

Study on the Influencing Factors of Population Size in China Mega-Cities

Yu-Wei GONG ^a, Yan-Ming LV ^{b,*} and Meng-Xue CHEN ^c

Sichuan Agricultural University, P.R.China

^a2583210946@qq.com, ^blvyanming0515@163.com, ^c18482112368@163.com

* Yan-Ming LYU

Keywords: Urban population size, cost factors, income factors

Abstract. It is of great significance for urban planning and management to identify the influencing factors on population size. We collect the data of resident population and its influencing factors in China four municipalities from 1997 to 2016 to establish a fixed effect model. The urban population size and its influencing factors are analyzed from the angle of cost and income. The results suggest that the wage level, the development of the second and third industries, and the medical level are the main income factors to promote population growth. In addition, the intensity of environmental pollution is the main cost factor to restrain population growth. However, there is a significant positive relationship between the real estate price and the urban population size, perhaps the population size growth causes the real estate prices rise.

1. Introduction

In recent years, the population size of large cities in China has increased rapidly. Such as Beijing, in 1997, the resident population size was 12.4 million, and by 2016, it was 21.73 million, with an average annual increase of 3.02% (Urban Statistical Yearbook of China in 1998 and 2017). Therefore, special concern has been given to the size of urban population, particularly the mega-cities, which is very significant to the development of urban economy as well as the rational allocation of resources and the protection of the environment. Social scholars have made a great deal of research on the scale of urban population. Some studies show that the rapid increase of the urban population led to the rapid increase of urban economy [1,2]. But other studies show that the rapid growth of the urban population will pose new challenges to urban employment, housing, resources and the environment[3,4]. So if the population size is not big enough, it will be not conducive to the urban economic development. However, if the size is excessive, it will lead to a lot of urban problems. Therefore, many scholars proposed the research topic of the optimal urban population size. Yarmohammadian N et al. (2014) studied the optimal and sustainable size of Iran largest cities, and found that Tehran was overpopulated between 1999 and 2012, exceeding 71 percent of its best population size and 5 percent of its sustainable population size; in addition, four major cities, Isfahan, Mashhad, Hiraz and Awaz, have surpassed their optimal size. Zhang Jie et al. (2017) based on economic benefits and energy efficiency, the problem of optimal urban scale in China is studied, they found that the cities with 2 million people have the highest efficiency in terms of economic benefits, social benefits and environmental quality. According to Camagni R et al. (2013), the urban population size is mainly affected by cost and benefit, and there is no exact optimal population size.

Finding out the influencing factors of urban population size is the key to study the optimal urban population scale and to solve a series of problems caused by the excessive urban population size and it is also helpful to urban planning and management[6,7]. Firmino Costa da Silva D et al. (2017) established econometric spatial models to analyze the data pertaining to 3659 Brazilian minimum comparable areas (MCAs) over the period 1970–2010, the results show that the amenities of nearby Mcas and the level of local human capital will affect the population growth of Brazil Mcas. Camagni R et al. (2013) set up a model of equilibrium urban size based on urban cost and urban benefit, then estimate on a sample of 59 European cities with data at FUA level, empirical results show that rent is the highest cost related to the city size, amenities are linked to larger cities, the city benefits from attracting highly educated professionals, and scientific networks can make cities more

balanced. Jeanty P W et al. (2010), considering that there is a spatial interdependence between estate price and population migration, a spatial simultaneous equation model that combines population change and housing value is estimated with data from the Michigan census area. Finally, it is found that if the community population increases, it will improve the housing value, but with the raise of the community housing value, the community is more likely to lose its population under the interaction of space.

The above discussions show that many scholars have examined urban population size and its influencing factors from different aspects, but it seems that considering the population growth from the aspect of cost and benefit is missed, which is important influencing factors that families choose target city to work and live in. For the difficulty of data collection, there are few researches on the relationship between urban population size and influencing factors using long panel data. This paper collected the data of resident population and its influencing factors in China four municipalities from 1997 to 2016 to establish a fixed effect model. The urban population size and its influencing factors are analyzed from the aspect of cost and benefit.

The findings from this study are expected to provide reference for regional government to enact effective measures to control population scale. The remainder of this paper is organized as follows. Section 2 introduces the research data. Section 3 lists the process of econometric model establishment. Section 4 presents econometric model analysis results. Section 5 draws the discussions and conclusions of the study.

2. Research Data

The population of the four municipalities is larger than other cities in China. In 2016, the total resident population of the four municipalities was 95.63 million, including 21.73 million in Beijing, 24.2 million in Shanghai, 15.62 million in Tianjin and 30.48 million in Chongqing. And the total resident population of China was 792.98 million; the total population of the four municipalities exceeds 12 percent of the total urban population of the country.

The four municipalities are typical representatives of Chinese large cities. On the one hand, population agglomeration promotes the rapid growth of urban economy; On the other hand, too large population scale also creates many social problems, such as traffic congestion, excessive consumption of resources, serious environmental pollution and so on. Therefore, this paper chooses the four municipalities to study the relationship between urban population size and its influencing factors. The analysis result can be used as reference for the management of population scale in other cities urbanization process.

It is commonly appreciated that in previous studies population scale topic is a comprehensive concept, as shown in table 1. This measure should contain cost indicators (e.g. environmental pollutants) and income indicators (e.g. education and wage). By referring to previous studies, wage level (Y_1) and employment(Y_2) are used as measures of the incomes of population growth, medical service(Y_3) and education level(Y_4) are considered important to reflect the performance of a city's public service, and population is attracted to quality public services, thus they are selected as another income indicators[11,12,13]. In referring to previous studies, the most typical cost indicator in measuring population scale are refer to environmental pollutants(C_1), and volume of industrial SO_2 is used to measure the level of environmental pollution. Other typical cost indicators include congestion(C_2) and housing rents(C_3),and road area is selected to employ the level of congestion, commercial house price is considered important to reflect the housing rents of a city[7,12,14,16]. The main features of the urban population cost-income indicators are demonstrated in table 2.

Table 1 Summary of typical studies on the influencing factors of population size

Researcher	Study scope	Model	variables
Glaeser E L et al. (2014)	Entire US metropolitan areas	Regression model	Population, median income and self-reported housing values
Au C C, Henderson J V (2006)	205 cities in China	Urban inverted U-shaped model	Population, output of a city, manufacturing to service ratio (MS), employment, capital, output (value-added) of independent accounting units, FDI, roads per capita, high school, distance to coast, distance to provincial capital, highway, area, doctors per capita, books per capita, telephone per 100 persons
Yang Xiaojun (2017)	257 cities in China	GMM	Population, education, cultural, medical, transportation, environmental
Camagni R et al. (2013)	59 European cities	Equilibrium city size model	Population, amenities, diversity, density, rent, malaise, networks, functions, sprawl
Richard Arnott (2006)		optimal city size model	Population, economies of scale in production, congestion
MA Cole&E Neumayer(2004)	54 countries	Econometric model	Population, SO ₂ , CO ₂ , age composition, urbanization rate, average household size

Table 2 Main features of the urban population cost-income indicators for analysis

Dimension	Indicator and definition	Unit
Population	Resident population	10 ⁴ persons
Income indicators	Wage(Y ₁): The average wage of employed staff and workers	10 ⁴ yuan
	Employment(Y ₂): The number of employed persons in second and third industries	10 ⁴ persons
	Medical service(Y ₃): The number of beds in hospitals and health centers	bed
	Education(Y ₄): The number of students in regular institutions of higher education	10 ⁴ persons
Cost indicators	Environmental pollution(C ₁): Industrial sulfur dioxide emissions	ton
	Congestion(C ₂): The reciprocal of per capita road area	m ²
	Rent(C ₃): The sale price of commercial housing	10 ⁴ yuan

The data for analysis about these indicators are collected from four municipalities for the period of 1997–2016, which are available from China Urban Statistical Yearbook, Chinese Statistical Yearbook, Beijing Statistical Yearbook, Tianjin Statistical Yearbook, Shanghai Statistical Yearbook, Chongqing Statistical Yearbook and China Environmental Statistical Yearbook. Meanwhile, in order to get a more intuitive empirical point of view, the paper standardized all the data to obtain the comparability of the data.

3. Establish Econometric Model

3.1 Regression model setting

The aim of this research is to analyze the influencing factors of population size, and a panel data model is used to achieve the aim. It can reflect the change law of both time and cross section at the same time, and reduce the adverse effects caused by multi-collinearity; meanwhile, the variables are subjected to logarithmic processing to reduce the influence of the heteroscedasticity and the linearization of the non-linear relation. The urban population size is taken as the explained variable and the cost-income indicators which influence the urban population size are taken as the explanatory variable, and the model is set as follows:

$$\ln N_{i,t} = B_{i,t} + \alpha_1 \ln Y_{1,i,t} + \alpha_2 \ln Y_{2,i,t} + \alpha_3 \ln Y_{3,i,t} + \alpha_4 \ln Y_{4,i,t} + \beta_1 \ln C_{1,i,t} + \beta_2 \ln C_{2,i,t} + \beta_3 \ln C_{3,i,t} + u_i$$

Where $N_{i,t}$ denotes the size of the population of the i city in t year, and $B_{i,t}$ is a constant term, $Y_{1,i,t}$ is the wage level of the i city in t year, $Y_{2,i,t}$ is the employment scale of the i city in t year, $Y_{3,i,t}$ is the medical care level of the i city in t year, $Y_{4,i,t}$ is the education level of the i city in t year, $C_{1,i,t}$ is the level of environmental pollution of the i city in t year, $C_{2,i,t}$ denotes the urban traffic congestion cost of the i city in t year, and $C_{3,i,t}$ denotes the urban housing cost of the i city in t year. $\alpha_1, \dots, \alpha_4, \beta_1, \dots, \beta_3$ are their correlation coefficients. u_i is the error term. $i=1, 2, 3, 4$ denotes four municipalities, and $t=1997, 1998, \dots, 2016$, represents different years of observation.

3.2 Model testing

The unit root test of panel data is the development of unit root test of time series, which mainly determines the stability of panel data. When applying panel data for analysis, the stability of panel data should be tested first, that is, the unit root test of panel data should be carried out. The main methods are Levin, Lin and Chu test (LLC test, 1992, 1993); Breitung test (2002); Hadri test (2002); Im, Pesaran and Shin test (IPS test, 2003), Fisher- ADF test and Fisher- PP test.

In the above several unit root test methods, we found that the IPS test takes into account the small sample characteristics of the test, and proposes a method to construct the panel unit root test statistics according to the mean value of the individual unit root test statistics. It is considered that the testing ability of IPS is stronger than other methods, and it is also more suitable for our research data.

For non-stationary panel data, panel data co-integration test is to determine whether there is a long-term equilibrium between them. There are two main types of tests: one is the Fisher (combined Johansen) test (MaddalaWu,1999) based on the Johansen cointegration test, the other is the Pedroni Pedroni,1999 and the Kao test (Kao,1999) based on the Engle Granger two-step test, both the Kao test and the Pedroni test are based on the hypothesis of testing the stationarity of the residuals.

3.3 Model selection

Panel data analysis model can be divided into two types: fixed effect model (FE) and random effect model (RE). The difference lies in whether the factors that do not change with time are related to the explanatory variables, and if so, the fixed effect model is chosen. On the contrary, the random effect model is selected. For the choice of the two models, the commonly used detection method is Hausman test. The basic idea is as follows: if the chi-square statistics are significant, the fixed effect model is used instead of the original hypothesis, otherwise the random effect model is adopted.

4. Results

4.1 Panel data unit root test

Direct regression of panel data is likely to lead to pseudo-regression and affect the accuracy of the results. Therefore, the panel data must be tested by unit root test and co-integration test. According to 3.2, we chose IPS test and use STATA12.0 as an analytical tool for testing, as is shown in table 3.

Table 3 Panel data unit root test results

variables	Original sequence			first difference		
	W[t-bar]	P-value	results	W[t-bar]	P-value	results
lnN	3.747	1.000	non-stationary	-1.004	0.058	stationary
lnY ₁	-0.136	0.446	non-stationary	-6.350	0.000	stationary
lnY ₂	3.710	1.000	non-stationary	-1.561	0.059	stationary
lnY ₃	2.013	0.978	non-stationary	-4.538	0.000	stationary
lnY ₄	-0.998	0.159	non-stationary	-6.439	0.000	stationary
lnC ₁	1.453	0.927	non-stationary	-4.137	0.000	stationary
lnC ₂	-0.268	0.394	non-stationary	-4.132	0.000	stationary
lnC ₃	-0.852	0.197	non-stationary	-1.451	0.073	stationary

Table 3 shows that the horizontal values of the variables measured in the panel data are non-stationary, while the first-order difference values are stationary, Therefore, each variable satisfies the integrated of order one I(1) and can be tested by co-integration.

4.2 Co-integration test results

According to 3.2, the Kao test is used for panel co-integration analysis, as is shown in table 4.

Table 4 Co-integration test results

method	Statistic	p-value
Kao test (Kao, 1999)	Modified Dickey-Fuller t	-3.5424
	Dickey-Fuller t	-2.2625
	Augmented Dickey-Fuller t	-1.8404
	Unadjusted modified	-2.3171
	Dickey-Fuller t	-1.9926
	Unadjusted Dickey-Fuller t	0.0232

Table 4 is the result of co-integration test, which significantly rejects the original hypothesis at 5% confidence level, indicating that there is a co-integration relationship among the sequences and they are all balanced data.

4.3 Regression Analysis Result

Select a panel data model by Hausman validation. and found that the Hausman statistic (obeying the chi-square distribution) is 151.55, and the Prob > chi2 is 0.0000, as a result, the original hypothesis should be rejected, that is to say, there are some implicit factors that do not change with time and related to the explanatory variables in the mechanism of determining the resident population size in cities. Further thinking, each city has its own unique history of development, humanistic spirit, geography and geomorphology, these inherent factors can have a lasting and stable effect on the explanatory variables. Therefore, a fixed effect model should be established for regression analysis, as shown in table 5.

Table 5 regression results of urban population size and its influencing factors

variables	Coef.	Std. Err.	T	variables	Coef.	Std. Err.	T
lnY ₁	0.0579* (0.094)	0.0440339	1.31	lnC ₁	-0.0593** (0.043)	0.0286777	-2.07
lnY ₂	0.503*** (0.000)	0.0741305	6.78	lnC ₂	-0.0182 (0.354)	0.0194572	-0.93
lnY ₃	0.316*** (0.000)	0.060658	5.21	lnC ₃	0.0609* (0.093)	0.0389897	1.56
lnY ₄	0.0166 (0.449)	0.0217597	0.76	_cons	3.859*** (0.000)	0.4368717	8.83

Note: P values in parentheses, N=72, R-sq. =0.916, Standard errors in parentheses * p<0.1, ** p<0.05, *** p<0.01;F-test: F (7, 61) = 94.49, Prob > F = 0.0000;

The analysis results show that R-sq = 0.916, F test result Prob > F = 0.0000, which indicate that the combined significance of all the coefficients is higher. In table 5, lnY₁ represents the wage level, lnY₂ represents the level of employment, the result shows that two variables are positively correlated with the urban population size, and passed the test of significance hypothesis of $\alpha = 0.1$ and $\alpha = 0.01$; lnY₃ represents the medical treatment level, and the result shows that there is a positive correlation to the urban population size, and passed the significance hypothesis test of $\alpha = 0.01$; lnY₄ represents the education level, and the correlation is also positive, but not significant; lnC₁ represents the intensity of environmental pollution, the result shows that there is a negative correlation to the urban population size, and passed the test of significance hypothesis of $\alpha = 0.05$; and lnC₂ represents the traffic congestion cost, the correlation is also negative, but not significant; lnC₃ represents housing cost, the result shows that the correlation is positive, and passed the significance hypothesis test of $\alpha = 0.1$; finally, the last line is the constant.

5. Discussion and Conclusion

From the above studies, we find that wage and education have a positive impact on urban population growth, but these effects are not always statistically significant, such as, the education has no significant positive effect on the urban population growth, which is not consistent with the results of Glaeser (2011). Not exactly in line with MA Cole & Neumayer (2004), we find that there is a negative correlation between SO₂ emissions and urban population scale, the correlation is more statistically significant than that of traffic congestion. The influence coefficient of housing cost on population growth is positive, which is similar to the conclusion of Camagni (2013) that rent is the highest cost related to city size, and the impact of housing cost on population growth is higher than that of wage level on population scale growth. Here, we think that the economically developed areas such as Beijing and Shanghai will attract the population, and housing, as a special rigid demand, will increase with the size of population increases. Moreover, a large number of employment opportunities and better medical care and wages are the main reasons for people to choose the four municipalities; the environmental pollution is also the main consideration for people to choose a city. These results are of great significance to the planning and management of the population in four municipalities.

The above results have important policy implication. In the development process of the four municipalities, we should exert the spillover effect of the central city, promote the common development of the surrounding small and medium cities, and narrow the income gap between the large and small cities, and the surrounding small cities should vigorously develop the service industry to increase the employment opportunities. Only by perfecting the medical system, promoting equal access to medical treatment in large and medium-sized cities, improving environmental pollution and providing comfortable living environment, can we retain the population and ease the size of the population in mega-cities and realize the sustainable development of the city.

Limitations of this study are appreciated for further research. Firstly, a more detailed analysis of

the cost and income factors of urban population scale should be analyzed. Secondly, due to the complexity of the spatial distance of the four mega-cities, the influence of space distance on the urban population size has not been considered, which are the further studies in the future.

References

- [1] Sarker S, Khan A, Mannan M M. Urban population and economic growth: South Asia perspective. *European Journal of Government and Economics*, 2016, 5(1): 64-75.
- [2] Frick S A, Rodríguez-Pose A. Change in urban concentration and economic growth. *World Development*, 2018, 105: 156-170.
- [3] M c GUIRK P, Argent N. Population growth and change: implications for Australia's cities and regions. *Geographical Research*, 2011, 49(3): 317-335.
- [4] Hinrichsen D, Robey B. Population and the environment: the global challenge. *Population Reports*, 2000.
- [5] Yarmohammadian N, Akbari N, Asgary A, et al. Optimal and Sustainable City Size by Estimating Surplus Function for Metropolitans of Iran. *International Journal of Business and Development Studies*, 2014, 6(1): 21-38.
- [6] Jie Z, Yang X, Fang Q. Optimal City Size in China: An Extended Empirical Study from the Perspective of Energy Consumption. *China City Planning Review*, 2017(02):24-30.
- [7] Camagni R, Capello R, Caragliu A. One or infinite optimal city sizes? In search of an equilibrium size for cities. *Annals of Regional Science*, 2013, 51(2):309-341.
- [8] Firmino Costa da Silva D, Elhorst J P, Silveira Neto R M. Urban and rural population growth in a spatial panel of municipalities. *Regional Studies*, 2017, 51(6): 894-908.
- [9] Jeanty P W, Partridge M, Irwin E. Estimation of a spatial simultaneous equation model of population migration and housing price dynamics. *Regional Science and Urban Economics*, 2010, 40(5):343-352.
- [10] Hesham Abdel In ahman, Fujita M. Product Variety, Marshallian Externalities, and City Sizes. *Journal of Regional Science*, 2006, 30(2):165-183.
- [11] Glaeser E L, Ponzetto G A M, Tobio K. Cities, skills and regional change. *Regional Studies*, 2014, 48(1): 7-43.
- [12] Au C C, Henderson J V. Are Chinese Cities Too Small? *Review of Economic Studies*, 2006, 73(3):549-576.
- [13] Yang Xiaojun. Impact of urban public service quality on population mobility. *Chinese population Science*, 2017 (02): 106 / 116 130.
- [14] Arnott R. Optimal city size in a spatial economy. *Journal of Urban Economics*, 2006, 6(1):65-89.
- [15] Wang D, Li B L. Study on the Control of Traffic Congestion Based on Hierarchical Road Capacity. *Trans Tech Publications*, 2015, 744: 1862-1865.
- [16] Cole M A, Neumayer E. Examining the impact of demographic factors on air pollution. *Population and Environment*, 2004, 26(1): 5-21.
- [17] Maddala G S, Wu S. A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and statistics*, 1999, 61(S1): 631-652.