms the previous conjecture that urban built-up area where transportation activity is more intensive is an important indicator that affects the public transport provided by government. For the sample cities with different conditions, the urban construction area largely determines the allocation and cost of urban public transportation. The larger the urban built-up area is, the more the amount of buses in urban cities are. It is initially considered that this instrumental variable is reasonable.

But whether the instrumental variable is valid or not requires a series of tests. The last 1–3 rows in Table 4 report the measurements of instrumental variable validity, and the fourth line of countdown shows the regression results of the endogenous explanatory variable (*logBN*) and the instrumental variable (*logUrbanarea*) in the first stage. Firstly, F-test indicates that the hypothesis, the regression coefficients of exogenous instrumental variable and endogenous core explanatory variables are both zero, should be rejected. Moreover, the instrumental variable is all significant at the 1% level in the result of each column, as shown in the last fourth row. Therefore, it is reasonable that the instrumental variable and the endogenous explanatory variable are highly linearly correlated. Secondly, the P value of Anderson Canon. corr. LM statistic obtained from the test for non-identification is far less than 1%. It rejects the null hypothesis that the instrumental variable is not identifiable, and holds that the instrumental variable is related to the endogenous explanatory variable. Thirdly, for weak instrumental variable test, the critical value of Cragg-Donald Wald F statistic should be compared with that of different confidence level intervals. At this point, the values are significantly greater than the critical value at the 10% significance level, so the null hypothesis of the weak instrumental variable is rejected.

In terms of the second stage empirical results of 2SLS method in Table 4, it is clear that the coefficient of *logUrbanarea* is negative and significant under each model, and the result is still stable with the addition of control variables. Besides, it is consistent with the regression results of the fixed effect model which verifies that the increase in the number of buses can significantly reduce the air quality index. However, the absolute values of the coefficient estimates of *logUrbanarea* are significantly greater than the regression results in Table 3. It implies that fixed effect model does not adequately estimate the improvement of urban public traffic on air quality. Without correcting the endogenous bias, the results are possibly underestimated. Furthermore, as previously argued, the coefficients of *logUrbanarea* also gradually increase with the addition of control variables. Moreover, the increase of absolute value of coefficient is much greater than that of fixed effect model without considering the instrumental variable, which further confirms the strong correlation between the number of bus vehicle and

**Table 4**
Regression results of 2SLS model.

| Dependent variable:*logAQI* | | (1) | (2) | (3) | (4) |
|---|---|---|---|---|---|
| *logBN* | | −0.417*** | −0.565** | −0.566*** | −0.678* |
| | | (-5.73) | (-3.26) | (-3.42) | (-2.87) |
| *logGRE* | | – | −0.062 | −0.0864 | −0.0879 |
| | | | (-1.53) | (-1.93) | (-1.94) |
| *logPassenger* | | – | – | −0.128*** | −0.131*** |
| | | | | (-4.34) | (-4.33) |
| *Temp* | | – | – | – | −0.0079*** |
| | | | | | (-1.38) |
| *Wind* | | – | – | – | −0.0003** |
| | | | | | (-0.07) |
| *Humi* | | – | – | – | −0.0004 |
| | | | | | (-0.38) |
| *Constant* | | 6.736*** | 7.024*** | 7.567*** | 7.796*** |
| | | (28.38) | (17.95) | (15.22) | (14.23) |
| *Time dummy variable* | | YES | YES | YES | YES |
| *Individual fixed effect* | | YES | YES | YES | YES |
| *N* | | 756 | 756 | 756 | 756 |
| $R^2$ | | 0.0637 | 0.1244 | 0.1268 | 0.4222 |
| Regression results in the first stage of 2SLS model | | | | | |
| *Endogenous variable* logBN | *Instrumental variable* log*Urbanarea* | 0.0010*** (7.11) | 0.0005*** (3.71) | 0.0005*** (3.96) | 0.0004** (3.16) |
| *F-test* | | 79.34 | 87.32 | 96.73 | 95.46 |
| | | [0.00] | [0.00] | [0.00] | [0.00] |
| *Anderson canon. corr. LM statistic* | | 494.965 | 489.655 | 483.709 | 487.994 |
| | | [0.00] | [0.00] | [0.00] | [0.00] |
| *Cragg-Donald Wald F statistic (10% maximal IV size)* | | 1427.808 (16.38) | 1382.491 (16.38) | 1334.107 (16.38) | 1361.984 (16.38) |

Note: t statistics in parentheses; *p < 0.1; **p < 0.05; ***p < 0.01.

**Table 5**
Placebo test results of 2SLS model.

| Dependent variable:*logInduswater* | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| *logBN* | −0.135<br>(-0.52) | −0.179<br>(-0.35) | −0.179<br>(-0.35) | −0.305<br>(-0.46) |
| *logGRE* | – | 0.0293<br>(0.15) | 0.0373<br>(0.17) | 0.0581<br>(0.24) |
| *logPassenger* | – | – | −0.0627<br>(-0.40) | −0.106<br>(-0.54) |
| *Temp* | – | – | – | −0.0266<br>(-0.75) |
| *Wind* | – | – | – | 0.0373*<br>(2.49) |
| *Humi* | – | – | – | −0.00405<br>(-0.91) |
| *Constant* | 10.08***<br>(4.97) | 10.31**<br>(3.06) | 10.45**<br>(2.82) | 12.10*<br>(2.26) |
| *Time dummy variable* | YES | YES | YES | YES |
| *Individual fixed effect* | YES | YES | YES | YES |
| *N* | 756 | 756 | 756 | 756 |
| $R^2$ | 0.8277 | 0.8303 | 0.8323 | 0.8213 |

Note: t statistics in parentheses; *p < 0.1; **p < 0.05; ***p < 0.01.

**Table 6**
Robustness test results of 2SLS model.

| Dependent variable:*logAQI* | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| *logBN* | −0.678*<br>(-2.87) | −0.877*<br>(-2.28) | – | −0.465***<br>(-4.40) |
| *logGRE* | −0.0879<br>(-1.94) | 0.0736<br>(1.53) | −0.134***<br>(-8.34) | – |
| *logPassenger* | −0.131***<br>(-4.33) | −0.165<br>(-3.72) | 0.0018<br>(0.13) | −0.0983<br>(-4.71) |
| *Temp* | −0.0079***<br>(-1.38) | −0.0053<br>(-0.83) | −0.0002<br>(-0.05) | −0.0081<br>(-1.58) |
| *Wind* | −0.0003**<br>(-0.07) | 0.0007<br>(-0.15) | −0.0029<br>(-0.91) | −0.0009<br>(-0.25) |
| *Humi* | −0.0004<br>(-0.38) | −0.0004<br>(-0.26) | −0.0009<br>(-0.96) | −0.0004<br>(-0.38) |
| *logIndusgas* | – | 0.0895**<br>(2.59) | – | – |
| *logBP* | – | – | −0.0293*<br>(-1.76) | – |
| *logRoad* | – | – | – | 0.0073<br>(0.29) |
| *Constant* | 7.796***<br>(14.23) | 8.251***<br>(10.74) | 7.542***<br>(11.43) | 8.451***<br>(8.66) |
| *Time dummy variable* | YES | YES | YES | YES |
| *Individual fixed effect* | YES | YES | YES | YES |
| *N* | 756 | 756 | 756 | 756 |
| $R^2$ | 0.4346 | 0.4546 | 0.4589 | 0.4006 |

Note: t statistics in parentheses; *p < 0.1; **p < 0.05; ***p < 0.01.

control variables. Secondly, introducing the instrumental variable (*logUrbanarea*) into the model, our findings are reasonable to different city size measures containing urban built-up area. Thirdly, further comparing with the regression results in column (4) in Table 4 and column (4) in Table 3, it is found that the bus annual passenger transport volume, temperature and wind speed are significantly enhanced and consistent with the results of the fixed effect model, which verifies that the instrumental variables better control the omitted variable bias and then increase the robustness of the results.

### 4.3. Further discussion

#### 4.3.1. Placebo test

The above empirical analysis shows that the increases in the number of bus will improve urban air quality. We further take a placebo test to confirm our findings. By choosing a variable that has no direct correlation with the explanatory variable, we use industrial waste water to replace the AQI as the dependent variable. If the coefficient is significant, there is a placebo effect (Luo et al., 2018). Table 5 presents the results of industrial waste water (*logInduswater*) as a dependent variable based on the fixed effect model with 2SLS method. In none of the regressions, the number of bus vehicles is a significant determinant of the pollution. Therefore, our main findings are free from placebo effects.

#### 4.3.2. Robustness test

*4.3.2.1. Omitted variable bias.* Instrumental variables can reduce omitted variable errors while ensuring exclusiveness. Besides, increasing control variables which restricts sample heterogeneity is also a good way to alleviate endogenous bias. Cities with more urban transportation investment could lead to more frequent traffic related to industrial activities, and thus indirectly brought some negative effects on emissions reduction (Duran-Fernandez and Santos, 2014). As previously argued, leaving out industrial pollution in the models may result in omitted variable bias. To address this issue, we introduce industrial waste gas emissions (*logIndusgas*) as an additional control variable. Table 6 tabulates the results. The result in column (2) indicates a positive correlation between industrial emissions and air quality. Further, compared with the result in column (1), the coefficient estimates for *logBN* remain unchanged. Therefore, the effect of bus vehicle on air quality is robust to the inclusion or exclusion of industrial emissions.

*4.3.2.2. Measurement errors.* Measurement errors may occur to the explanatory variable or control variables. So far, we have used *logBN* as the core explanatory variable, following Badami and Haider (2007). However, taking the demographic factor into account, we replace *logBN* by *logBP*, which is defined as the number of buses per ten thousand people, and re-estimate model (3). The result in column (3) of Table 6 shows that *logBP* is negatively correlated with *logAQI*. However, the absolute value of *logBP* is much less than that of *logBN*. This testifies that using *logBP* as core explanatory variable is not as effective as *logBN* on air quality, because excessive number of buses with no corresponding number of passengers would increase exhaust emissions, and then deteriorate air quality.

Regarding the urban traffic construction, the gradual improvement of road infrastructure leads to more alternative channels between the origin and destination (Ahmad et al., 2017). To some extent, increasing road area could ease traffic congestion and have an emission-alleviating effect (Sun et al., 2019). According to the negative correlation between green area and the number of buses in the empirical results, the effect of increasing road area and green area on air quality is considered to be the same. Therefore, we use road area instead of green area to do robustness test. Compared with column (1) in Table 6, the results in column (4) of both variables are negative whereas the correlation of *logRoad* with *logAQI* is insignificant. More importantly, the estimated effects of *logBN* remain the same.

*4.3.2.3. Estimation method bias.* Turning to problem of model setting error, except for setting instrumental variables (IV), we decide to establish dynamic panel data model. The data samples in this paper are typical large N small T panel data, therefore, the differential GMM can be used to overcome the endogeneity in the model.

As shown in Table 7, the coefficient estimates for all variables are basically consistent with the results of fixed effect model, indicating that the results are still robust after using the dynamic panel data model with differential GMM. There is no model setting error in benchmark model. In addition, the coefficients of lagged dependent variables ($logAQI_{t-1}$) are positive and statistically significant at least at 1% level. This suggests that the urban air pollution in the current period will be significantly affected by the past pollution and this effect is continuous and sustainable. Thus, the treatment of urban pollution is a long-term process.

**Table 7**
Robustness test results of dynamic panel data model with GMM method.

| Dependent variable:*logAQI* | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| *logAQI*$_{t-1}$ | 0.359*** | 0.310*** | 0.300*** | 0.294*** |
| | (48.52) | (34.85) | (31.22) | (20.28) |
| *logBN* | −0.121*** | −0.104*** | −0.117*** | −0.117*** |
| | (-22.22) | (-18.60) | (-20.80) | (-16.92) |
| *logGRE* | – | −0.0437*** | −0.0379*** | −0.0433*** |
| | | (-11.58) | (-9.02) | (-6.80) |
| *logPassenger* | – | – | −0.0217*** | −0.0253*** |
| | | | (-15.85) | (-16.64) |
| *Temp* | – | – | – | −0.00499*** |
| | | | | (-3.82) |
| *Wind* | – | – | – | 0.0000825 |
| | | | | (0.63) |
| *Humi* | – | – | – | −0.000780*** |
| | | | | (-4.41) |
| *Constant* | 3.626*** | 3.860*** | 4.041*** | 4.227*** |
| | (59.07) | (51.76) | (51.44) | (50.67) |
| *Time dummy variable* | YES | YES | YES | YES |
| *Individual fixed effect* | YES | YES | YES | YES |
| *N* | 756 | 756 | 756 | 756 |

Note: t statistics in parentheses; *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

## 5. Conclusions and policy implications

Air quality deterioration has become a severe issue which has already jeopardized public health, and the government has rose environmental governance to an unprecedented level. However, with the income increase, the consumption of motor vehicles, which is the largest contributor to anthropogenic pollutant emissions in urban environments, is still growing at a relatively high speed (Yang et al., 2018; Querol et al., 2001). Unfortunately, it's not affordable for most developing counties using alternative fuels or adopting new technologies to alleviate air contamination related to traffic exhaust (Luo et al., 2018). In this context, more public transportation is crucial. Meanwhile, decision makers face trade-offs between the constructions of various public transit that can be manipulated to improve travel condition. Unfortunately, no previous research efforts have been made to explore the roles played by bus vehicle in affecting air quality although few studies examined the impacts of other public transport such as urban metro and BRT.

Based on the panel data of 63 cities in China from 2004 to 2015, this paper uses the instrumental variable (IV) to overcome the endogeneity of *logBN* on the basis of fixed effect model. Four findings are important. First, increasing the urban public transport will lead to beneficial environmental effects. More specifically, for every 1% increase in the amount of bus vehicle, the level of AQI drops by 0.082%. And the coefficients of *logUrbanarea* gradually increase with the addition of control variables in all models, which further confirms the strong correlation between the number of bus vehicle and control variables. Second, after introducing the instrumental variable into the fixed effect model, the absolute values of $logAQI_{t-1}$ estimates are significantly greater than that of fixed effect model which implies that fixed effect model with 2SLS method would adequately estimate the improvement of urban public traffic on air quality. What's more, our findings are reasonable to different city size measures containing urban built-up area. Third, the coefficients of lagged dependent variable ($logAQI_{t-1}$) in the dynamic panel data model are significantly greater than 0.2, suggesting that the air pollution has a lag effect. Fourth and lastly, our results robust to placebo test, omitted variable bias, measurement errors and estimation method bias. All of these results are conducive to the following policy recommendations.

For rapidly urbanized cities confronted with formidable challenge of air pollution owing to vehicle emissions, it is essential to strengthen public bus system construction. The government should make long-term plans to deliver more buses and improve the auxiliary facilities of public transport system such as increasing bus lanes and so on. Our findings appeal to obtain a win-win situation for urban transport development and environmental pollution control when planning Bus Transit system. What's more, using different city size measures containing urban built-up area in China as a case studying, our findings are also robust to different developing countries under consideration between serous air contamination and urbanization. Furthermore, government should be aware of the persistence of air pollution. The introduction of high-pollution projects will have a lasting impact on urban environment. At the time of project approval, attention should be paid to the research on environmental protection.

## Acknowledgements

We would like to express our sincere gratitude to the editor and anonymous referees for their insightful and constructive comments. Especially, we would like to thank the experts who participated in the evaluation and improvement of this manuscript. This work is funded by the National Natural Science Foundation of China (Funding No.71673230 and 71303199), and the Fundamental Research Funds for the Central Universities at Xiamen University (Grand no. 20720191006).

## References

Ahmad, F., Mahmud, S.A., Yousaf, F.Z., 2017. Shortest processing time scheduling to reduce traffic congestion in dense urban areas. IEEE Transactions on Systems Man & Cybernetics Systems 47 (5), 838–855.
Anselin, L., Lozano-Gracia, N., 2009. Errors in Variables and Spatial Effects in Hedonic House Price Models of Ambient Air Quality. Spatial Econometrics. Physica-Verlag HD.
Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev. Econ. Stud. 58, 277–297.
Badami, M.G., Haider, M., 2007. An analysis of public bus transit performance in Indian cities. Transp. Res. A Policy Pract. 41, 961–981.
Basagaña, X., Triguero-Mas, M., Agis, D., 2018. Effect of public transport strikes on air pollution levels in Barcelona(Spain). Sci. Total Environ. 2018 (610), 1076–1082.
Bel, G., Holst, M., 2018. Evaluation of the impact of bus rapid transit on air pollution in Mexico city. Transp. Policy 63, 209–220.
Chen, Y., Whalley, A., 2012. Green infrastructure: the effects of urban rail transit on air quality. Am. Econ. J. Econ. Policy 4, 58–97.
Colvile, R.N., Hutchinson, E.J., Mindell, J.S., 2001. The transport sector as a source of air pollution. Atmos. Environ. 35, 1537–1565.
Duran-Fernandez, R., Santos, G., 2014. A GIS model of the national road network in Mexico[J]. Res. Transp. Econ. 46 (2), 36–54.
Fattahi, M., 2015. The role of urbanization rate in the relationship between air pollution and health expenditures: a dynamic panel data approach. Naunyn-Schmiedeberg's Arch. Pharmacol. 53, 68–72.
Frost, M., Linneker, B., Spence, N., 1996. The spatial externalities of car-based worktravel emissions in greater London, 1981 and 1991. Transp. Policy 3, 187–200.
Ghose, M.K., Paul, R., Banerjee, S.K., 2004. Assessment of the impacts of vehicular emissions on urban air quality and its management in Indian context: the case of Kolkata (Calcutta). Environ. Sci. Policy 7, 345–351.
Giovanis, E., 2018. The relationship between teleworking, traffic and air pollution. Atmos. Pollute. Res. 1, 1–14.
Gómez-Perales, J.E., Colvile, R.N., Fernández-Bremauntz, A.A., et al., 2007. Bus, minibus, metro inter-comparison of commuters' exposure to air pollution in Mexico city. Atmos. Environ. 41, 890–901.
Kumar, R., Gupta, K., Durai, B.K., 2011. Development of Real World Driving Cycles at delhi Bus Rapid Transit Corridor and Air Pollution Scenario.
Lefèvre, B., 2009. Long-term energy consumptions of urban transportation: a prospective simulation of "transport–land uses" policies in Bangalore. Energy Policy 37, 940–953.
Leightnerab, J.E., 2008. Capturing climate's effect on pollution abatement with an improved solution to the omitted variables problem. Eur. J. Oper. Res. 191 (2), 540–557.
Lozano, A., Granados, F., Guzmán, A., 2014. Impacts of modifications on urban road infrastructure and traffic management: a case study. Procedia - Social and Behavioral Sciences 162, 368–377.
Lu, M., Sun, C., Zheng, S., 2017. Congestion and pollution consequences of driving-to-school trips: a case study in Beijing. Transp. Res. D Transp. Environ. 50, 280–291.
Luo, Z., Wan, G.H., Wang, C., Zhang, X., 2018. Urban pollution and road infrastructure: a case study of China. China Econ. Rev. 49, 171–183.
Mao, R., Duan, H., Dong, D., et al., 2017. Quantification of carbon footprint of urban roads via life cycle assessment: case study of a megacity- Shenzhen, China. J. Clean. Prod. 166, 40–48.

Pan, L., Yao, E., Yang, Y., 2016. Impact analysis of traffic-related air pollution based on real-time traffic and basic meteorological information. J. Environ. Manag. 183, 510–520.
Querol, X., Alastuey, A., Rodriguez, S., et al., 2001. Monitoring of PM10 and PM2.5 around primary particulate anthropogenic emission sources. Atmos. Environ. 35, 845–858.
Reddy, A.K.N., Anand, Y.P., D'Sa, A., 2000. Energy for a sustainable road/rail transport system in India. Energy for Sustainable Development 4, 29–44.
Richardson, B.C., 2005. Sustainable transport: analysis frameworks. J. Transp. Geogr. 13, 29–39.
Sahai, H., 2010. Application of some multivariate analysis techniques to data on effect of air-pollution on morbidity. Biom. J. 20, 557–573.
Saito, I., Ishihara, O., Katayama, T., 1991. Study of the effect of green areas on the thermal environment in an urban area. Energy Build. 15, 493–498.
Souza, G., Gomes, E., 2015. Management of agricultural research centers in Brazil: a DEA application using a dynamic GMM approach. Eur. J. Oper. Res. 240 (3), 819–824.
Sun, C., Luo, Y., Li, J., 2018. Urban traffic infrastructure investment and air pollution: evidence from the 83 cities in China. J. Clean. Prod. 172, 488–496.
Sun, C., Zhang, W., Luo, Y., Xu, Y., 2019. The improvement and substitution effect of transportation infrastructure on air quality an empirical evidence from China's rail transit construction. Energy Policy 129, 949–957.
Wehner, B., Wiedensohler, A., 2003. Long term measurements of submicrometer urban aerosols: statistical analysis for correlations with meteorological conditions and trace gases. 3, 1699–1733.
Xu, Y., Masui, T., 2009. Local air pollutant emission reduction and ancillary carbon benefits of SO control policies: application of AIM/CGE model to China. Eur. J. Oper. Res. 198 (1), 315–325.
Yang, M., Ma, T., Sun, C., 2018. Evaluating the impact of urban traffic investment on $SO_2$, emissions in China cities. Energy Policy 113, 20–27.
Zaman, K., Moemen, A.E., 2017. The influence of electricity production, permanent cropland, high technology exports, and health expenditures on air pollution in Latin America and the Caribbean Countries. Renew. Sustain. Energy Rev. 76, 1004–1010.
Zegas, C., 1998. The costs of transportation in Santiago de Chile: analysis and policy implications. Transp. Policy 5, 9–21.
Zheng, S., Kahn, M.E., 2013. Understanding China's urban pollution dynamics. J. Econ. Lit. 51, 731–772.
Zugravusoilita, N., Millock, K., Duchene, G., 2008. The factors behind $CO_2$ emission reduction in transition economies. Social Science Electronic Publishing 75, 461.