A Study of Income Segregation in Large Chinese Cities

A thesis submitted in partial satisfaction of the requirements for the degree Master of Urban and Regional Planning

by

Jiren Zhu

2013
ABSTRACT OF THE THESIS

A Study of Income Segregation in Large Chinese Cities

by

Jiren Zhu

Master of Urban and Regional Planning
University of California, Los Angeles, 2013
Professor Paavo Monkkonen, Chair

Income segregation is an issue of great importance to scholars and policy makers because it is thought to exacerbate inequities of social outcomes such as education, social capital, and employment. China makes an ideal case study for the topic of income segregation. It has the largest population of any country, some of the highest levels of income inequality in the world, and many megacities. Using data from 20 large cities in China, this paper measures levels and patterns of income segregation in large Chinese cities with new measurement techniques. Findings show that there is a negative correlation between segregation levels of studied cities and their level of economic development and size, a sharp contrast to patterns in cities of the United States and Hong Kong. Additionally, the shape of the segregation profile varies greatly across the cities. There are three clear groups of cities based on their internal income segregation patterns. Possible explanations for the differences among these groups include the size of the built-up area, level of economic development, and the size of the immigrant population.
The thesis of Jiren Zhu is approved.

Anastasia Loukaitou-Sideris

Rui Wang

Paavo Monkkonen, Committee Chair

University of California, Los Angeles

2013
# Table of Content

1. Introduction ............................................................................................................. 1

2. Literature Review .................................................................................................... 5

3. Research Design: Data and Methodology ............................................................. 10
   3.1 Data and Framework ......................................................................................... 10
   3.2 Methodology .................................................................................................... 12

4. Analysis of Findings .............................................................................................. 14
   4.1 Correlation Analysis on Ordinal Index ............................................................ 14
   4.2 Interpreting Pairwise Index .............................................................................. 19

5. Conclusion ................................................................................................................ 31

Appendix I .................................................................................................................... 34

Appendix II ................................................................................................................... 36

References .................................................................................................................... 38
1. Introduction

Income segregation refers to the uneven geographic distribution of income groups within a certain area, and is a complex, multidimensional phenomenon (Reardon and Bischoff 2010). Empirical evidence suggests that income segregation may lead to inequity of social outcomes, such as the phenomenon of “urban village”, the existence of “gated community” and so on. Different income groups are isolated from each other for a number of reasons. Low-income households will locate in neighborhoods with relatively lower average income than will high-income households because they are unable to afford the high price of housing where high-income households select to live. However, high-income households place more value on environment, public resources, and traffic accessibility in their living areas; hence they are usually willing to pay a higher price to live in neighborhoods located in central areas of the city, which consist of households with higher income.

If a person’s social, economic, or physical outcomes are indirectly affected by the average income of the neighborhoods he/she lives in (Jencks and Mayer1990, Sampson, Raudenbush and Earls 1997, Leventhal and Brooks-Gunn 2000, Morenoff 2003, Sampson, Raudenbush and Sharkey 2008), then income segregation will cause more unequal outcomes between low- and high-income households than their expected differences in income alone. In a highly segregated region, higher-income households may gain more benefits than lower-income households not only because
of the difference in their own incomes, but also because of the differences in their respective neighbors’ incomes (Reardon and Bischoff 2010).

While income segregation is empirically entangled with racial segregation in the United States, China presents a totally different case study in this field. On the one hand, there is limited racial segregation in China. Thus, research on urban segregation in China focuses on socioeconomic segregation (Zhang 1995). On the other hand, due to the late start of relevant studies and limited data access, scholars seldom conduct analyses of income segregation that covers multiple metropolitan areas or provide a broad picture of the segregation issue in China.

Understanding income segregation in Chinese cities is important. The region has some of the highest levels of income inequality in the world and the largest population of any country, as well as many megacities. Thus, the spatial location of these different income groups has major implications for social outcomes and governance. Compared to the early studies of urban spatial segregation in western countries in the 1950s, China’s income segregation issue has mostly been studied since the 1980s (Sun and Wu 2009). This is largely due to the changes in the economic system and housing policies that were based on the reform and opening-up strategy accepted in 1978. However, from another point of view, the transitional economy and politics distinguishes China from other countries because of its unique features of urban income segregation. Indeed, before the market reform, urban households enjoyed
relatively equal incomes, housing conditions and other forms of social welfare such as education, medicine and facilities. Industries usually provided housing for workers to live around the work place, while teachers clustered in the school apartments. Most large cities displayed a mixed residential situation rather than socioeconomic segregation created by income difference (Yeh et al.1995, Li and Wu 2007).

The traditional socio-spatial structure of cities was transformed during the 1990s when the housing distribution system changed from a welfare-oriented one to a housing market. The Central Government of China concurrently implemented a wide range of reform across government-owned corporations after the 1980s, breaking down the original structure of enterprises and allowing a number of different corporate forms which resulted in a mixture of public and private capital (Zou and Ouyang 2008). These policy changes are thought to have led to the rapid development of income segregation in China, especially in large cities. Consequently, poor migrants flood into big cities like Beijing and Shanghai for job opportunities, shaping clustering areas of low-income households located on both outskirts and the downtown segments; meanwhile wealthy households tended to concentrate in the gated communities of city centers (Ma and Xiang 1998, Hu and Kaplan 2001, Wu 2005). It seems that large cities in China are heading down the path of income segregation, similar to what has happened in many US cities in recent decades.

Yet empirical evidence on the phenomenon is limited. This study is the first to
measure the levels and patterns of the residential segregation of different income
groups in a large number of Chinese large cities (20), utilizing recently developed
measurement techniques that allow for the disaggregation of segregation across
income distribution. The segregation analysis focuses on the dimension of evenness in
segregation outlined by Massey and Denton (1988). The unit of neighborhood defined
by the author with income reported over nine categories in diverse cities of China is
applied to the whole study.

The study outlines the general patterns of residential segregation in large Chinese
cities, finding that different income groups are spatially segregated in varying degrees.
The most important finding is that the segregation level tends to decrease as the city
becomes larger and more developed. This pattern is in sharp contrast to that found in
the United States where segregation levels are higher in more developed and larger
cities than in less developed and smaller cities (Reardon et al. 2006, Reardon el al.
2009, Reardon and Bischoff 2010). A second finding is that there is a significant
variation in the patterns of segregation across the income distribution between these
cities, and there are three clear groups identified to describe the features of
segregation patterns. For each type of segregation pattern, explanations are proposed.

This paper proceeds as follows. The following section is a review of empirical
literature on residential segregation in China, with a basic comparison to the related
studies in US. The third section presents data, methodology and research design.
Findings and related analysis are specified in the fourth part. The paper concludes with a brief summary of findings, implications and questions for further research.

2. Literature Review

Residential disparity is believed to occur in two instances. First, there is an effective ranking system in society, which differentiates residents according to status or prestige; second, this ranking system is matched by divisions in the housing market (Abu-Lughod 1969). Since the race issue inevitably influences the construction of socio-spatial structure, the study of socioeconomic structure segregation in US is always associated with racial segregation (Wilson 1978, Massey and Denton 1993, Bayer, McMillan and Rueben 2003, Reardon and Bischoff 2010). Massey and Denton (1988) have proposed five different dimensions of urban segregation and their basic measurements, which largely improved the research on measuring segregation levels.

However, as early as 1950s, Bell (1954), Duncan and Duncan (1955) measured segregation between two population groups using a dissimilarity index and an isolation index. After several decades of development, researchers have largely improved the calculation of indices and now employ spatial indices and entropy indices to address the issues of multiple groups and ordinal variables (Morgan 1975, Sakoda 1981, Jargowsky 1996, Reardon and Firebaugh 2002, Reardon and O’Sullivan 2004, Reardon and Bischoff 2010, Monkkonen and Zhang 2011).
In the United States, the study of spatial segregation started since the 1950s and focused on both racial and socioeconomic issues, and was combined with a complete development of research methodology. In contrast, the research on urban spatial segregation in China has only been around since the 1980s and has concentrated on socioeconomic segregation. The content of Chinese scholars’ study can be generally divided into three categories.

The first one is the theoretical research on socio-spatial structure segregation. Sun and Wu (2009) points out that few scholars have developed independent theories on the segregation of Chinese cities prior to 2000; rather the literature was based on the summary of academic articles in this field from 1986 to 2007. Instead, researchers usually described and borrowed theories from Western schools of thought, such as the Neoclassical School, Ecology School, Behaviorism and so on. The method of factorial analysis from the Ecology School has influenced Chinese scholars greatly and has been largely applied to the segregation studies (Yu 1986, Feng and Zhou 2003, Li and Wu 2006). However, since Wu (2001) has proposed that socio-spatial integration should be the theoretical foundation of segregation studies, Chinese scholars have begun to focus on the unique features of transitional China without employing Western theories. Examples include, the study of socio-spatial differentiation in “Post-Socialist” cities (Li, Wu and Xue 2006) and the analysis of urban social space in the socialist transitional countries (Wei and Yan 2006), which injected new ideas in this realm, though the theoretical system has not been completely shaped (Sun and
The analysis of socio-spatial structure continually moves forward from static research to dynamic study, and is starting to focus on the models and features of socio-spatial structure segregation, which shapes the second category of segregation studies in China. Yu (1986) used the method of factorial analysis to study the socio-spatial structure of segregation in Shanghai’s urban area, indicating that population density and social status (job, housing, and education level) are mainly responsible for the phenomenon in the core area of Shanghai. Xu, Hu and Yeh et al. (1989) applied principal component and cluster analysis to discuss the socio-spatial structure of Guangzhou in 1985 by utilizing survey data of travelling and housing, discovering that population density, scientific level, the ratio of workers and cadres, housing quality and household structure create Guangzhou’s residential structure.

Zheng, Xu and Chen et al. (1995) employed 1990 National Census Data to analyze Guangzhou’s residential communities from a comparative aspect, pointing out urban policies on economy, housing and urban planning with the urban historic background playing a key role in shaping socio-spatial structure in Guangzhou. Besides Shanghai and Guangzhou, Beijing as the other big metropolis also has received great attention from many researchers. Feng and Zhou (2003) combined a factorial ecology approach with census data in 1982 and 2000, showing that the socio-spatial structure of Beijing tends to be gradually complicated. Factorial analysis reflects that the floating
population became a key factor in reconstructing the socio-spatial structure of Beijing in 2000. This conclusion is consistent with the research outcomes from the data analysis of the questionnaire survey in Beijing (Gu et al. 2003).

The third category focuses on measuring segregation levels of cities based on different socioeconomic elements. With the release of 2000 National Census Data, more and more Chinese scholars devote themselves to this field through measuring segregation level of big cities. By utilizing data units at the neighborhood level, Li and Wu (2006, 2006) have studied socio-spatial differentiation and residential segregation in the meso level of Shanghai’s urban space through factorial analysis and regression analysis. They conclude that in the lower spatial level, residents with stratified socioeconomic status are being sorted into stratified neighborhoods; however, the extent of socio-spatial differentiation is still much lower than that of Western countries. Logan and Li (2012) creatively employed GIS spatial method and index of dissimilarity to measure the segregation level of Beijing’s urban area. They argue that the factor of housing tenure greatly impacts residential segregation patterns.

On the other hand, Xing, Wang and Cao (2004) selected 523 residential community projects since 1990 to analyze the residential structure of Xi’an. They divided the houses into different categories based on prices and calculated residential segregation of Xi’an, indicating that the city’s condition is on the path towards reconstruction and differentiation; however, the changes in the housing market and income segregation
mainly cause the phenomenon. Xing (2005) has proposed a similar conclusion through a study of commercial housing price in Tianjin’s six urban districts. Sun and Wu (2008) get the “U” curve line of residential segregation index in Shanghai’s urban area through the analysis of house renting price, finding that low income households and high income households tend to be more segregated than middle income class. Last but not least, segregation studies grounded on data of social investigation have also obtained effective outcomes. Lu (2005) has measured 15 communities in Hefei and asserts that the gap between rich and poor people in Hefei is not so significant. However, Yang and Wang (2006) reach a totally different conclusion in Shanghai through testing eight communities in Pudong New Area.

In addition to the three categories discussed above, in recent years, several researchers have sought to study the segregation issue from a large scale, utilizing latest data and advanced methods. Liu, Yan and Cao (2010) have analyzed housing type variation and its influencing factors in transitional urban China, which covers 88 cities across the country. They conclude that household income and the registration system (hukou) impact housing choice significantly, which may lead to the phenomenon of residential segregation. However, due to the wide range of studied cities, the segregation levels and patterns of 88 cities are scarcely discussed in their research. On the other hand, Monkkonen and Zhang (2011) applied new measurements to Hong Kong, China, which has high population density and high level of income inequality. Their study finds out that high-income households are much more isolated than low-income
households in Hong Kong, which is quite different from many big cities in the United States (Massey and Denton 1993, Pendall and Carruthers 2003, Reardon et al. 2006, Reardon and Bischoff 2010). Though Hong Kong’s context is distinct from most cities in mainland China, it is still worthwhile studying metropolitan areas of mainland China from a broad perspective through new measurements.

3. Research Design: Data and Methodology

3.1 Data and Framework

This study measures income segregation of family households in the dimension of evenness proposed by Massey and Denton (1988) across 20 large cities in China. These cities are selected based on different population size and variant level of development; however, they share three common characteristics: (1) they are municipalities directly under the Chinese Central Government or province’s capital city; (2) they are all large cities with a population of more than a million people in urban areas; (3) there are at least 60 studied geographical units (neighborhoods) in each urban area of the city, the reasons for which are discussed below.

Data for this project comes from 2000 China Township Population Census Data, which is prepared by the China Data Center in the University of Michigan. Since the 2010 China Township Population Census Data is not yet available, this data set is the most recent statistical resource. It includes GIS maps 2000 township (for all
townships of mainland China) and a number of socioeconomic attributes attached to data points of the maps. Unlike census tract data in the United States, the 2000 China National Census Data is simply collected by main streets of urban districts and towns, and then attached to the points on GIS maps. Therefore, in order to create data tabulations for the study, I employed the function of proximity in GIS mapping to shape a network of each city’s data points into neighborhoods (Logan and Li 2012). The neighborhoods are larger than the census tracts in America, usually with 30,000 people in each neighborhood. In order to make the study more accurate and representative, cities with fewer than 60 neighborhoods in the urban area are not selected. The urban area of each city is defined based on the newest administrative division of State Council by 2012, which provides a uniform manner to limit the boundary of the urban area in each city.

Figure 1 presents an example of Chongqing to show creating neighborhoods from original data points with GIS mapping.
3.2 Methodology

My calculation method is based on the entropy index, which is a weighted average of the heterogeneity calculated for each sub-unit of a city and measures the difference between heterogeneity of the city for variable attributes. It overcomes the limitation of just measuring separation between two groups by index of dissimilarity and allows for measurement of segregation among multiple groups (Reardon and Firebaugh 2002). However, for measuring the separation of different income groups, Reardon et al. (2009) indicate that the multi-group entropy index fails to seize ordinal nature of the data, leading to a greater gap between high-income groups and low-income groups than it does between middle-income groups and high-income groups. Instead, they propose to employ ordinal entropy index on cumulative income groups to calculate...
the segregation level. In addition, a pairwise entropy index is also applied to the present study since the outcomes calculated by ordinal entropy index have a correlation with how to categorize income data. This two-group entropy index, will calculate the segregation value of each cumulative income category compared to the rest of the groups rather than taking a weighted average of these measures. It is expected to reduce the bias brought by subjective income categories of data and support the value of ordinal entropy index (Monkkonen and Zhang 2011). Therefore, the study will utilize these two indexes to measure segregation levels of variant income groups across 20 large cities in China.

Variant income groups are classified into three basic categories (high-, middle-, low-income household) based on the classification method utilized by National Statistics Bureau in 2000. The method divides sample households into five equally sized groups, and then sets up the reference standard depending on each group’s average income level (China Statistical Yearbook 2001). By applying this method, a high-income household is defined as a family household with an annual income of more than 100,000 RMB (roughly 16,000 dollars); the low-income household is a family household with an annual income of less than 20,000 RMB (roughly 3,000 dollars); and a middle-income family household has an annual income in the range from 20,000 RMB to 100,000 RMB.
4. Analysis of Findings

4.1 Correlation Analysis on Ordinal Index

Table 1 summarizes the values of ordinal entropy index in these 20 large cities, combining the population, GDP, and GDP per capita of each city. The detailed calculated function of ordinal entropy index is presented in the appendix II. The range of ordinal entropy index values in 20 cities is from 0.06 to 0.18. By observing the data, I find that it is difficult to generalize a correlation between the value of ordinal entropy index and other indicators. For instance, Shanghai and Changsha have similar index values; however, the former city has a population size and GDP ten times the corresponding attributes of the latter city. Moreover, Guiyang is fairly similar to Lanzhou based on their population sizes, GDP and GDP per capita; however, Lanzhou’s value of ordinal entropy index is only half of Guiyang’s value.

<table>
<thead>
<tr>
<th>City</th>
<th>Ordinal Entropy Index</th>
<th>Population in Urban Area (10,000 people)</th>
<th>GDP per Capita (Yuan)</th>
<th>GDP (Million Yuan)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shanghai</td>
<td>0.0590</td>
<td>1136.82</td>
<td>36,054</td>
<td>40,986</td>
</tr>
<tr>
<td>Beijing</td>
<td>0.0869</td>
<td>974.14</td>
<td>23,942</td>
<td>23,323</td>
</tr>
<tr>
<td>Chongqing</td>
<td>0.1489</td>
<td>896.49</td>
<td>8,770</td>
<td>7,862</td>
</tr>
<tr>
<td>Wuhan</td>
<td>0.0703</td>
<td>749.19</td>
<td>16,109</td>
<td>12,068</td>
</tr>
<tr>
<td>Guangzhou</td>
<td>0.1189</td>
<td>566.68</td>
<td>38,207</td>
<td>21,651</td>
</tr>
<tr>
<td>Shenyang</td>
<td>0.0947</td>
<td>485.04</td>
<td>19,336</td>
<td>9,378</td>
</tr>
<tr>
<td>Chengdu</td>
<td>0.0946</td>
<td>335.86</td>
<td>19,944</td>
<td>6,698</td>
</tr>
<tr>
<td>Harbin</td>
<td>0.1088</td>
<td>303.72</td>
<td>18,109</td>
<td>5,499</td>
</tr>
<tr>
<td>Changchun</td>
<td>0.0817</td>
<td>292.83</td>
<td>21,110</td>
<td>6,181</td>
</tr>
<tr>
<td>Nanjing</td>
<td>0.0750</td>
<td>289.52</td>
<td>26,789</td>
<td>7,755</td>
</tr>
<tr>
<td>Dalian</td>
<td>0.1518</td>
<td>267.78</td>
<td>29,506</td>
<td>7,901</td>
</tr>
<tr>
<td>Jinan</td>
<td>0.0978</td>
<td>264.46</td>
<td>25,010</td>
<td>6,614</td>
</tr>
<tr>
<td>Qingdao</td>
<td>0.1760</td>
<td>234.60</td>
<td>26,808</td>
<td>6,289</td>
</tr>
<tr>
<td>Taiyuan</td>
<td>0.0980</td>
<td>233.20</td>
<td>12,642</td>
<td>2,948</td>
</tr>
<tr>
<td>Guiyang</td>
<td>0.1634</td>
<td>186.92</td>
<td>11,538</td>
<td>2,156</td>
</tr>
<tr>
<td>Lanzhou</td>
<td>0.0850</td>
<td>181.54</td>
<td>14,908</td>
<td>2,706</td>
</tr>
<tr>
<td>Hangzhou</td>
<td>0.0986</td>
<td>179.18</td>
<td>37,831</td>
<td>6,778</td>
</tr>
<tr>
<td>Changsha</td>
<td>0.0600</td>
<td>175.41</td>
<td>23,673</td>
<td>4,152</td>
</tr>
<tr>
<td>Tianjin</td>
<td>0.1209</td>
<td>682.05</td>
<td>20,422</td>
<td>13,928</td>
</tr>
<tr>
<td>Xi’an</td>
<td>0.1067</td>
<td>393.47</td>
<td>15,288</td>
<td>6,015</td>
</tr>
</tbody>
</table>

Table 1 Overall Ordinal Entropy Index Value with Different Attributes
Since there is a close relationship between GDP and population size in each city, GDP per capita is recognized to be a more reasonable indicator which reflects the economic development of a city. By drawing data from 2001 Urban Statistical Yearbook of China, I put the 20 cities’ values of ordinal entropy index on scatter graphs, matched with their GDP per capita data and population amount. As visual analysis hardly reflects anything, a correlation analysis is employed in the scatter graphs. Additionally, in order to study the correlation between segregation level and changes of population and economic development, I add the data of GDP per capita change and population size change from 1990 to 2000 to the scatter graphs. The percentage change of GDP per capita from 1990 to 2000 reflects the level of city’s economic development in these 10 years to a certain degree; on the other hand, population growth and movement in 10 years can be inferred through the analysis of percentage change in the city’s population size from this time period. Last but not least, log value of the data is utilized to shrink the range of statistical numbers.
Figure 2 Correlation Analysis based on Ordinal Index and GDP per Capita
(Source: Author’s calculation with data from 2001 Urban Statistic Yearbook of China)

Figure 3 Correlation Analysis based on Ordinal Index and GDP per Capita Change
(Source: Author’s calculation with data from 2001 Urban Statistic Yearbook of China)
Figure 4 Correlation Analysis based on Ordinal Index and Population Amount
(Source: Author’s calculation with data from 1991 and 2001 Urban Statistic Yearbook of China)

Figure 5 Correlation Analysis based on Ordinal Index and Population Amount Change
(Source: Author’s calculation with data from 1991 and 2001 Urban Statistic Yearbook of China)

Figures 2 to 5 show the relationship between segregation level calculated by ordinal entropy index and the attributes I have applied. The slopes of the trend lines in all four
figures are negative, which indicates that when either the amount of population or GDP per capita in the city increases, the segregation level of the city decreases. In other words, more developed and larger cities tend to be less segregated across different income groups, while less developed and relatively smaller cities have a higher level of income segregation. This outcome, is totally different from circumstances found in Hong Kong and cities in the United States where segregation levels increase consistently with population size and economic development (Reardon et al. 2009, Reardon and Bischoff 2010, Monkkonen and Zhang 2011). One possible explanation for the unique feature of Chinese large cities is that the spatial turnover between original residents and wealthy residents in developed cities like Beijing and Shanghai shapes mixed-living conditions across the city. Those original households, though typically low- or middle-income, are able to live in the central urban area, which has the highest housing prices, by owning houses passed down from previous generations. However, due to the rapid urbanization processes of these cities, great many high-income households choose to live in suburban areas of cities to enjoy a high-amenity environment and natural views; this, largely decreases the segregation level of suburban areas in which low-income households are clustered based on the low housing prices (Wu and Phelps 2008). Compared to the most developed cities, some relatively less developed cities, such as Qingdao and Guiyang, have a shorter urban history and slower urbanization processes, and thus households are greatly segregated by spatial difference of housing price related to household income.
Nonetheless, several limitations require us to conduct a further analysis to study the issue. Though the trend lines all show a negative correlation, the $R^2$ in each figure reflects that none of the four indicators is an important explanation of the variation. One possible reason is that the research sample of the present study is too small to obtain significant outcomes, since a strong correlation is usually based on the test of more than 100 sample cities. Moreover, the value of ordinal entropy index is more or less influenced by the way in which income data are categorized. It is difficult for the ordinal index to indicate the segregation level of each income category in a city. Therefore, the pairwise entropy index is going to be applied to the project in the next step, which allows us to easily visualize segregation levels across the income distribution.

4.2 Interpreting Pairwise Index

Pairwise index is calculated for each cumulative category of income, and rather than taking a weighted average of these measures, a polynomial function is estimated to represent the curve of segregation values across the income distribution, and an index is calculated based on this curve (Monkkonen and Zhang 2011). According to the different shapes of the index curves across income distribution, I divide 20 cities into three groups to represent diverse segregation patterns.
The first group of cities with a similar segregation pattern is presented in Figure 6. It is obvious that these eight cities share a similar shape of segregation index curve across cumulative income groups. The segregation levels of these cities increase in correlation with the low-, mid-, and high-income categories; segregation levels are highest in the high-income category, while low-income households are in the least segregated group which is dispersed in the urban area. Jinan can be used as an example to specifically explain the segregation characteristics. Figure 7 below displays the distribution of low-income households, middle-income households and high-income households in Jinan’s urban area. Low-income households are not only located across the peripheral area, but are also dispersed across the urban core area. Compared to the low-income households’ locations, the middle-income group becomes more segregated and it is mainly concentrated around the urban center. However, for the high-income households, they are highly centralized in the urban
core area and isolated from the other two groups of households. From the low-income category to the high-income class, there is a vivid transformation process from even dispersion to uneven concentration and the segregation level continues to increase.

Figure 7 Segregation Level in Jinan across different Income Groups
(Source: Pictures made by author with 2000 China Township Population Census Data)
As capital cities of their provinces, each city in the first category is the political, economic and cultural center. The capitals have more job opportunities, better resources and a larger market, which attracts many people from other areas within the province every year. This group of people, known as “floating people”, is mostly comprised of low-income households trying to find good jobs to improve their quality of life in the capital city; they are expected to have played a key role in shaping socio-spatial structure of the cities in this category (Yi 2004, Ding 2009, Song 2011, Song et al. 2011). Based on the common fact that the urban central area is still the job center of the city, floating people choose to either rent houses in the central area because of the advantages of proximity or to buy houses in the periphery to settle down for long-term development (Liu, Yan and Cao 2010). Consequently, a considerable proportion of low-income households are dispersed across the city, causing a low segregation level in the low-income group. On the other hand, the city’s built-up area is found to be a possible factor that can influence its segregation level. A built-up area, by definition, is a city’s urbanized area which has generally complete municipal infrastructures. National Population Statistics are usually conducted based on the built-up area since most people reside there. According to 2001 Urban Statistical Yearbook of China (See Table 2, Appendix I), these eight cities on average have relatively small built-up areas among the whole studied objectives (the average built-up area of eight cities is 176km², compared to the whole studied cities’ average built-up area which is 235 km²). Transportation cost may be largely reduced by the small built-up area of the city, and thus the middle-income bracket is distributed in a
wide range of places instead of being concentrated in the urban center to save on housing expenditures. Therefore, it seems like only high-income households select to live in the central area of the city, which has the most expensive housing and best municipal amenities; low-income households and middle-income households intersperse within the city, creating the segregation profiles seen above of cities across cumulative income groups in the first category.

For the second category of the cities, Shenyang, Changchun, Tianjin and Chongqing share similar characteristics of segregation index curve on cumulative income groups, which show that the distribution of high-income households and low-income households are more segregated than the middle-income group (Figure 8). Meanwhile, high-income households have a much higher segregation level than the low-income family does.

Figure 8 Segregation Profile across Cumulative Income Groups (Second Category)
(Source: Author’s calculation with 2000 China Township Population Census Data)
In the case of Changchun (Figure 9), GIS maps tell us low-income households tend to be clustered in the peripheral area of the city; in contrast, high-income households are still concentrated in the urban center, which is highly segregated from the suburban area of Changchun. However, compared to the distribution of low-income households, the segregation level of high-income households is much higher. Middle-income households are dispersed across the urban area, indicating a relatively lower segregation level.
The difference of built-up area between cities in the first category and the second category may be the main cause for the changes in segregation patterns. Compared to the cities in the first category, the four objectives in the second group have a much larger built-up area, where average built-up area is roughly 256km². As the built-up area is increased in the city, transportation cost becomes a significant factor that influences both household’s location choice and land prices. Due to the increased commuting time and expenses caused by a larger built-up area, middle-income households tend to live close to the central area of the city instead of being dispersed across the city area. They are clustered around the urban center, illustrating an ascending tendency of segregation pattern within this income range. On the other hand, as the negative correlation between housing prices and the distance to the
central area of a city become more significant, lots of low-income households are kept from living close to the central area and they are concentrated in the city’s peripheral areas with the lowest housing prices (Glaeser and Gottlieb, 2009). Therefore, the segregation level of low-income households in cities of the second category is higher than in cities of the first category, though they are similar in that high-income households are centralized in the urban center and highly isolated from the other two groups.

Figure 10 below shows the levels of segregation across the income distribution for the third-tier of cities. The segregation profile across the income distribution has changed slighter compared to the curves of the other two groups for low-income households to high-income households. Reardon et al. (2006) present similar graphs for several US cities, all of which have a flat U-shape with less variation. Lower-income households generally experience a similar level of segregation as high-income households in these cities. However, with the distinct economic development and population size, though two cities share the similar shape of segregation index curve, they can have totally different segregation level across the cumulative income groups. Like Beijing and Dalian, their segregation profiles on cumulative income groups share similar features, Beijing’s segregation level is remarkably lower than Dalian’s level from low-income group to high-income group.
I use another pair of cities – Shanghai and Guiyang, to describe the characteristics of the segregation pattern in this group of cities more specifically. According to Figure 10, the segregation profiles of Shanghai and Guiyang are similar in shape, though the segregation level of Shanghai is much lower than the standard in Guiyang despite of different income groups. Figures 11 and 12, show low-income households and high-income households in Shanghai are dispersed across different urban districts; while the low-income group is concentrated in periphery of Guiyang, the high-income group is largely located in the central area of the city. The two groups of residents, are segregated from each other in Guiyang, but are more interspersed in Shanghai.
Indeed cities in the third category can be divided into two groups based on their population size, economic development and built-up area. Beijing, Shanghai and
Dalian are several of the most developed cities in China with huge population sizes and large built-up areas; however, Guiyang, Lanzhou and Taiyuan are in the bottom of the development standard of studied cities. Therefore, the possible shaping reasons behind their segregation profile may also be totally different. For the former type of cities, as I have mentioned before, due to the historical factor, there are a number of original residents living in ancestral houses located in the central area of cities like Beijing and Shanghai, even though the residents earn less and cannot afford the market housing prices there. However, the crowded living conditions and poor air quality caused by population pressure in the urban center of these cities lead many high-income households to move to the suburban areas. This phenomenon of “Counter Urbanization” has been found in a great many cities in developed countries (Berry 1980, Dahms and McComb 1999, Kahsia and Schaeffer 2010). As a result, the spatial turnover of original residents and some high-income households creates mixed living conditions across the city, largely decreasing the city’s overall segregation level. Nonetheless, as the most developed cities in China, these cities have the largest amount of floating people, meanwhile generating the highest percentage of high-income households compared to cities in other tiers. The huge gap between poor people and rich people caused by the features of the cities keep the segregation level of low-income household and high-income household higher than middle-income household does.

For the latter type of cities, low-income households account for a large part of the
total amount of people due to the relatively low level of economic development in Northwest China. A considerable proportion of low-income households barely just reached the level of having enough food and clothing in 2000 (Hussain, 2003). Hence even though the average built-up area of these cities is only 146km², most of the low-income households may still be isolated in the peripheral areas of the city and highly segregated from the other two groups. However, middle-income households have more flexibility in choosing their living places since the average housing price in these cities is affordable, except in the central area, and the transportation costs are relatively low. The dispersed distribution of middle-income households causes the segregation profile in the range of middle-income to consistently go down. On the other hand, it seems like most high-income households in these cities are involved in the business of mineral resources, judging from the ample mineral resources in the three cities (coal in Taiyuan, quartz in Lanzhou and phosphorite in Guiyang). Besides the better transportation and living conditions, they need to be clustered in the central area of the city for business transaction and market development. Consequently, high- and low-income households may be spatially separated from one another, creating the U-shape profile of segregation pattern in this group of cities.

Another notable point which cannot be neglected in this category is that there are still differences between cities in the same segregation pattern. Beijing, Shanghai and Dalian share a similar segregation pattern across different income groups, but the overall segregation level of Dalian is much higher than the level of the other two
cities. However, Dalian has a much smaller population and a lower standard of economic development than Beijing and Shanghai. Coincidentally, Guiyang’s overall segregation level is much higher than the level of Taiyuan and Lanzhou, though it has a smaller population and the lowest level of economic development out of those three cities. This relationship is consistent with what I have found in the previous correlation analysis, which indicates that there is a negative correlation between the segregation level of a city and democratic and economic factors.

5. Conclusion

The paper presents an analysis of segregation levels across spatial scales and the income distribution in 20 large cities in China. It applies a new methodology based on scholars’ relevant studies of US cities to measure the segregation levels and patterns in China from a broad perspective. According to the ordinal entropy index value of each city, I find that there should be a correlation between the segregation level and democratic and economic factors since the housing market and floating populations contribute much to the development of socio-spatial segregation in US cities (Tiebout 1956, Mills and Hamilton 1994, Monkkonen and Zhang 2011). I also find a negative correlation between the overall segregation level and the city’s population size and economic development. Households in cities with bigger populations and higher levels of economic development tend to be less segregated by income, which is in sharp contrast to the relevant findings of the cities in the United States and Hong Kong where bigger cities are consistently found to be more segregated, presumably
because more competitive land markets lead to greater neighborhood differentiation.

In addition, the shape of the segregation profile in different cities across the income distribution divides the studied cities into three categories according to their distinct segregation patterns. The first category of cities shows an obvious increase in segregation with family household income. These cities, such as Wuhan, Nanjing and Hangzhou, are all regarded as the economic center in the province and attract lots of floating people annually. These people, combined with small built-up areas, are expected to be the main reasons to explain the segregation pattern of cities in this class. Another group of cities: Chongqing, Tianjin, Shenyang and Changchun, has a relatively more segregated low-income group and less segregated middle-income households. This might be caused by these cities having larger built-up areas compared to the first tier of cities. For the third group, a number of cities share a flat U-shape with less variation in segregation level across cumulative income groups. They have a similar segregation level in low-income groups and high-income groups, and a lower segregation level in middle-income households. However, based on the different economic development and population size, the shaping reasons behind the segregation level in different cities of this group might be totally distinct. As the most developed cities in China, the spatial turnover between original residents and high-income households is probably responsible for the segregation pattern of Beijing and Shanghai; for cities like Guiyang, Taiyuan and Lanzhou, the segregation pattern is possibly based on the less developed economy and the big gap between rich and poor people. Last but not least, a notable point from the analysis of pairwise index
outcomes is that the different segregation levels of cities in the same segregation pattern also indicate the negative correlation between the segregation level and democratic and economic elements in those cities, which is surprisingly consistent with my first major finding of the study.

Although what causes the difference of segregation patterns and levels between the large cities in China still needs be confirmed in further studies, the overall segregation level and the broad profile of these typical large cities brings an important twist into the existing literature. I wish that this study can provide the foundation for further research on the issue of Chinese socio-spatial structure segregation from a large scale. After the release of the 2010 China Township Population Census Data, it would be exciting to conduct a comparative study utilizing the data in a new generation to analyze the changing levels and patterns of segregation in large Chinese cities from 2000 to 2010. Additionally, the different profile of segregation pattern across income groups raises the important question of how does the segregation pattern influence the residents in large cities of China?
Appendix I

Figure 13 Scatter Graph of Ordinal Entropy Index Value in Each Studied City based on Its Population in Urban Area

Figure 14 Scatter Graph of Ordinal Entropy Index Value in Each Studied City based on Its GDP per Capita
<table>
<thead>
<tr>
<th>Rank</th>
<th>City</th>
<th>Built-up Area (km²)</th>
<th>Rank</th>
<th>City</th>
<th>Built-up Area (km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Shanghai</td>
<td>550</td>
<td>11</td>
<td>Xi'an</td>
<td>181</td>
</tr>
<tr>
<td>2</td>
<td>Beijing</td>
<td>488</td>
<td>12</td>
<td>Taiyuan</td>
<td>177</td>
</tr>
<tr>
<td>3</td>
<td>Guangzhou</td>
<td>431</td>
<td>13</td>
<td>Hangzhou</td>
<td>177</td>
</tr>
<tr>
<td>4</td>
<td>Tianjin</td>
<td>386</td>
<td>14</td>
<td>Harbin</td>
<td>168</td>
</tr>
<tr>
<td>5</td>
<td>Chongqing</td>
<td>262</td>
<td>15</td>
<td>Lanzhou</td>
<td>163</td>
</tr>
<tr>
<td>6</td>
<td>Dalian</td>
<td>234</td>
<td>16</td>
<td>Changchun</td>
<td>159</td>
</tr>
<tr>
<td>7</td>
<td>Chengdu</td>
<td>231</td>
<td>17</td>
<td>Jinan</td>
<td>120</td>
</tr>
<tr>
<td>8</td>
<td>Shenyang</td>
<td>217</td>
<td>18</td>
<td>Qingdao</td>
<td>119</td>
</tr>
<tr>
<td>9</td>
<td>Wuhan</td>
<td>210</td>
<td>19</td>
<td>Changsha</td>
<td>119</td>
</tr>
<tr>
<td>10</td>
<td>Nanjing</td>
<td>201</td>
<td>20</td>
<td>Guiyang</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>4691 km²</td>
<td></td>
<td>Average</td>
<td>235 km²</td>
</tr>
</tbody>
</table>

**Table 2 Built-up Areas of Studied Cities**
(Source: 2001 Urban Statistic Yearbook of China)

<table>
<thead>
<tr>
<th>City</th>
<th>Multigroup Entropy Index</th>
<th>Population in Urban Area (10,000 people)</th>
<th>GDP per Capita (Yuan)</th>
<th>GDP (Million Yuan)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shanghai</td>
<td>0.0693</td>
<td>1136.82</td>
<td>36,054</td>
<td>40,986</td>
</tr>
<tr>
<td>Beijing</td>
<td>0.0775</td>
<td>974.14</td>
<td>23,942</td>
<td>23,323</td>
</tr>
<tr>
<td>Chongqing</td>
<td>0.1231</td>
<td>896.49</td>
<td>8,770</td>
<td>7,862</td>
</tr>
<tr>
<td>Wuhan</td>
<td>0.0614</td>
<td>749.19</td>
<td>16,109</td>
<td>12,068</td>
</tr>
<tr>
<td>Guangzhou</td>
<td>0.1029</td>
<td>566.68</td>
<td>38,207</td>
<td>21,651</td>
</tr>
<tr>
<td>Shenyang</td>
<td>0.0906</td>
<td>485.04</td>
<td>19,336</td>
<td>9,378</td>
</tr>
<tr>
<td>Chengdu</td>
<td>0.0809</td>
<td>335.86</td>
<td>19,944</td>
<td>6,698</td>
</tr>
<tr>
<td>Harbin</td>
<td>0.1092</td>
<td>303.72</td>
<td>18,106</td>
<td>5,499</td>
</tr>
<tr>
<td>Changchun</td>
<td>0.0861</td>
<td>292.83</td>
<td>21,110</td>
<td>6,181</td>
</tr>
<tr>
<td>Nanjing</td>
<td>0.0793</td>
<td>289.52</td>
<td>26,789</td>
<td>7,755</td>
</tr>
<tr>
<td>Dalian</td>
<td>0.1607</td>
<td>267.78</td>
<td>29,506</td>
<td>7,901</td>
</tr>
<tr>
<td>Jinan</td>
<td>0.0764</td>
<td>264.46</td>
<td>25,010</td>
<td>6,614</td>
</tr>
<tr>
<td>Qingdao</td>
<td>0.1557</td>
<td>234.60</td>
<td>26,808</td>
<td>6,289</td>
</tr>
<tr>
<td>Taiyuan</td>
<td>0.0807</td>
<td>233.20</td>
<td>12,642</td>
<td>2,948</td>
</tr>
<tr>
<td>Guiyang</td>
<td>0.1364</td>
<td>186.92</td>
<td>11,538</td>
<td>2,156</td>
</tr>
<tr>
<td>Lanzhou</td>
<td>0.0701</td>
<td>181.54</td>
<td>15,687</td>
<td>1,128</td>
</tr>
<tr>
<td>Hangzhou</td>
<td>0.1006</td>
<td>179.18</td>
<td>37,831</td>
<td>2,156</td>
</tr>
<tr>
<td>Changsha</td>
<td>0.1364</td>
<td>175.41</td>
<td>23,673</td>
<td>4,152</td>
</tr>
<tr>
<td>Tianjin</td>
<td>0.1296</td>
<td>682.05</td>
<td>20,422</td>
<td>3,928</td>
</tr>
<tr>
<td>Xi'an</td>
<td>0.0771</td>
<td>393.47</td>
<td>15,288</td>
<td>6,015</td>
</tr>
</tbody>
</table>

**Table 3 Multi-group Entropy Index Value with Different Attributes**
Appendix II

Appendix II lists the formulas of three segregation indices applied in the study: the multi-group entropy index (Reardon and Firebaugh 2002), the ordinal entropy index (Reardon 2009), and the rank-order entropy index (Reardon et al. 2006).

1. The multi-group entropy index (H) is essentially a weighted average of the entropy of sub-units of the city compared to the citywide entropy. It is estimated as follows:

\[ H = 1 - \frac{1}{TE} \sum_{j=1}^{J} t_j E_j \]  

where

\( T = \) the total number of residents;
\( t_j = \) number of residents in sub-unit \( j \) (\( j \) indexes sub-units);
\( E = \) the overall entropy of the city; and
\( E_j = \) the entropy in block \( j \).

The entropy for the whole city is calculated as follows:

\[ E = \sum_{m=1}^{M} \pi_m \log M \frac{1}{\pi_m} \]  

where

\( \pi_m = \) proportion of the population in income group \( m \); and
\( M = \) number of income groups.

The entropy for each sub-unit is calculated as follows:

\[ E_j = \sum_{m=1}^{M} \pi_{jm} \log M \frac{1}{\pi_{jm}} \]  

Where

\( \pi_{jm} = \) proportion in group \( m \) in block \( j \).
2. The ordinal entropy index (\( \Lambda \)) is calculated as follows:

\[
\Lambda = \sum_{j=1}^{J} \frac{t_j}{T} \cdot \frac{v - v_j}{v}
\]  

(4)

Where \( v \) is the entropy calculated using cumulative income groups \( (c_m) \), which are defined below. Note that log of base two is used so that the index has a maximum value of one.

\[
v = -\frac{1}{M-1} \sum_{m=1}^{M-1} c_m \log_2 c_m + (1 - c_m) \log_2 (1 - c_m)
\]

(5)

Cumulative income shares \( (c_m) \) are the sum of the proportion of the population in income groups \( (k) \), which are less than and equal to each income category \( m \). The formula is as follows:

\[
c_m = \sum_{k=1}^{M} \pi_k
\]

(6)

As with the multi-group index, the entropy based on cumulative income groups is also calculated for sub-units of the city indexed by \( j \):

\[
v_j = -\frac{1}{M-1} \sum_{m=1}^{M-1} c_{jm} \log_2 c_{jm} + (1 - c_{jm}) \log_2 (1 - c_{jm})
\]

(7)

\[
c_{jm} = \sum_{k=1}^{M} \pi_{jk}
\]

(8)

3. The pairwise segregation index \( H(g) \) and entropy values \( E(g) \) are computed that compare the segregation of households at each point \( g \) in the income distribution:

\[
H(g) = 1 - \frac{1}{T_E(g)} \sum_{j=1}^{J} t_j E_j(g)
\]

(10)

Where \( T \) and \( t_j \) are defined as in the multi-group entropy index, and \( E(g) \) and \( E_j(g) \) are the entropy values for cumulative income group at point \( g \) on the distribution citywide and in sub-unit \( j \), defined as follows:

\[
E(g) = g \log_2 \frac{1}{g} + (1 - g) \log_2 \frac{1}{1-g}
\]

(11)

\[
E_j(g) = \pi_{jg} \log_2 \frac{1}{\pi_{jg}} + (1 - \pi_{jg}) \log_2 \frac{1}{1-\pi_{jg}}
\]

(12)
References


Song Lan (2011). *Differentiation and Character Analysis of Xi’an City’s Residential Space Based on GIS*, Master’s Thesis in Urban Planning and Design, Xi’an University of Architecture and Technology.


of the Residential Areas in Metropolitan Xi’an since the 1990s, City Planning Review, 28 (6): 68-73.


