Urban Systems: Understanding and Predicting the Spatial Distribution of China's Population

*Pengfei Li, Ming Lu**

Abstract

With urbanization and population migration, some Chinese cities fall into decline whereas others prosper. Using nighttime light data, we redefine the city based on economic function and evaluate the city size distribution in representative countries. The results provide evidence not only for Zipf's law, but also for a distortion in China's current city size distribution. This study proposes a feasible method to predict urban population distribution based on the role of geographical factors in regional development, following the idea of spatial equilibrium. This prediction suggests that the divergence of city size in China tends to be pronounced, with inter-regional income disparity being narrowed and the city size distribution following Zipf's law. The Chinese government should further relax restrictions on population inflow into large cities and prepare for more migration in the future.

Key words: geographical factors, population distribution, urban system, Zipf's law
JEL codes: O21, R12, R23

I. Introduction

Predicting future city size is difficult. In practice, incorrect predictions may cause two directions of problems. One is the insufficient supply of infrastructure and public services, restricting city development. The other is excessive public investment, which places a heavy burden on local public finance. In China, the former problem occurs in large cities with continuous population inflow, and the latter happens in small- and medium-sized cities with population outflow, especially in inland provinces.

The reasons for the above phenomena are the following. First, there is no simple

*Pengfei Li, Post-doctoral Fellow, Antai College of Economics and Management, Shanghai Jiao Tong University, China. Email: lipfchn@163.com; Ming Lu (corresponding author), Professor, Antai College of Economics and Management; China Institute for Urban Governance, Shanghai Jiao Tong University, China. Email: luming1973@sjtu.edu.cn. The authors acknowledge research support from the National Natural Science Foundation of China (Nos. 72073094 and 71834005), Shanghai Institute for National Economy, Shanghai Institute of International Finance and Economics, and the China Merchants Charitable Foundation.

and feasible method of predicting population distribution in economics and geography. Future urban development is difficult to predict, especially in countries experiencing rapid urbanization. Second, the potential population growth in each city is mainly determined by the overall population of a country. However, in practice, city size is usually planned by local government, which ignores inter-relationships across cities. Third, the central government often ignores the agglomeration economies of large cities, and sometimes prefers an even population distribution. This preference makes it difficult for the government to make urban development policies in line with migration trends. Fourth, local governments in cities with population outflow have a strong motivation to attract migrants, whereas governments in cities with population inflow worry about urban problems, such as congestion, which are seemingly related to rapid population growth. In China, urban planning is largely affected by the heritage of a planned economy with strong government intervention, resulting in serious practical problems, such as overbuilding infrastructure in cities with population outflow.

Most existing literature concerning city size distribution provides empirical evidence for whether Zipf's law holds and focuses more on administrative cities (Anderson and Ge, 2005; Kausik and Basu, 2009). However, in China's statistics, there are mainly two levels of administrative cities: four provincial-level municipalities directly under the central government, and 293 prefectural-level cities. A municipality or a prefectural-level city usually consists of an urban district (*shiqu*) and several conuties or county-level cities. There are 394 county-level cities, which are relatively economically independent from the administrative cities that govern the counties, so the administrative cities could not reflect the economically integraded cities. In this paper, a city is redefined based on the economic function of city according to nighttime light data, and the city size distribution is discussed accordingly. A "nighttime light city" (NLC) refers to connected urban areas measured by nighttime light. The boundary of the NLC is determined by economic activities and does not coincide with administrative boundaries. Some large NLCs may refer to metropolitan areas that cover a large municipality/prefecture and the surrounding small- and medium-sized counties. Most of the NLCs correspond to economically independent counties, which are disconnected from the urban districts of the prefectures. Based on this concept, we examine city size distribution in representative countries (including Japan, South Korea, India, Indonesia, and China) and find a distortion in China's city size distribution. Compared with the existing literature, this paper also provides a simple and feasible method to predict future population distribution. First, we establish an econometric model to study the effects of geographical factors on urban economic growth, and then use it to predict the future GDP growth rate of each city and the spatial distribution of economic activities over the country. Then, based on the idea of spatial equilibrium, the spatial distribution of the population is predicted, which is roughly consistent with that of economic activities. This prediction provides a scientific benchmark for both spatial planning of public investment at the country level and the planning of each city.

The rest of this paper is organized as follows. Section II re-evaluates city size distribution in representative countries using nighttime light data. Section III provides the evidence for the distortion of China's city size distribution and we find that the size of the 30 largest metropolitan areas in China has deviated from Zipf's law. Section IV proves the effect of geographical factors on urban economic growth, and forecasts China's urban population distribution in 2035 based on the idea of spatial equilibrium. Section V concludes the paper.

II. Re-evaluating city size distribution and Zipf's law

To understand urban system in a country, we need to re-evaluate the basic rules of city size distribution. In the literature, Zipf's law is widely used for discussions of city size distribution and has been regarded as an important benchmark for urban systems. However, the empirical conclusions based on different data are divergent. In this paper we use nighttime light data to analyze city size distribution, laying a foundation for re-evaluating China's city size distribution.

Zipf's law was first proposed by Zipf (1949). As shown in Equation (1), Zipf's law implies that the population of the first largest city (the primary city) is i times that of the ith largest city:

$$POP_1 = iPOP_i, \tag{1}$$

where POP_1 and POP_i represent the population of the first and ith largest cities. To make an empirical analysis of city size distribution, Equation (1) can be converted into Equation (2):

$$\ln Rank_j = \beta_0 + \beta_1 \ln POP_j + \varepsilon_j, \tag{2}$$

where $\ln Rank_j$ represents the logarithm of size ranking of city j, $\ln POP_j$ represents the logarithm of population size of city j, and ε_j is the random error term. Zipf's law holds for city size distribution when β_1 is equal to –1.

Zipf's law and the power law have attracted most discussion and approval in theoretical and empirical research (Rozenfeld et al., 2011; Huang and Yost-Bremm, 2018; Su, 2020). In fact, Zipf's law is a special case of the power law whose parameter is equal to 1. In Equation (2), the power law holds for city size distribution when β_1 is significantly negative, including the case where β_1 is equal to –1. In reality, β_1 is related to economic, demographic, and geographic factors (Rosen and Resnick, 1980), and is not equal to –1 in the majority of cases. We therefore emphasize the goodness of fit (R^2)[1] for Equation (2) and focus less on whether β_1 is equal to –1 in the following analysis. That is to say, this paper made no effort to distinguish between Zipf's law and the power law.

Some researchers accept Zipf's law (Gabaix, 1999; Rozenfeld et al., 2011), and others are skeptical about it (Eeckhout, 2004, 2009; Bee et al., 2013). The main difference between these two strands of literature lies in sample selection and city definition. Many studies supporting Zipf's law use truncated data and ignore small cities (Kausik and Basu, 2009; Li and Sui, 2013; Arshad et al., 2018), whereas studies using data from all cities found that Zipf's law was not applicable (Eeckhout, 2004; Anderson and Ge, 2005). It was also shown in previous studies that city size measured by urban function followed Zipf's law. For example, Berry and Kozaryn (2012) found that cities measured by economic clusters fitted Zipf's law well. Rozenfeld et al. (2011) also found that city size distribution followed Zipf's law using urban agglomeration data. Jiang and Jia (2011) demonstrated that Zipf's law was applicable when city size was defined by the number of crossroads that are within the same city and are closer than a threshold value. Veneri (2016) suggested that the city size distribution in most countries conformed to Zipf's law using data for urban functional areas. Generally, Zipf's law is more applicable when cities are defined by urban function, which can extend beyond the administrative boundary.

To better understand the rules of city size distribution, we should therefore first choose a proper definition of the city. In existing studies, the definition of the city based on urban function has been widely accepted and nighttime light data have been used to measure economic activity (Henderson et al., 2012). This study uses the data from Jiang (2020) to define the NLC from the perspective of economic function. Specifically, based on the deblurred nighttime light data, Jiang (2020) first delineated human settlement areas based on a luminosity threshold equal to zero, which means that pixels with positive luminosity values were all considered human settlement areas. Second, human settlement areas with short distances between them were aggregated into one polygon. Third, with the Global Rural–Urban Mapping Project (GRUMP) database,[2] 1,527 polygons in 1992 that either cover these GRUMP units or turn to be the most relevant ones near the units were defined as NLCs and named according to their corresponding

[1]The goodness of fit of a regression model describes how well it fits a set of observations, and is expressed as R^2. The range of R^2 is [0, 1]. The larger R^2 is, the better the regression model fits these observations.

[2]This database contains geocoded locations of over 70,000 human settlement areas across the world, their names, populations, and the higher administrative divisions to which they belong (Jiang, 2020).

GRUMP units. Finally, with the gridded population data from the LandScan Global Population Database,[3] the population of an NLC was "the sum of all cells falling within or intersecting with the city contour" (Jiang, 2020, p. 9).[4]

To compare different definitions of the city, we explored city size distribution with both NLC and functional urban area (FUA), which is another definition of city developed by the Organisation for Economic Co-operation and Development (OECD) and the European Union (EU). A FUA consists of one urban center and its commuting zones. Based on gridded population data, each grid cell over the map is assigned a population value, and the urban center is composed of contiguous cells with high population density. The commuting zones are defined as all municipalities with at least 15 percent of their employed residents working in the certain city core (OECD, 2012).[5] We drop the NLCs whose population size is less than 50,000, consistent with the required minimum population size of a FUA. The FUA database covers OECD countries and Columbia, whereas the database of NLC only covers Asian countries. Japan and South Korea are both included in these two databases, so we employ city-level data for these two countries to compare different definitions of a city (FUA and NLC) and study the city size distribution in Figure 1.

Figure 1 shows that the NLC corresponds to a greater number of cities and presents a greater goodness of fit for Equation (2) than the FUA. In 2016, the population of Tokyo reached 13.6 million (administrative city),[6] 35.88 million (FUA), and 54.47 million (NLC) under different definitions. That is to say, Tokyo's economic impacts on surrounding areas are still seriously underestimated even under the definition of FUA. Similarly, the population of Seoul reached 26.49 million (NLC) and 24.05 million (FUA). Functional urban area and NLC have advantages in different aspects. Functional urban area focuses more on the labor market, whereas NLC concerns more on the economic relationship and connectivity across neighboring cities, which is conducive to exploring the leading effects of large cities on the surrounding cities.

---

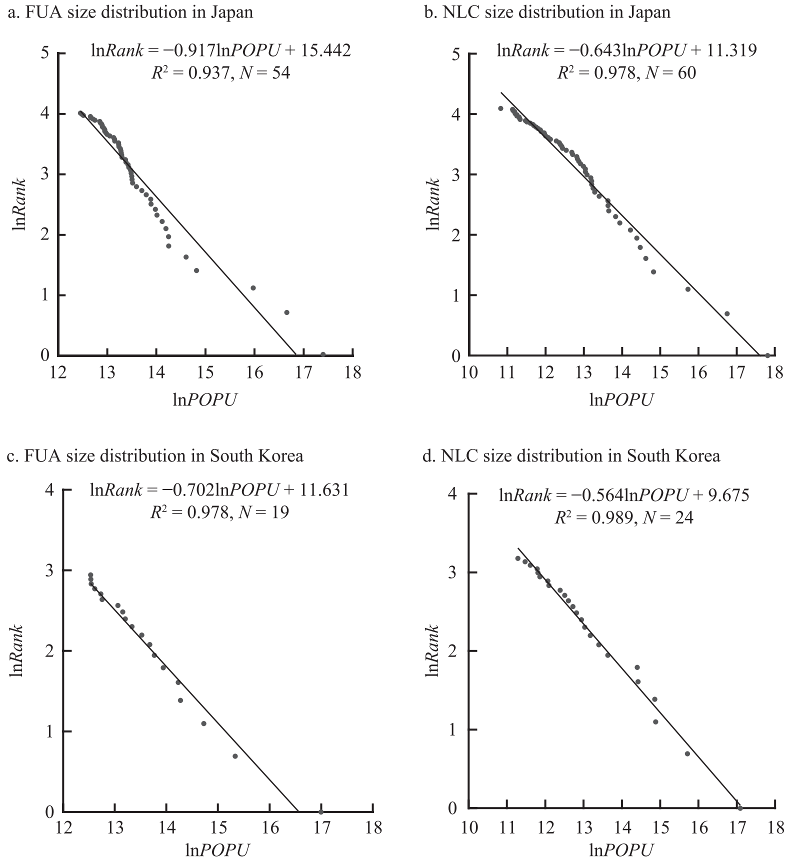
[3]The LandScan Global Population Database is a global gridded population distribution estimation developed by Oak Ridge National Laboratory, in the US. LandScan uses a geographic information system (GIS) and remote-sensing data and technologies with a dasymetric modeling approach to distribute a certain population to cells. For more details, refer to https://landscan.ornl.gov/documentation.

[4]We greatly appreciate the data from Jiang (2020). Nighttime light data start in 1992, and some NLCs had expanded and were adjacent to other NLCs over time. Jiang (2020) divided these NLCs where the luminosity values were the lowest at the neighborhood level. This division has no effect on the conclusions of this paper. More details of this procedure can be obtained from the authors upon request.

[5]For more details regarding FUA, refer to https://stats.oecd.org/OECDStat_Metadata/ShowMetadata.ashx?Dataset=CITIES&ShowO-nWeb=true&Lang=en (online; cited March 2020).

[6]This is an official estimation from the Statistics Bureau of Japan, available from: https://dashboard.e-stat.go.jp/en/dataSearch (online; cited March 2020).

Figure 1. Size distribution of functional urban areas (FUAs) and nighttime light cities (NLCs) in Japan and South Korea in 2016

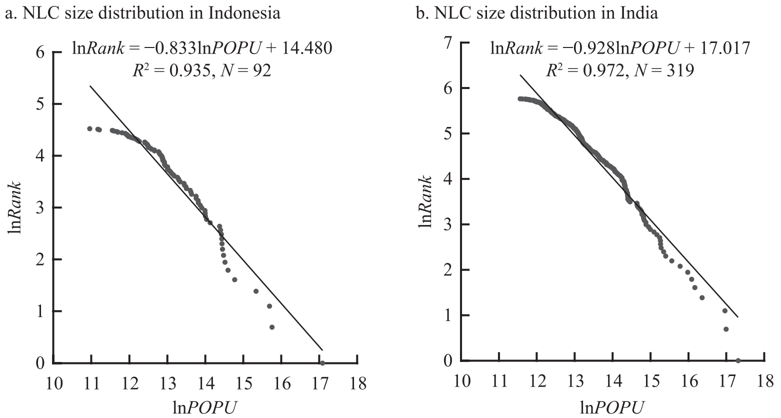


Sources: The data on FUA come from OECD (2012) and the data on NLCs are from Jiang (2020).

Notes: This figure presents the relationship between ln*POPU* and ln*Rank* using scatter diagrams. In panels a and c, ln*POPU* is the logarithm of population size of FUAs. In panels b and d, ln*POPU* is the logarithm of population size of NLCs in Japan and South Korea. In panels a and c, ln*Rank* is the logarithm of the size ranking of FUAs in Japan and South Korea. In panels b and d, ln*Rank* is the logarithm of the size ranking of NLCs in Japan and South Korea. The equation in each panel is the regression result of Equation (2) estimated by the data on FUA (panels a and c) or NLC (panels b and d). Each point represents a FUA (panels a and c) or an NLC (panels b and d) and the solid line represents the fitted values from the regression equation. R^2 indicates the goodness of fit of each regression equation. *N* is the number of points in each panel.

In developing countries, such as India and Indonesia, the goodness of fit of NLC for Equation (2) is still above 93 percent as shown in Figure 2. Owing to some unavailable data, developing countries must adopt an objective indicator to define cities and plan urban development, and the NLC has obvious advantages in this respect.

Figure 2. Nighttime light city (NLC) size distribution in India and Indonesia in 2016



Source: The data on NLC come from Jiang (2020).

Notes: In panels a and b, ln*POPU* is the logarithm of population size of NLCs in India and Indonesia; ln*Rank* is the logarithm of size ranking of NLCs in India and Indonesia. Each point represents a NLC and the solid line represents the fitted values from the regression equation. The equation in each panel is the regression result of Equation (2) estimated by the data on NLC. The meanings of R^2 and *N* are the same as in Figure 1.

III. Distortion of city size distribution in China

In most of the literature, administrative cities in China represent the prefecture-level cities and municipalities directly administrated by the central government. There are three ways to measure Chinese city size. One is the registered population (Anderson and Ge, 2005; Peng, 2010; Wen, 2016; Fang et al., 2017) according to the *China City Statistical Yearbook* (NBS, 1986–2013) or *Fifty Years of Cities in New China* (NBS, 1999). These studies mainly focused on the dynamic changes of city size distribution through long-term data. Wen (2016) found that China's city size distribution followed Zipf's law gradually from 1990 to 2010. Fang et al. (2017) found that the policies of restricting the development of large cities from 1985 to 2000 led to the convergence of city size, but this convergence disappeared after 2000 when restrictions were relaxed. The second measurement of city size is the number of permanent residents according to the Chinese population census data (Kausik and Basu, 2009; Chen and Lu, 2019). Chen and Lu (2019) found that the size of large cities in China was relatively small compared to the fitted line of Zipf's law. The third measurement is based on urban geographical characteristics. For example, Farrell and Nijkamp (2019), based on satellite images and administrative population, found that, from 1982 to 2010, China's city size distribution followed Zipf's law gradually. Long (2016) defined cities in terms of the number of road junctions, and concluded that China's large cities were too small.

Measuring city size using permanent residents has advantages in public service, taxation, and urban governance for local government. However, this definition ignores the interaction across cities. The urbanization of land is also faster than that of the population in China (Lu and Wan, 2014). A definition based on geographical characteristics therefore creates difficulty in measuring the development of Chinese cities through satellite images or number of junctions. In this section, therefore, we evaluate the city size distribution in China by comparing NLCs and administrative cities.

1. Comparison between nighttime light cities and administrative cities in China

Table 1 shows a comparison between NLCs and five administrative cities in China. The differences are obvious. The NLCs whose core city is Guangzhou (NLC-Guangzhou) is the largest NLC in China both by area and population, which includes Foshan, Shenzhen, and other cities (Jiang, 2020). This shows that Guangzhou has built close economic interactions with surrounding cities. The ratio of the size of NLC-Guangzhou area to the size of administrative Guangzhou area (area ratio of Guangzhou) is 2.02, while those of Shanghai and Beijing are only 1.01 and 0.4, respectively. The area ratio of Beijing indicates its limited economic influence on the surrounding areas. In addition, the ratio of the population in the NLC to the population in the core city (population ratio) is significantly higher than the area ratio of representative cities in Table 1, which means that, although the connection between the core city and surrounding areas could be quite low, the NLC covers most of the local population, which agglomerates in the core city.

Table 1. Comparison between nighttime light cities (NLCs) and five selected administrative cities in China in 2016

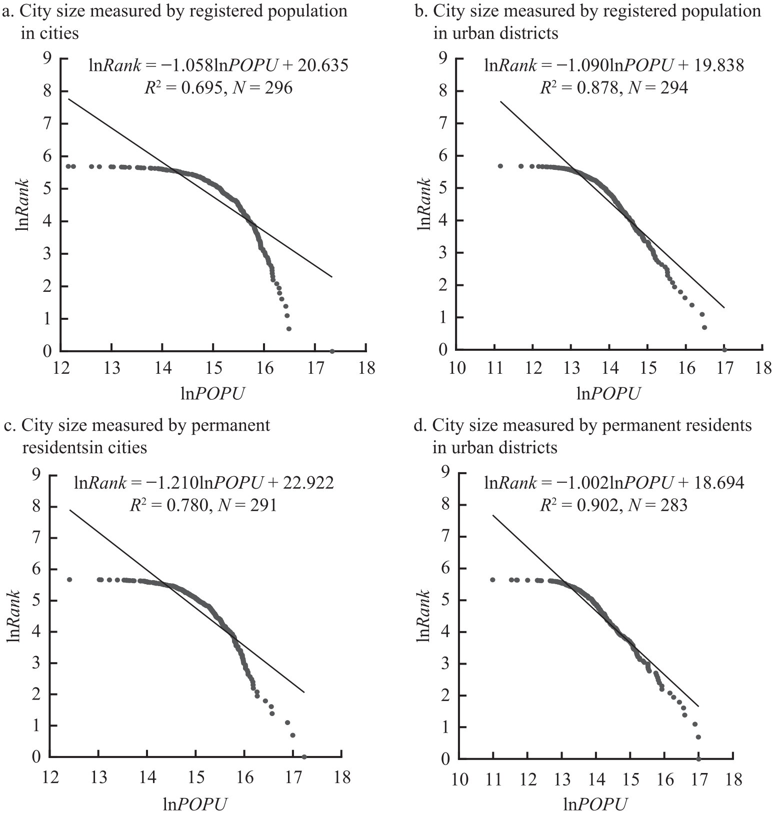
City name	Area (km^2)		Area ratio	Population (million)		Population ratio
	(1)	(2)	(3)	(4)	(5)	(6)
	Administrative	NLC	NLC/Administrative	Administrative	NLC	NLC/Administrative
Guangzhou	7,434	15,030	2.02	13.77	45.52	3.30
Shanghai	6,341	6,410	1.01	24.17	24.14	1.00
Beijing	16,411	6,615	0.40	21.72	19.82	0.91
Lanzhou	13,086	644	0.05	3.70	2.58	0.70
Hegang	14,679	150	0.06	1.04	0.55	0.50

Sources: Administrative city data come from the NBS (2017) and the data on NLC come from Jiang (2020).

Notes: Guangzhou, Shanghai, and Beijing are the core cities of three main metropolitan areas. Lanzhou is a typical city, the development of which is restricted by terrain. Hegang is a famous resource-exhausted city. These five cities represent cities at different stages of development. The population of an administrative city is taken to be the permanent resident population, which is calculated according to GDP and GDP per capita. Columns (2) and (5) indicate the area and population of the corresponding NLCs.

In China, an administrative city usually includes an urban district and several surrounding counties and some other county-level cities. In most cases, the majority of the population and the economic activities of administrative cities are concentrated in the urban district. To further compare the definition of an NLC with an administrative city, we first measue city size, using the permanent residents and registered population of prefectural-level cities and their urban districts, respectively, and study administrative city size distribution. This is shown in Figure 3.

Figure 3. Administrative city size distribution in China in 2016



Source: Data on administrative city come from NBS (2017).

Notes: In panels a, b, c, and d, ln*POPU* represents the logarithm of population size of registered population in cities, registered population in urban districts, permanent residents in cities and permanent residents in urban districts, respectively; ln*Rank* denotes the logarithm of size ranking of registered population in cities, registered population in urban districts, permanent residents in cities, and permanent residents in urban districts, respectively. Each point indicates a city (panels a and c) or urban district (panels b and d) and the solid line represents the fitted values from the regression equation. The equation in each panel is the regression result of Equation (2) estimated by the data on administrative cities (panels a and c) or urban districts (panels b and d). The meanings of R^2 and N are the same as in Figure 1. Permanent residents in each city and urban district are calculated based on the data on GDP and GDP per capita.

In fact, some counties and county-level cities have a close economic connection with urban districts whereas others do not. In China, the location of the economic center is usually close to the seat of local government. However, the average distance between the seat of government of a county or county-level city and the seat of an administrative city government is 77.5 km[7] and most counties and county-level cities have their own industrial zones and economic centers. It is therefore not reasonable to count an urban district and all corresponding counties and county-level cities as one economic city. In Figure 3, the size distribution of administrative cities is also far from Zipf's law. Although the size distribution of urban districts fits Zipf's law well, the permanent residents in all urban districts only account for 43.7 percent[8] of the total urban permanent residents in China, showing that the cities defined by the urban districts seriously underestimate the actual population size of urban systems. On the one hand, some counties and county-level cities should be regarded as single cities considering population size or economic development. On the other hand, urban districts in some administrative cities are closely connected with surrounding areas. Therefore, if one defines a Chinese city by urban district, neither the number of cities nor the size of the urban population can represent the real urban systems of China. It would also be a mistake to conclude that China's city size distribution are already in line with Zipf's law. In short, the relationship between counties or county-level cities and the urban district is complex and has been neglected under the administrative definition.

Figure 4 shows the size distribution of NLCs in China. Not only is the goodness of fit of the NLCs for Zipf's law higher than that of the urban districts but also there is a greater number of NLCs and population size than urban districts. In particular, the number of NLCs with more than 50,000 people reaches 662,[9] showing that the number of China's economic activity agglomeration areas is greater than that of administrative cities. Shanghai is the largest administrative city (NBS, 2017), whereas the NLC around Guangzhou is the largest NLC (Jiang, 2020).

Figures 3 and 4 also show a common phenomenon, namely that the size of China's large cities is always lower than the fitted line. This finding is consistent with those of Long (2016) and Chen and Lu (2019). In Figure 4, the goodness of fit reaches 94.8 percent, but the cities that are ranked higher than 30 (points whose ln*Rank* are smaller than 3.4) suddenly deviate from the fitted line. In Figures 1 and 2, some large cities (points with low ln*Rank*) might also be under the fitted line,[10] but this phenomenon is not common. Furthermore, the deviation of large cities from the fitted line in Figure 4 is more significant

[7]Authors' calculation based on the latitude and longitude of the seat of local government.

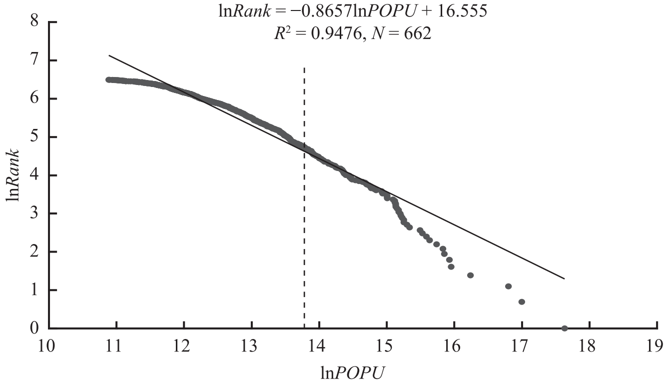
[8]Calculated using data from *China City Statistical Yearbook* (NBS, 2018).

[9]Authors' calculation based on data from Jiang (2020).

[10]The smaller the ln*Rank*, the larger the city.

than that in India and Indonesia (as shown in Figure 2). This is because China still uses the *hukou* system and has policies restricting the population inflow into large cities.

Figure 4. Nighttime light city (NLC) size distribution in China in 2016



Source: Data on NLC come from Jiang (2020).

Notes: ln*POPU* is the logarithm of population size of NLCs in China; ln*Rank* is the logarithm of the size ranking of NLCs in China. Each point indicates a NLC and the solid line represents the fitted value from the regression equation. The equation is the regression result of Equation (2) estimated by the data on NLC. The meanings of R^2 and *N* are the same as in Figure 1. The vertical dashed line indicates that the population equals 1 million.

The slope of the fitted line in Figure 4 is obviously smaller than that in Figure 3 but is larger than –1. This phenomenon is mainly due to the fact that there are too many small- and medium-sized cities. In Figure 4, the number of cities on the left-hand side of the vertical dashed line is greater than on the right-hand side. In particular, there are 554 cities on the left-hand side of the vertical dashed line with an average population of 0.32 million, while the right-hand side contains 108 cities with an average population of 3.34 million.[11] The large number of cities on the left lowers the overall slope. In other words, the flat slope is the result of encouraging the development of small and medium cities. In fact, many small cities in China have been experiencing population outflow from 2000 to 2010 (Long and Wu, 2016), and this phenomenon has also occurred in many developed countries. When migration restrictions are relaxed in the future, China's city size distribution might approach Zipf's law, which will be discussed in Section IV.

2. Large cities in large countries

The existing literature mainly focuses on the city size distribution in a single country, which is difficult to compare across countries. It is therefore difficult to identify a

[11] Authors' calculation using data from Jiang (2020).

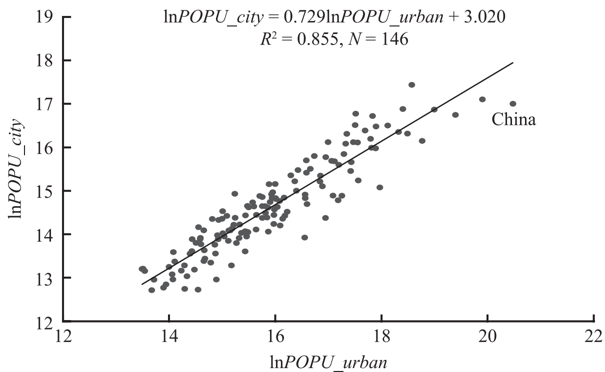
universal basic rule in urban systems. In fact, Zipf's law can also be used to analyze the city size distribution across countries. Based on Equation (1), when summing up the urban population in all cities we can write total urban population of a country as:

$$TP = (1 + 1/2 + \ldots + 1/n)\, P_1, \tag{3}$$

where P_1 represents the population of a primary city. We use Equation (3) to explore the city size distribution combining cross-country data, and two hypotheses can be drawn from Equation (3). First, the more cities there are in a country, the lower is the proportion of the primary city's population in the total urban population of a country (urban primacy ratio) (Hypothesis 1). Second, in large countries with a large number of cities, $1/n$ gradually tends to be 0, which means increasing the number of cities would not significantly change the urban primacy ratio. That is to say, the larger the urban population is in a large country, the larger the population is in its primary city (Hypothesis 2).

According to the above hypotheses, the urban primacy ratio reflects not only the city size distribution in a single country but also the comparison across countries. Figure 5 shows the relationship between the administrative primary cities and the total urban population for 146 countries in 2016. Surprisingly, R^2 is 0.855, which means that, to a large extent, one can predict the population of a primary city accurately based on the total urban population in a country.[12]

Figure 5. Relationship between urban population and primary cities in 146 countries in 2016



Source: The data on urban population and primary cities come from World Bank Database.

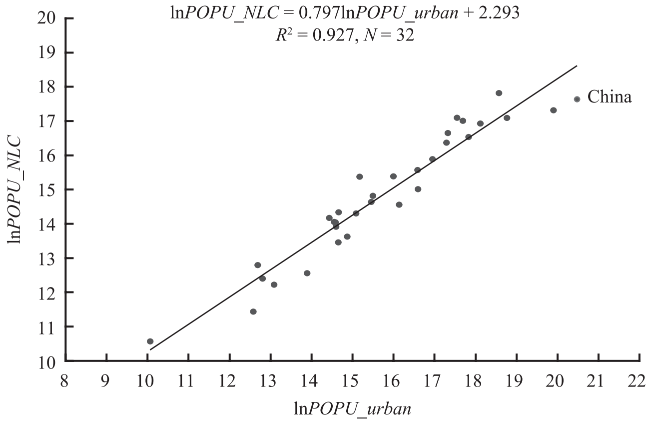
Notes: ln*POPU_urban* indicates the logarithm of total urban population in each country and ln*POPU_city* represents the logarithm of population of the primary city in each country. The equation is the regression result estimated using the data for 146 countries. R^2 indicates the goodness of fit of this regression result. N is the number of countries.

[12]Chen and Lu (2019) shown the positive relationship between primary city population and national population. Here, the relationship between national urban population and primary city population is derived from Zipf's law.

We could verify the above two hypotheses using the slope of the fitted line in Figure 5. First, the positive slope of the fitted line in Figure 5 indicates that the larger the total urban population a country has, the larger the primary city is. Second, the slope of the fitted line in Figure 5 is 0.729, which means that the greater the total urban population a country has, the lower the urban primacy ratio is.

China is below the fitted line (Figure 5), showing that the urban primacy ratio in China is low, consistent with previous results. Moreover, Figure 6 shows the relationship between the primary NLCs and the total urban population in Asian countries.

Figure 6. Urban population and primary nighttime light cities (NLCs) in Asia in 2016



Sources: Urban population data in each country are from the World Bank Database and the data on NLC come from Jiang (2020).

Notes: ln*POPU_urban* indicates the logarithm of total urban population in each country and ln*POPU_NLC* represents the logarithm of the population of the primary NLC in each country. The equation is the regression result estimated using the data for 32 countries. R^2 indicates the goodness of fit of this regression result. N is the number of countries. Countries or regions with only one NLC are excluded.

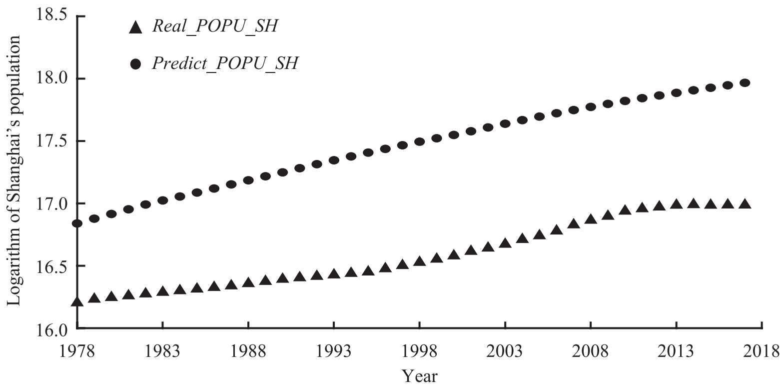
Compared with Figure 5, the primary NLC in China is closer to the fitted line, but is still below it. This indicates that, although the population of NLC-Guangzhou reached 45.52 million in 2016, the urban primacy ratio in China is still low compared to the level it should have as an Asian country.

3. City size evolution of Shanghai

Although NLC-Guangzhou is the largest NLC in China, Shanghai has always been regarded as the largest city in population and economic scale. According to the comparison in Table 1, the main problem regarding Shanghai is less economic connection with surrounding areas. Is the population in Shanghai approaching the fitted line of China, as shown in Figure 5, as a result of Yangtze River Delta integration?

In Figure 7, we estimate the benchmark population of a primary city (represented by a dot) based on the total urban population in China and the regression coefficient (0.729) and intercept term (3.020) in Figure 5. The benchmark population indicates that the population in the primary city corresponds to China's total urban population in each year. The triangle sign line represents the logarithm of real permanent residents in Shanghai. An obvious deviation exists between the dot line and the triangle sign line. Before 1995, Shanghai's real permanent resident population gradually deviated from the benchmark, which corresponded to the policy of restricting the large-city population in this period. However, from 1995 to 2010, the gap gradually narrowed, which was also related to the large-scale population inflow into Shanghai. It is noteworthy that Shanghai has strictly controlled the population inflow since 2013, with an increasing gap between the two lines.

Figure 7. Comparison between benchmark population and real population in Shanghai



Sources: Data for real permanent residents in Shanghai (*Real_POPU_SH*) come from the Shanghai Municipal Bureau of Statistics (2018). The predicted population in Shanghai (*Predict_POPU_SH*) was calculated by the authors using China's urban population data from 1979 to 2016 (World Bank Database) and the regression equation in Figure 5.

IV. Predicting population distribution in China

We are not only interested in the current deviation of China's city size distribution from Zipf's law, but also the future evolution of China's urban systems and the spatial distribution of the population. Prediction of the population distribution is valuable for urban planning and the corresponding provision of infrastructure and public services.

Re-examination of Zipf's law using a variety of data confirms that China's large cities are not large enough. In fact, China's central government has proposed reform of the *hukou* system and a development plan for the metropolitan areas. It is therefore inevitable that population will be further concentrated in areas around large cities. This migration process will bring some significant challenges and more research needs to be done. In this regard, in this section we attempt to shed light on population predictions and propose that all cities should prepare for free migration in the future.

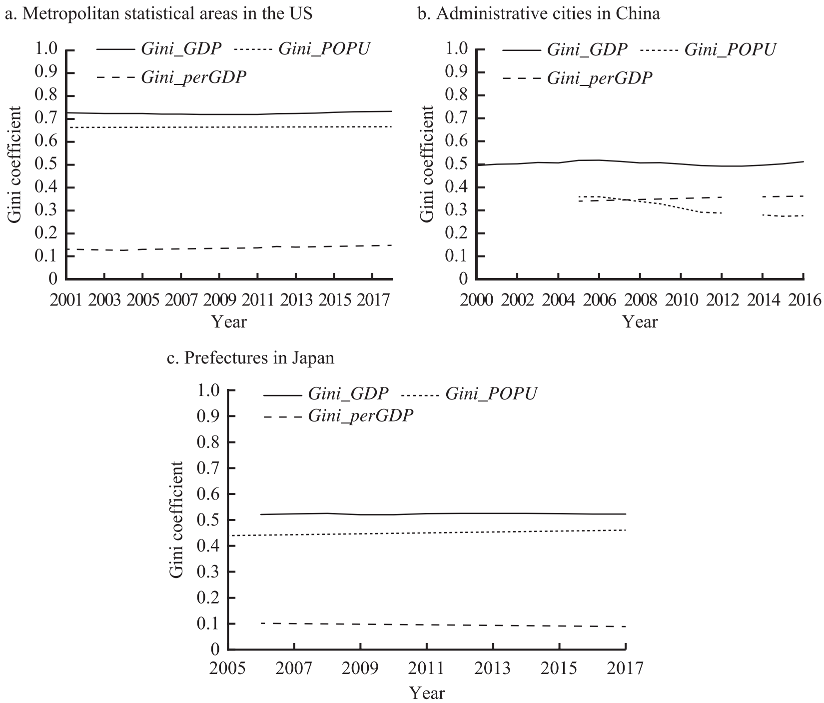
In the existing literature, researchers employ different models, including the bi-regional, net migration rate, and Hamilton–Perry models, among others, to forecast urban population, based on historical data such as population, mortality rate, immigration numbers, and the cohort change ratio (Renski and Strate, 2013; Smith et al., 2013; Wilson, 2016). The key assumption of these methods is that the past population growth pattern remains stable in the future. In countries where the population has been flowing freely, these methods are reasonable. However, China is transforming from population restriction to free migration, and its regional income disparity is still very large. These methods may therefore not be able to predict the future of population distribution in China.

A simplified method of prediction is to calculate the size of a primary city according to the fitted line in Figure 5 based on the assumption that China's urbanization rate will reach 80 percent, and then we can obtain the size of other cities according to Zipf's law. However, this method assumes that city size distribution in the future follows Zipf's law absolutely, which is unrealistic. We therefore propose a more practicable method to predict city size distribution in China.

Many factors drive inter-regional migration, and regional income disparity is one of the most important. With the process of population migration, regional per capita income would converge to a steady state, i.e. "spatial equilibrium" (Roback, 1982; Liu et al., 2018). Specifically, population migration from less-developed regions to developed regions increases natural resources per capita (e.g. land and tourism resources) in less-developed regions. The people's income in less-developed regions would therefore increase. In developed regions, the labor supply is sufficient due to population inflow, leading to low wage growth. Population migration would therefore narrow regional income disparity, which, in turn, weakens the motivation for migration from less-developed regions. Finally, the spatial distribution of the population would coincide with that of economic activities, and GDP per capita is largely equalized, although not completely.

Figure 8 employs the Gini coefficient to represent regional disparity and show the rule of spatial equilibrium. The Gini coefficients for GDP, population, and GDP per capita of the metropolitan statistical areas (MSAs) in the US, administrative cities in China, and prefectures in Japan were calculated separately for each year. A large Gini coefficient indicates a high degree of spatial agglomeration. As shown in Figure 8, economic activities (GDP) and population are highly concentrated in a few areas in the US and Japan, whereas GDP per capita across regions is much lower and is steady. In China, the spatial agglomeration of economic activities (GDP) and population is much lower than that in the US and Japan, but the Gini coefficient of GDP per capita is higher and is affected by regional policies. Interestingly, the Gini coefficient of GDP per capita in China decreases with the process of population agglomeration. If migration restriction is relaxed in the future, regional income disparity would therefore decrease and the population distribution would approach that of economic activities, reaching spatial equilibrium.

Figure 8. Comparison of economic activities and population concentration in the US, China, and Japan, 2000–2017



Sources: Data for the metropolitan statistical areas in the US are from the US Bureau of Economic Analysis, available from: https://apps.bea.gov/itable/iTable.cfm?ReqID=70&step=1. Administrative city data for China come from the NBS (2006–2018). Data for prefectures in Japan are from the Statistics Bureau of Japan, available from: https://dashboard.e-stat.go.jp/en/dataSearch.

Notes: *Gini_GDP*, *Gini_POPU*, and *Gini_perGDP* represent the Gini coefficients for the city-level GDP, population, and GDP per capita, respectively.

To predict the population distribution in the future, we must therefore first understand the factors affecting urban economic growth. Then, it is assumed that the population distribution gradually converges to that of economic activities because of inter-regional migration, which can roughly equalize inter-regional income per capita. Thus, city size distribution can be predicted using the idea of spatial equilibrium.

According to the core-periphery model, the relationship between the distance to the economic center and market potential follows a " ∽ "-shaped curve (Fujita and Krugman, 1995; Fujita and Mori, 1996; Fujita et al., 1999; Lu, 2017). Lu et al. (2019) also found that the distribution of economic activities and population in China was significantly related to the distance to major seaports (Tianjin, Shanghai, and Hong Kong) and national central cities (Beijing, Tianjin, Shanghai, Guangzhou, Chongqing, Chengdu, Wuhan, Zhengzhou, and Xi'an). We therefore construct Equation (4) to predict the city-level annual GDP growth rate:

$$GDP_{g,c} = \alpha_0 + \alpha_1 Distance\_Port_c + \alpha_2 Distance\_Port_c^2 + \alpha_3 Distance\_Port_c^3 + \alpha_4 Distance\_City_c + \alpha_5 Distance\_City_c^2 + \gamma \boldsymbol{X}_c + \eta_c, \tag{4}$$

where $GDP_{g,c}$ indicates the GDP growth rate in city c, $Distance\_Port_c$ indicates the minimum distance to one of three major seaports, $Distance\_Port_c^2$ is the square of $Distance\_Port_c$, $Distance\_Port_c^3$ is the cube of $Distance\_Port_c$, $Distance\_City_c$ indicates the minimum distance to one of nine national central cities, $Distance\_City_c^2$ is the square of $Distance\_City_c$, $\boldsymbol{X}_c$ is a vector that contains $Investment_g$, $National\_central\_city$, $\ln GDP_{2000}$, $\ln Manufacturing_{2000}$, and $\ln Service_{2000}$, and η_c is the random error term. Table 2 shows the definitions of the variables in this study.

Table 2. Variables and definitions

Variable	Definition
Dependent variable	
GDP_g	Annual GDP growth rate from 2000 to 2015
$Manufacturing_g$	Annual manufacturing GDP growth rate from 2000 to 2015
$Service_g$	Annual service GDP growth rate from 2000 to 2015
Independent variable	
Distance_Port	Minimum distance to one of three major seaports (Tianjin, Shanghai, and Hong Kong)
$Distance\_Port^2$	Square of *Distance_Port*
$Distance\_Port^3$	Cube of *Distance_Port*
Distance_City	Minimum distance to one of nine national central cities (Beijing, Tianjin, Shanghai, Guangzhou, Chongqing, Chengdu, Wuhan, Zhengzhou, and Xi'an)
$Distance\_City^2$	Square of *Distance_City*
$Investment_g$	Annual fixed asset investment growth rate from 2000 to 2015
National_central_city	Whether one of the nine national central cities
$\ln GDP_{2000}$	Logarithm of GDP in 2000
$\ln Manufacturing_{2000}$	Logarithm of manufacturing GDP in 2000
$\ln Service_{2000}$	Logarithm of service GDP in 2000

Sources: Data for $Investment_g$, *National central city*, $\ln GDP_{2000}$, $\ln Manufacturing_{2000}$, and $\ln Service_{2000}$ come from NBS (2001, 2016). Data for GDP_g, $Manufacturing_g$, $Service_g$, $Investment_g$ come from authors' calculation based on NBS (2001, 2016). Data for *Distance_Port* and *Distance_City* come from authors' calculation based on the latitude and longitude of the seat of government of each city.

We find that the distance to major seaports has a significant effect on GDP growth rate in each city (Table 3). Based on *China City Statistical Yearbook* data from the NBS (2001, 2016), it was found that the long-term GDP growth rate in each city follows a "∽"-shaped curve with increasing distance to major seaports (*Distance_Port*).

Table 3. Geographical factors and city-level annual GDP growth rate from 2000 to 2015

Variable		GDP_g		$Manufacturing_g$	$Service_g$
	(1)	(2)	(3)	(4)	(5)
Distance_Port	1.262 (3.772)	–6.026 (3.804)	–9.547*** (3.666)	–10.158** (4.480)	–10.359** (4.531)
$Distance\_Port^2$	0.304 (5.551)	8.971 (5.480)	12.622** (5.251)	14.117** (6.418)	10.589 (6.484)
$Distance\_Port^3$	–0.422 (2.398)	–3.363 (2.330)	–4.789** (2.235)	–5.250* (2.731)	–3.416 (2.760)
Distance_City	3.230* (1.686)	3.616** (1.597)	5.308*** (1.602)	6.542*** (1.958)	7.674*** (1.975)
$Distance\_City^2$	–4.151*** (1.325)	–3.647*** (1.257)	–4.721*** (1.232)	–7.620*** (1.505)	–6.166*** (1.519)
$Investment_g$		0.169*** (0.030)	0.166*** (0.029)	0.280*** (0.036)	0.108*** (0.036)
National_central_city			4.002*** (0.813)	3.570*** (0.978)	4.648*** (1.029)
$\ln GDP_{2000}$			–0.909*** (0.180)		
$\ln Manufacturing_{2000}$				–1.591*** (0.194)	
$\log Service_{2000}$					–0.830*** (0.213)
Constant	13.236*** (0.657)	10.305*** (0.816)	24.218*** (3.049)	30.742*** (3.239)	25.014*** (3.388)
Observations	259	259	259	259	259
R^2	0.096	0.194	0.298	0.519	0.186

Source: See Table 2 for data sources.

Notes: ***, **, and * represent significance at the 1, 5, and 10 percent levels, respectively. Standard errors are reported in parentheses. In columns (1), (2), and (3), the dependent variable is GDP_g. The dependent variable in column (4) is $Manufacturing_g$ and in column (5) if is $Service_g$. See Table 2 for the definitions of all variables.

As shown in column 2 of Table 3, the coefficients of the distance to the major seaports are statistically significant after accounting for the average annual investment growth rate from 2000 to 2015. During this period, large-scale investment, with strong government intervention, was allocated to central and western regions (Lu et al., 2019). In column (3), the impact of geographical factors (*Distance_Port*, $Distance\_Port^2$

and *Distance_Port*^{3}) is consistent with the theory of new economic geography, and R^2 reaches 0.298, showing that our prediction is convincing.

The fitted values of 259 cities were calculated based on column (3) in Table 3 using the coefficients of all variables except $Investment_g$, which is endogenous and strongly controlled by the government. Then, the long-term predicted annual growth rates of each city were obtained after adjusting the intercept term,[13] and we assume this growth rate is constant from 2015 to 2035.

The second step is to calculate the GDP of each city in 2035 based on the predicted GDP growth rate and total GDP in 2015. Then, according to the forecast total population of China in 2035 (UN, 2019), the total urban population in these 259 cities is calculated by assuming that their population share in total population in 2015 is constant throughout 2015–2035. The results show that in 2035 the total GDP of 259 cities is expected to be RMB171.7 trillion, and the GDP per capita reaches RMB136,140 (US$21,858) using the average exchange rate in 2015.[14] This prediction is consistent with the goal of becoming a moderately developed country, proposed by the Chinese government. To achieve this goal, the compound annual growth rate of GDP per capita from 2015 to 2035 is 4.42 percent, and that of total GDP is 4.64 percent.[15] Table 4 shows the predicted annual growth rate of some cities in China based on the results of Table 3.

Table 4. Predicted annual growth rate of some selected cities in China, 2015–2035 (%)

Top 10	Predicted growth rate	Last 10	Predicted growth rate
Tianjin	7.894	Jiamusi	–1.765
Shanghai	6.965	Hegang	–1.729
Guangzhou	6.566	Shuangyashan	–0.994
Beijing	6.525	Jixi	–0.116
Sanya	6.481	Jiayuguan	0.692
Wuhai	6.204	Yichun	1.369
Zhengzhou	6.164	Heihe	1.452
Xi'an	6.027	Mianyang	1.499
Fangchenggang	5.974	Harbin	1.592
Tongling	5.97	Deyang	1.665

Source: Authors' calculation based on the fitted values of column (3) in Table 3 using the coefficients of all variables except $Investment_g$ and adjusted intercept term.

[13]For each city, 19 is added to the fitted value. First, adjusting the intercept term will not change the relative GDP growth rate, and has little impact on the predicted city size. Second, at the 19th National Congress of the Communist Party of China, it was proposed that China would basically realize socialist modernization by 2035. After adjusting the intercept item, the GDP per capita in 2035 reaches approximately US$20,000, which is consistent with this goal.

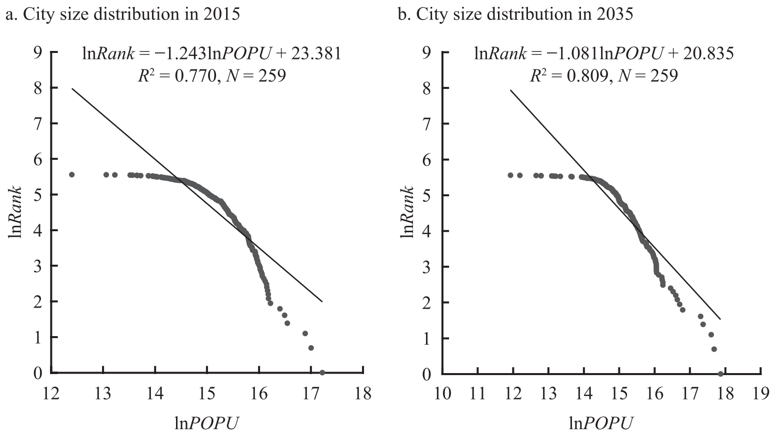
[14]Authors' calculation based on summing the predicted city-level GDP in 2035.

[15]Authors' calculation based on the predicted GDP in 2035 and the GDP in 2015 (NBS, 2016).

Third, according to the idea of spatial equilibrium, an assumption of regional inequality in 2035 is advanced here; that is, GDP per capita in Shanghai is 1.5 times that in Guizhou in 2035,[16] and GDP per capita disparity between the other cities and Guizhou follow proportional adjustment. This assumption is easy to understand. Actually, the Gini coefficient of urban GDP per capita is 0.11[17] when the assumption holds, which is similar to that of the US and Japan currently.

Finally, according to the GDP in each city and total urban population in 2035, the city size distribution in 2035 can be calculated based on the assumption mentioned above (Figure 9).

Figure 9. China's city size distribution in 2015 and 2035



Sources: City data in 2015 come from the NBS (2016). City data in 2035 come from the authors' calculation.

Notes: In panels a and b, ln*POPU* represents the logarithm of permanent residents in 2015 and 2035 respectively; ln*Rank* represents the logarithm of size ranking of permanent residents in 2015 and 2035 respectively. Each point indicates a city and the solid line represents the fitted values from the regression equation. The equation in each panel is the regression result of Equation (2) estimated by the data on administrative city (panel a) and predicted city (panel b). The meanings of R^2 and N are the same as in Figure 1.

[16]Guizhou is one of the poorest provinces in China and includes four administrative cities (Guiyang, Zunyi, Liupanshui, and Anshun). According to the 2016 *China City Statistical Yearbook* (NBS, 2017), Shanghai's GDP per capita in 2015 was RMB103,796, approximately 2.4 times that of average of the four administrative cities in Guizhou. The GDP per capita disparity between Shanghai and Guizhou reveals an apparent regional disparity in China.

[17]Authors' calculation based on the predicted city-level GDP per capita in 2035.

When the assumption holds, the population in Shanghai (metropolitan area), the largest city in China, will reach 57.39 million in 2035.[18] The population continues to flow to eastern coastal areas and core cities from the central and western regions. The size of 130 cities decreases by more than 200,000 from 2015 to 2035 based on this prediction.[19] Among these 130 cities, 103 cities have fewer permanent residents than registered population in 2015, 88 cities belong to the central and western regions, and 17 cities belong to the northeast region (as shown in the appendix). These cities should therefore prepare for population decline in advance.

As mentioned in the introduction, it is difficult to predict population distribution. The experience of many countries is that the spatial distribution of population would coincide with that of economic activities in the case of free migration. The aim of this paper is to discuss the spatial distribution of urban population in the future from the perspective of population redistribution in the entire country based on the idea of spatial equilibrium. There is no doubt that this method ignores many factors, including terrain, actual income, and cost of living. However, when free migration is allowed (consistent with recent reform), GDP per capita will be largely equalized, although not completely, and this method is also of great practical importance. The population distribution would be closer to Zipf's law with the convergence of regional GDP per capita in the future, and this process would bring significant challenges for all cities.

V. Conclusions

To understand and predict the population distribution in China, the basic rules of urban systems were evaluated in this paper. Based on different definitions, city size distributions in representative countries were studied using nighttime light data. An NLC reflects the urban economic functions and has the advantage of international consistency for its calculation method, which is applicable for all countries. More important, the NLC size distribution in different countries proves Zipf's law.

How to apply the basic rules of urban systems to China is the main focus of this study. Since the reform and opening up of China, the policies restricting the population growth of large cities and encouraging the development of small cities have had a

[18]Authors' calculation based on the predicted city-level GDP in 2035 and the assumption of regional inequality in 2035.
[19]See the appendix for the information for these 130 cities.

negative impact on urbanization and urban systems (Lu and Wan, 2014; Fang et al., 2017). The distortion in the urban system results in a loss in economic efficiency (Au and Henderson, 2006; Wang, 2010; Lu, 2017).

Zipf's law suggests that the size of China's large cities has been restricted. Beijing and Shanghai, two cities with stringent household registration systems, lag far behind Guangzhou in terms of economic connection with surrounding areas. In addition, the restriction of Shanghai's household registration system after 2013 has directly aggravated this distortion. Based on the idea of spatial equilibrium, we predicted the urban population distribution of China in 2035 and found, with the distribution of population converging to that of economic activities, that income disparity across cities is narrowed and city size distribution gradually approaches those that would be expected from Zipf's law. To reduce the distortion in China's city size distribution, we propose the following three suggestions.

First, China should accelerate reform of the *hukou* system in large cities. According to the *Key Tasks for New Urbanization Construction in 2019*[20] released by the National Development and Reform Commission in China, cities with population between 3 and 5 million in the core urban area should relax restrictions on new migrants. In the future, *hukou* system reform should focus more on cities with more than 5 million urban permanent residents in the core urban area. According to the data in the *China Urban Construction Statistical Yearbook* (NBS, 2020), there were 10 of 281 cities with more than 5 million permanent residents in the core urban area. If we subtract the registered population from the permanent resident population, the subtracted number indicates the level of population migration. This shows that the total net number of migrants of these 10 cities is 41.17 million, which accounts for 49.14 percent of total migrants of the 281 cities. That is, most migrants are still restricted by the *hukou* system, while the current *hukou* system reform focuses more on small cities. The *hukou* system in large cities still has a huge effect on interregional migration and should be reformed as soon as possible.

Second, China should break administrative boundaries and develop metropolitan areas around megacities. The development of metropolitan areas in Beijing and Shanghai, which is restricted by provincial-level boundaries, has fallen behind Guangzhou and Shenzhen. Three reforms are worth undertaking: (i) promote land

[20]See http://www.gov.cn/zhengce/zhengceku/2019-09/29/5435018/files/56eef378a89749a68d386cd3ba8b7159.pdf (online; cited March 2020).

system reform and break the boundary constraints of urban land-use allocation; (ii) strengthen the cooperation between megacities and surrounding areas through highway and rail transit, forming a highly efficient spatial network; and (iii) achieve the complementarities between megacities and their surrounding areas in industrial structure and resource integration.

Third, China should follow the basic rules of population migration and market forces in resource allocation. The performance assessment of local government officials in underdeveloped areas should be changed based on per capita income and quality of life. Less-developed regions should focus on the development of modern agriculture, tourism, natural resources, and other industries according to their local comparative advantage. At the same time, the restrictions on population migration should be removed. Developed regions should update their understanding of population inflow and the scale economy in modern urban development. Along with *hukou* system reform, China should also increase the supply of land, infrastructure, and public services in large cities to accommodate more migrants and alleviate urban problems.

References

Au, C. and V. Henderson, 2006, "Are Chinese cities too small?" *Review of Economic Studies*, Vol. 73, No. 2, pp. 549–76.

Anderson, G. and Y. Ge, 2005, "The size distribution of Chinese cities," *Regional Science and Urban Economics*, Vol. 35, No. 6, pp. 756–76.

Arshad, S., S. G. Hu and B. N. Ashraf, 2018, "Zipf's law and city size distribution: A survey of the literature and future research agenda," *Physica A: Statistical Mechanics and Its Applications*, Vol. 492, No. 15, pp. 75–92.

Berry, B. J. and A. O. Kozaryn, 2012, "The city size distribution debate: Resolution for US urban regions and megalopolitan areas," *Cities*, Vol. 29, No. 1, pp. S17–23.

Bee, M., M. Riccaboni and S. Schiavo, 2013, "The size distribution of US cities: Not Pareto, even in the tail," *Economics Letters*, Vol. 120, No. 2, pp. 232–7.

Chen, Z. and M. Lu, 2019, "Urban systems and urban development in the People's Republic of China," in G. H. Wan and M. Lu, eds, *Cities of Dragons and Elephants: Urbanization and Urban Development in China and India*, Oxford: Oxford University Press, pp. 81–102.

Eeckhout, J., 2004, "Gibrat's law for (all) cities," *American Economic Review*, Vol. 94, No. 5, pp. 1429–51.

Eeckhout, J., 2009, "Gibrat's law for (all) cities: Reply," *American Economic Review*, Vol. 99, No. 4, pp. 1676–83.

Fang, L., P. Li and S. F. Song, 2017, "China's development policies and city size distribution: An analysis based on Zipf's law," *Urban Studies*, Vol. 52, No. 12, pp. 2818–34.

Farrell, K. and P. Nijkamp, 2019, "The evolution of national urban systems in China, Nigeria and India," *Journal of Urban Management*, Vol. 8, No. 3, pp. 408–19.

Fujita, M. and P. R. Krugman, 1995, "When is the economy monocentric? von Thünen and Chamberlin unified," *Regional Science and Urban Economics*, Vol. 25, No. 4, pp. 505–28.

Fujita, M. and T. Mori, 1996, "The role of ports in the making of major cities: Self-agglomeration and hub-effect," *Journal of Development Economics*, Vol. 49, No. 1, pp. 93–120.

Fujita, M., P. R. Krugman and T. Mori, 1999, "On the evolution of hierarchical urban systems," *European Economic Review*, Vol. 43, No. 2, pp. 209–51.

Gabaix, X., 1999, "Zipf's law for cities: An explanation," *The Quarterly Journal of Economics*, Vol. 114, No. 3, pp. 739–67.

Henderson, J., A. Storeygard and D. Weil, 2012, "Measuring economic growth from outer space," *American Economic Review*, Vol. 102, No. 2, pp. 994–1028.

Huang, X. Y. and C. Yost-Bremm, 2018, "The power law within a metropolitan area," *Cities*, Vol. 72, pp. 201–5.

Jiang, Y., 2020, "Spatial dynamics and driving forces of Asian cities," *ADB Economics Working Paper Series* No. 618 [online; cited August 2020]. Available from: https://www.adb.org/publications/spatial-dynamics-driving-forces-asian-cities.

Jiang, B. and T. Jia, 2011, "Zipf's law for all the natural cities in the United States: A geospatial perspective," *International Journal of Geographical Information Science*, Vol. 25, No. 8, pp. 1269–81.

Kausik, G. and B. Basu, 2009, "City size distributions for India and China," *Physica A: Statistical Mechanics and Its Applications*, Vol. 388, No. 13, pp. 2682–88.

Lewis, W. A., 1954, "Economic development with unlimited supplies of labour," *The Manchester School*, Vol. 22, No. 2, pp. 139–91.

Li, S. and D. Sui, 2013, "Pareto's law and sample size: A case study of China's urban system 1984–2008," *GeoJournal*, Vol. 78, No. 4, pp. 615–26.

Liu, Y. L., M. Lu and K. H. Xiang, 2018, "Balance through agglomeration: A race between geography and policy in China's regional development," *China & World Economy*, Vol. 26, No. 6, pp. 72–96.

Long, Y., 2016, "Redefining Chinese city system with emerging new data," *Applied Geography*, Vol. 75, pp. 36–48.

Long, Y. and K. Wu, 2016, "Shrinking cities in a rapidly urbanizing China," *Environment and*

Planning A, Vol. 48, No. 2, pp. 220–22.

Lu, M., 2017, *The Power of Space*, Shanghai: Truth & Wisdom Press (in Chinese).

Lu, M. and G. H. Wan, 2014, "Urbanization and urban systems in the People's Republic of China: Research findings and policy recommendations," *Journal of Economy Surveys*, Vol. 28, No. 4, pp. 671–85.

Lu, M., P. F. Li and H. Y. Zhong, 2019, "The new era of development and balance: Spatial political economics of new China's 70 years," *Guanli Shijie* (*Management World*), Vol. 35, No. 10, pp. 11–23.

NBS (National Bureau of Statistics of China), 1986–2013, 2001, 2006–2018, 2016, 2017, 2018, 2019, *China City Statistical Yearbook*, Beijing: China Statistics Press (in Chinese).

NBS (National Bureau of Statistics of China), 1999, *Fifty Years of Cities in New China*, Beijing: China Statistics Press (in Chinese).

NBS (National Bureau of Statistics of China), 2020, *China Urban Construction Statistical Yearbook*, Beijing: China Statistics Press (in Chinese).

OECD (Organisation for Economic Co-operation and Development), 2012, *Redefining "Urban": A New Way to Measure Metropolitan Areas*, Paris: OECD Publishing.

Peng, G. H., 2010, "Zipf's law for Chinese cities: Rolling sample regressions," *Physica A: Statistical Mechanics and its Applications*, Vol. 389, No. 18, pp. 3804–13.

Renski, H. and S. Strate, 2013, "Evaluating alternative migration estimation techniques for population estimates and projections," *Journal of Planning Education and Research*, Vol. 33, No. 3, pp. 325–35.

Roback, J., 1982, "Wages, rents, and the quality of life," *Journal of Political Economy*, Vol. 90, No. 6, pp. 1257–78.

Rosen, K. T. and M. Resnick, 1980, "The size distribution of cities: An examination of the Pareto law and primacy," *Journal of Urban Economics*, Vol. 8, No. 2, pp. 165–86.

Rozenfeld, H. D., D. Rybski, X. Gabaix and H. A. Makse, 2011, "The area and population of cities: New insights from a different perspective on cities," *American Economic Review*, Vol. 101, No. 5, pp. 2205–25.

Smith, S. K., J. Tayman and D. A. Swanson, 2013, *A Practitioner's Guide to State and Local Population Projections*, Dordrecht: Springer.

Shanghai Municipal Bureau of Statistics, 2018, *Shanghai Statistical Yearbook*, Beijing: China Statistics Press.

Su, H. L., 2020, "On the city size distribution: A finite mixture interpretation," *Journal of Urban Economics* [online; cited March 2021]. Available from: https://doi.org/10.1016/j.jue.2019.103216.

UN (United Nations), 2019, *World Population Prospects 2019*, New York: United Nations Press.

Veneri, P., 2016, "City size distribution across the OECD: Does the definition of cities matter?" *Computers, Environment and Urban Systems*, Vol. 59, pp. 86–94.

Wang, X. L., 2010, "Urbanization path and city scale in China: An economic analysis," *Jingji Yanjiu* (*Economic Research Journal*), Vol. 45, No. 10, pp. 20–32.

Wilson, T., 2016, "Evaluation of alternative cohort-component models for local area population forecasts," *Population Research and Policy Review*, Vol. 35, No. 2, pp. 241–61.

Wen, Z. Q., 2016, "Zipf's law analysis of urban scale in China," *Asian Journal of Social Science Studies*, Vol. 1, No. 1, pp. 53–8.

Zipf, G. K., 1949, *Human Behavior and the Principle of Least Effort*, Cambridge: Addison-Wesley Press.

Appendix

Prediction for cities that decrease by more than 200,000 population from 2015 to 2035

City	Region	Population outflow (million)	City	Region	Population outflow (million)
Nanyang	Central	4.48	Yiyang	Central	1.18
Fuyang	Central	4.36	Anyang	Central	1.17
Zhoukou	Central	4.31	Jining	Eastern	1.16
Baoding	Eastern	3.72	Zhangjiakou	Eastern	1.13
Harbin	Northeast	3.53	Shuangyashan	Northeast	1.12
Shaoyang	Central	3.37	Heihe	Northeast	1.07
Nanchong	Western	3.33	Leshan	Western	1.06
Ganzhou	Central	3.33	Wenzhou	Eastern	1.03
Suihua	Northeast	3.27	Kunming	Western	1.02
Shangqiu	Central	3.13	Maoming	Eastern	1.02
Handan	Eastern	3.12	Mudanjiang	Northeast	1.00
Huanggang	Central	3.07	Jiujiang	Central	0.92
Dazhou	Western	2.85	Xinzhou	Central	0.90
Xingtai	Eastern	2.80	Fuzhou (Jiangxi)	Central	0.89
Heze	Eastern	2.78	Huainan	Central	0.88
Qiqihar	Northeast	2.76	Hegang	Northeast	0.87
Zhumadian	Central	2.71	Yichun (Heilongjiang)	Northeast	0.86

(Continued on the next page)

(continued)

City	Region	Population outflow (million)	City	Region	Population outflow (million)
Mianyang	Western	2.58	Xuchang	Central	0.81
Linyi	Eastern	2.48	Ankang	Western	0.74
Bozhou	Central	2.37	Chuzhou	Central	0.73
Jingzhou	Central	2.36	Zigong	Western	0.71
Yuncheng	Central	2.28	Ya'an	Western	0.69
Xinyang	Central	2.26	Yueyang	Central	0.69
Qujing	Western	2.16	Guilin	Western	0.68
Shangrao	Central	2.15	Changde	Central	0.68
Suzhou	Central	2.10	Changzhi	Central	0.65
Bazhong	Western	2.05	Baoji	Western	0.63
Lu'an	Central	2.03	Datong	Central	0.63
Xiaogan	Central	2.03	Loudi	Central	0.60
Weinan	Western	2.02	Shiyan	Central	0.60
Xinxiang	Central	1.88	Puyang	Central	0.60
Yongzhou	Central	1.85	Tieling	Northeast	0.59
Zhanjiang	Eastern	1.82	Luohe	Central	0.58
Jiamusi	Northeast	1.77	Qingyuan	Eastern	0.57
Tianshui	Western	1.75	Jinzhong	Central	0.56
Yulin	Western	1.74	Shanwei	Eastern	0.56
Pingdingshan	Central	1.72	Qinzhou	Western	0.55
Baoshan	Western	1.72	Qitaihe	Northeast	0.55
Zunyi	Western	1.66	Changchun	Northeast	0.53
Huaihua	Central	1.64	Baiyin	Western	0.52
Hengyang	Central	1.59	Yunfu	Eastern	0.51
Yichun (Jiangxi)	Central	1.58	Xining	Western	0.50
Deyang	Western	1.56	Huludao	Northeast	0.49
Guigang	Western	1.55	Heyuan	Eastern	0.46
Linfen	Central	1.51	Jilin	Northeast	0.44
Yibin	Western	1.49	Lanzhou	Western	0.43
Meizhou	Eastern	1.46	Suizhou	Central	0.39
Suining	Western	1.45	Yuxi	Western	0.38

(Continued on the next page)

Pengfei Li, Ming Lu / 35–62, Vol. 29, No. 4, 2021

(continued)

City	Region	Population outflow (million)	City	Region	Population outflow (million)
Neijiang	Western	1.44	Jingmen	Central	0.37
Jieyang	Eastern	1.36	Xianning	Central	0.36
Shijiazhuang	Eastern	1.36	Jiaozuo	Central	0.35
Kaifeng	Central	1.35	Anshun	Western	0.33
Ziyang	Western	1.33	Weifang	Eastern	0.33
Jixi	Northeast	1.31	Chenzhou	Central	0.32
Anqing	Central	1.31	Siping	Northeast	0.30
Guangyuan	Western	1.29	Chaozhou	Eastern	0.29
Luzhou	Western	1.25	Cangzhou	Eastern	0.26
Meishan	Western	1.24	Chaoyang	Northeast	0.26
Luoyang	Central	1.24	Huangshi	Central	0.25
Hengshui	Eastern	1.23	Wuzhong	Western	0.24
Ji’an	Central	1.23	Shaoguan	Eastern	0.22
Guang’an	Western	1.22	Zhaoqing	Eastern	0.22
Shantou	Eastern	1.21	Zhangjiajie	Central	0.22
Xianyang	Western	1.19	Xiangyang	Central	0.20
Hanzhong	Western	1.19	Liaocheng	Eastern	0.20

Source: Authors’ calculation.

Notes: China is divided into the eastern, central, western, and northeast regions. The eastern region includes Beijing, Tianjin, Hebei, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, and Hainan provinces. The central region includes Shanxi, Anhui, Jiangxi, Henan, Hubei, and Hunan provinces. The western region includes Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, and Xinjiang provinces. The northeast region includes Liaoning, Jilin, and Heilongjiang provinces. The population outflow is calculated as the result of subtracting the predicted population in 2035 from permanent residents in 2015.

(Edited by Zhinan Zhang)