The Impact of City Size on Occupational Wages and Migration

BY

ZHAO GU
B.S., Beijing University of Technology, 2007
M.S., Acadia University, Canada, 2009

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics
in the Graduate College of the
University of Illinois at Chicago, 2016

Chicago, Illinois

Defense Committee:

Dr. Joseph J. Persky, Chair and Advisor
Dr. Helen Roberts
Dr. George Karras
Dr. Houston H. Stokes
Dr. Nebiyou Tilahun, Department of Urban Planning and Policy

ACKNOWLEDGMENTS

First and foremost I would like to express sincere gratitude to my Ph.D. advisor Professor Joseph Persky. I deeply appreciate his continuous mentoring, support and patience during my graduate studies and research at the University of Illinois at Chicago. I am truly indebted to him for providing me with flexibility to try different research topics based on my interests throughout my research. Without his guidance and advising, this thesis would never have been accomplished.

I would like to thank my committee members Professor Helen Roberts, Professor George Karras, Professor Houston Stokes and Professor Nebiyou Tilahun, for their valuable advice, suggestions and time. I also want to thank Professor Ali Akarca, for his kind help and encouragement during my Ph.D. study.

Lastly, I would like to thank Yuchen for his continuous help, understanding and encouragement. I would like to thank my parents for their unconditional trust and support.

ZG

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

<u>FIGURE</u> <u>PAGE</u>

LIST OF FIGURES (Continued)

FIGURE **PAGE**

SUMMARY

This dissertation studies the economic phenomenon of city size at the occupation level, focusing on two topics. The first topic is the relationship between city size and wage. According to agglomeration economy theories, workers engaged in some occupations tend to spatially cluster in big cities whereas some others are more likely to locate in smaller areas. As various occupations show different patterns with respect to city size, it is interesting to discover how the city size influences the wages at the occupation level. Besides a comprehensive descriptive analysis on wage and city size for each individual occupation, two counterfactual experiments are constructed by leveraging probabilistic modeling to quantify the effects of city size on wages for each occupation. In addition, a regression model is introduced to fully understand what characteristics associated with occupation lead to the wage premium.

The second topic is to explore the relationship between migration and city size. With a thorough data analysis over migration and metropolitan population for each individual occupation, it is found that workers from different occupations have diverse migration patterns. Thus, a comprehensive study on migration is conducted by dividing occupations into three categories: occupations attracted to large cities, occupations preferred to stay in small areas and the rest. Furthermore, a number of key attributes related to occupation are examined in order to understand why various occupations differ significantly in terms of migration. The model identifies that average education level, industry coverage and average residential population are three statistically significant attributes on the tendency of migrating to large metropolitan areas.

CHAPTER 1

OVERVIEW

City size is an important topic in urban and regional economy. Many regional factors are highly correlated with city size, such as education, regional economic growth, infrastructure and facilities. It is critical and interesting to understand the patterns of city size by considering a few important economic factors. In this dissertation, I will study the economic phenomenon of city size at the occupation level, focusing on two main topics. The first topic that I study is the relationship between city size and wage. Based on agglomeration economy theories, workers engaged in some occupations tend to spatially cluster together in big cities for higher productivity, more efficient communication and positive externalities, whereas some occupations are more likely to locate in smaller areas. As wage gaps usually exist between big cities and small cities and various occupations show different patterns with regard to city size, it is important to discover how the city size influences the wages at the occupation level. With a comprehensive descriptive analysis on wage and city size for each individual occupation, I demonstrate the relationships between city size and wages for top occupations in the United States and give detailed explanations using agglomeration economy theories. Two counterfactual experiments are implemented by leveraging probabilistic modeling and they are used to quantify the effects of city size on wages for each occupation. In addition, since different occupations present various patterns on wage gaps introduced by city size, I introduce a regression model to fully understand what characteristics associated with occupation lead to the wage premiums.

Since the relationship between city size and wage is quantified in the first topic, a natural extension is to explore the dynamics of city size. In the second topic, I establish the relationship between migration and city size. With a thorough data analysis over migration and metropolitan population for each individual occupation, I discover that workers from different occupations have diverse migration patterns. Thus, I conduct a comprehensive study on migration by dividing occupations into three categories: occupations attracted to large cities, occupations preferred to stay in small areas and the rest. Furthermore, it is critical to analyze why various occupations differ significantly in terms of migration. I study a number of key attributes related to occupation, including average age, average education level, industry coverage, overall occupational prestige, average wage, average residential population and unemployment rate, and establish the relationship between these occupational attributes and migration. My model accurately identifies that average education level, industry coverage and average residential population are three statistically significant attributes on the tendency of migrating to large metropolitan areas. This finding is still robust after removing multicollinearity effects among attributes and this clearly demonstrates that these three factors are indeed crucial to the migration patterns.

CHAPTER 2

CITY SIZE AND WAGE

2.1 Introduction

Previous research have documented that larger cities provide higher wages. Glaeser (2001) claimed that wages of workers were 33 percent higher in metropolitan areas than nonurban counterparts in 2000. Baum-Snow and Pavan (2012) pointed out that the wage gap among workers is wider in the larger city from 1980 to 2007. They proposed a theoretical model and provided a solution on discovering important factors related to city wages in a quantitative way. However, the majority of previous research consider workers across all occupations as a whole. Based on agglomeration economy theories, various occupations have drastically different intrinsic characteristics thus present different patterns in terms of city size, location and economy. To the best of my knowledge, no previous literature focused on the relationship between the city size and wages of different occupations.

Rosenthal and Strange (2004) shows that the characteristics of occupations determine their distributions across cities. Some occupations cluster together in large cities to take advantage of benefits of agglomeration economies and high productivities bring the higher wages. But for some occupations, especially traditional occupations are found less and less in large metropolitan areas but may concentrate in smaller cities, so the wage gap between large city and small city would be lower. As occupation is a key factor on determining the nature of jobs and affects wage significantly, incorporating occupation in the study is critical to understand the relationship between city size and wage and the analyses measured based on occupation can be more fine-grained and comprehensive than treating all workers as a whole. In this paper, I investigate how the city size affects wages of different occupations.

The hypothesis of this study is that city size should have different effects on wages of different occupations, because of the characteristics of occupations, agglomeration economies and other externalities. To test this hypothesis and understand the relationship between city size and wages for each individual occupation, I first extract data from Integrated Public Use Microdata Series (IPUMS) collected by Ruggles et al. (2015) and select years between 1980 and 2010 on the individual levels. For each individual record, a number of variables related to wage, city and occupation are extracted, including age, education, city size, industry, occupational standing, etc.

Based on the data collected in this chapter, I conduct a descriptive analysis on some representative occupations by correlating city size with wages. The results show that different occupations have various patterns about wage and city size. For skill intensive industries and service-oriented occupations such as finance, IT and lawyers, city size boosts the wages across all years from the obtained data. On the contrary, traditional manufacturing related occupations, e.g., assemblers and machine operators, display the negative relationship between city size and wage, namely wage decreases as the size of city increases. This clearly demonstrates that the relationship between city size and wage is highly affected by occupations of the workers.

The following content is organized as follows. Section 2.2 briefly introduces the agglomeration economy; Section 2.3 is the literature review; Section 2.4 states the problem; Section 2.5 presents the data used in this study; The results on descriptive analysis of the problem are illustrated in Section 2.6 and conclusion is in the last section.

2.2 Agglomeration Economy

Agglomeration economies, which are generally referred as location-specific economies of scale, were first acknowledged by Weber (1909). Marshall (1920) then provided more detailed descriptions about the sources of agglomeration economies. In his statement, increasing returns to scale must be obtained by the firms in the agglomeration economies. Based on that, he identified that information spillovers, local non-traded inputs and a local skilled labor pool are three reasons why such economies of scale can be achieved.

The sources of agglomeration economies given by Marshall only described the scenario that firms within the same industry cluster together in space to achieve localized external economies of scale. For the areas that firms in different industries cluster together geographically, Ohlin (1933) and Hoover (1937) employed a new classification to describe the nature of agglomeration economies. According to this classification, agglomeration economies can be divided into three types, namely internal returns to scale, localization economies, and urbanization economies.

Internal returns to scale are firm-specific economies of agglomeration (McCann (2001)). They are generated due to that a high level of investment and people takes place at a particular location. Hence these economies are internal to the firm and location specific. Internal increasing returns to scale are mostly found in manufacturing and services (Harvey (2009)). It can be found in both light industries and heavy or high-technology industries, but in heavier industries, internal scale economies are higher.

Localization economies are industry-specific economies of agglomeration. Localization economies come from geographically concentrated firms in the same industry. These clustered firms are connected by the technology they use, the products they provide, the markets they serve and the knowledge they share. The concentration of firms in space makes them more productive, and the competitions among firms also lead to productivity growth because firms are forced to innovate and improve.

Urbanization economies are the agglomeration economies which accrue to firms from different sectors. In the urbanization economies, the various activities which are not directly related to the sector experiencing internal returns to scale and localization economies (Karlsson (2010)), cluster in the local economy, providing services for the firms and workers of this sector. In the Hoover typology, urbanization economies are defined as city-specific economies of agglomeration and as cities grow bigger, urbanization economies become more important.

In large cities, agglomeration economies raise up the average productivity of firms and workers. These higher productivities arise from a variety of advantages in urbanization economies. For example, similar firms share suppliers or exchange ideas and technologies; firms in different industries share public goods and facilities; they can also gain from the experiences and innovations among one another, and labor markets are able to provide more narrowly specialized workers.

2.3 Literature Review

There are many literatures about the relationship between cities and skills. Glaeser (2001) showed that the wages of workers in cities are 33% higher than nonurban areas. While there is no doubt about the positive relationship between wages and city size, they argued that the urban wage premium is not only the result of higher productivity in cities. They used two different data sets NLSY and PSID to examine the urban wage premium. The NLSY data supported the wage level effect and the PSID data showed the wage growth effect. Therefore, the authors claimed that the urban wage premium is a combination of the wage level effect and wage growth effect. Based on this evidence, the authors further suggested that skills accumulated faster in metropolitan areas.

From the finding of Glaeser (2001), Glaeser and Resseger (2010) pointed out that skill learning effects are stronger in cities with more skills and not significant in cities with lower levels of skill. This strong connection between productivity and city population is the evidence for the existence of agglomeration economies. According to the view of agglomeration economy, they indicated that city density is important because proximity makes workers spread knowledge more quickly and firms share information more efficiently. The authors concluded that bigger cities are more attractive to workers with higher level of skills and human capital accumulates more quickly.

Davis and Dingel (2014) proposed a framework for both theoretical and empirical applications on the comparative advantages of cities. The distribution of skills and sectors across cities are simultaneously introduced in their model and they assume that individuals' comparative advantage determines sectoral employment. Based on the data in the year of 2000, they found that larger cities are skill abundant and specialize in skill-intensive activities. Besides these findings, they further claimed that their model makes more precise predictions compared to prior studies. First, they argued that cities' skill and sectoral distribution will be substantially overlapping and this conclusion is more realistic than previous research describing cities are entirely sorted along skills or polarized with regard to sectoral composition. Second, they claimed that cities' skill and sectoral distributions will demonstrate systematic variation based on the monotone likelihood ratio property.

Classic research on agglomeration economies focused on the overall effect of local determinants. Combes and Gobillon (2014) further identified the impact of specific mechanisms that lead to agglomeration effects, using aggregate regional data and individual data. In addition, they studied the determinants of other local outcomes that are employment and firm location selections. In the analysis, they discussed a number of empirical issues and proposed solutions to deal with these problems. The empirical concerns they addressed include endogeneity at local and individual levels, the choice of a productivity measure between wage and TFP, spatial scale, characteristics of firms and functional forms. The authors also discussed the approaches on quantifying agglomeration mechanisms.

There are growing interests on the research of city size and wage. Baum-Snow and Pavan (2012) studied the reasons of higher wages and productivity in larger cities through an empirical approach. The idea was to decompose the log wage growth into a number of components. The analysis demonstrated that within job wage growth had more contributions on the city size wage gap than that of between job wage growth. They set up a conceptual environment and counterfactual simulations on the National Longitudinal Survey of Youth 1979 (NLSY79) data suggested that returns to experience and wage level effects are the most critical factors with regard to the city size wage gap difference. Specifically, differences in wage intercepts for different city locations are the most important for medium and small cities, and differences in returns to experiences are more crucial for the city wage premia between large and small cities. In addition, they discovered that there are a few independent minor negative components on city size wage premia including sorting on unobserved ability within education group and differences in labor market search frictions.

In terms of finding that firms and workers have higher average productivity in larger cities (Rosenthal and Strange (2004); Melo et al. (2009)), Combes et al. (2012) discussed two main explanations, which are firm selection and agglomeration economies. In order to distinguish the effect of these two factors, they build up a model which incorporates a standard model of agglomeration and a generalized version of the firm selection. Their model predicted that, on the one hand, firm selection in larger cities left-truncates the distribution of the productivity; on the other hand, agglomeration economies that improve interactions in larger cities right-shifts the distribution. The authors applied the model prediction to the French data and found that firm selection cannot explain the differences in productivity across areas in France.

De la Roca and Puga (2012) used a different perspective to explain the question that wages are higher in bigger cities. They used a panel data set for Spain to examine three potential reasons. The first one they considered is spatial sorting of workers with higher abilities. The second reason they proposed is that bigger cities have some advantages and the third one is that workers in bigger cities can learn and gain more experience. Their finding is that the spatial sorting of more productive workers is not the main factor generating wages differentials across different city sizes. They argued that workers in bigger cities do not have higher initial productivity than smaller cities. The higher wages in bigger cities are due to the fact that working in bigger cities can allow workers to accumulate more valuable experience and enhance abilities. Compared with previous literature, authors emphasized more on the learning benefit from bigger cities, which caused higher wages in big cities.

Baum-Snow and Pavan (2013) studied the relationship between wage inequality and city size that shows a significant monotonic pattern from 1979 to 2007. Through controlled experiments from the data for the given time range, the authors identified that at least 23% of the overall log wage inequality can be explained by the city size feature alone with controlling observed skills. They claimed that skill groups and industries in larger cities show bigger increases in their wage dispersion in larger cities compared with smaller cities, and this generates the city size specific component of inequality growth. In the meantime, they suggested that the increases of demand for observed skills in larger cities leads to increasing growth on prices and quantities of observed skills in larger locations relative to smaller locations. The experiments also demonstrated that about one-third of the city size effect can be explained by the disproportionate industrial factor and agglomeration economies are the key reason of the rapid growth since 1979.

Previous literatures about city size and wages offered different explanations to state that workers in larger cities have higher wages. Productivity advantages in cities, efficiency gains, cost savings and knowledge spillovers are provided as reasons. However, Echeverri-Carroll and Ayala (2011) questioned that city size is an important determinant of high wages after controlling for knowledge spillovers, because they noticed that cities with large human capital spillovers in US are not the largest cities. They used 5 percent PUMS of the 2000 US census data to examine the effects of city size and intellectual spillovers on productivity. They found that even after controlling for learning spillover effects, each additional 100000 inhabitants in the local labor market caused individual hourly wages to increase by 0.12%. According to the results, the authors concluded that city size determines urban wages.

2.4 Problem Statement

This study investigates the relationship between city size and the average wages for different occupations in various industries. I explore how city size affects wage changes for different occupations. There are many literatures on city size and wages, and the main finding is that the wages of workers are higher in larger cities. The explanation they provide about the urban wage premium associated with large cities is that workers are more productive in bigger cities. Most previous literatures only focus on the effects of city size on the overall wage in cities and treat all population of cities as a whole. Some research such as Baum-Snow and Pavan (2013) considered the factor of skills and divided all population into different groups by skill levels. However, this is a sketchy division. A specific skill level can be found in various occupations or industries, and even one occupation could have people with different skill backgrounds. Thus, the coarse granularity of existing research works on cities overlooked the intrinsic differences among industries and the effects of various occupations and industries are not incorporated in the previous analyses.

2.5 Data Description

I obtained the data across 30 years from 1980 to 2010 on the individual level from Integrated Public Use Microdata Series (IPUMS) (Ruggles et al. (2015)) to conduct the analyses on city size and wage incorporating the characteristics of occupations. Specifically, the data is extracted from the annual American Community Survey (ACS) and the sample years are 1980, 1990, 2000 and 2010 with a 10-year interval. The sample density of the years 1980 and 1990 is 5.0% whereas the sample density of other years is 1.0%. In this study, city size is measured by the population, in hundreds, for all identifiable metropolitan areas from IPUMS. Since wage is the labor market variable, metropolitan area is the natural unit for the labor market.

A number of factors, including age, city, education, wage and occupation, are also extracted for the further study. To remove the effects of inflation in different years, wages are adjusted to the amount they would have represented in 1999, corresponding to CPI. Thus, the dollar amounts of wage are comparable and consistent across various years.

As the Census Bureau reorganized the occupational classification system across different years, the occupation definitions change with each census year. In order to have a consistent occupational classification system for the study, all the samples in the data set are aligned to the 1990 occupation scheme thus the evaluation on occupations has the maximum consistency over the 30-year study. Meyer and Osborne (2005) proposed a detailed comparison on occupational category methods. In addition, the obtained data set is composed of microdata, namely each record represents a person. Compared with the aggregated data from previous research, the microdata provides a more refined granularity on city and wage levels and the experiments that are carried out on the individual level leads to a better representation of the whole economy system.

2.6 Descriptive Analysis

2.6.1 Skill Intensive Occupations

In order to study the effects of city size on individual wage over time for different occupations, I select the most popular 50 occupations in the U.S. to conduct analysis. As shown in previous study (Glaeser (2001)), job skill is one of the most critical factors on wages in agglomeration economies, and bigger cities are more attractive to workers with higher level of skills. Since different occupations require various levels of skills, I first study the impacts of city size on wages for a few occupations that need high level of skills and discover their underlying links with agglomeration economies.

Financial services industry is one of the most typical industries with significant agglomeration economies. No matter retail banking, consumer financing or commercial banking, investment banking, they are all spatially concentrated to take advantage of narrow specialization and agglomeration economies. As seen in the United States, finance industry is clustered in the east coast, especially New

York City. I first conduct a descriptive analysis on the finance industry with intensive skilled workers with regard to city size and wages.

Figure 1 illustrates the relationship between the average wage in the finance industry and city size. It is shown that the average wage in finance industry is increasing by city size in different time period. In the figure, all cities are assigned to five bins with equal range according to the log city population. The bin value 1 represents the smallest bin with range of log population [5.273, 6.471], which corresponds to the city population [19500, 64570]. Bin 5 is the largest bin and its range is [10.062, 11.259], corresponding to the city population [2344345, 7762800]. In the data, the smallest city in year 2010 is the Waterbury in Connecticut State with population of 40900, whereas the largest city in year 2010 is the New York in New York State and it has 7658000 residents. The vertical axis is the average wage in the finance industry.

From the figure, it is clear that the average wages across four years all increase with the increase of city sizes. Specifically, the average wages in the city size bins from 1 to 3 in the year of 1980 have similar values, whereas larger cities in bins 4 and 5 have higher wages. In addition, this pattern is more significant in recent years, as the differences between wages in higher bins and wages in lower bins increase over time. In other words, the result is consistent with the theory of agglomeration economy.

Similar to the finance industry, firms on information technology and computer science are densely spatially concentrated. In the United States, most IT firms cluster in the San Francisco Bay area in California and formed the world famous Silicon Valley. To study this phenomenon of agglomeration economy, I further conduct an analysis on the IT industry and investigate the relations between average

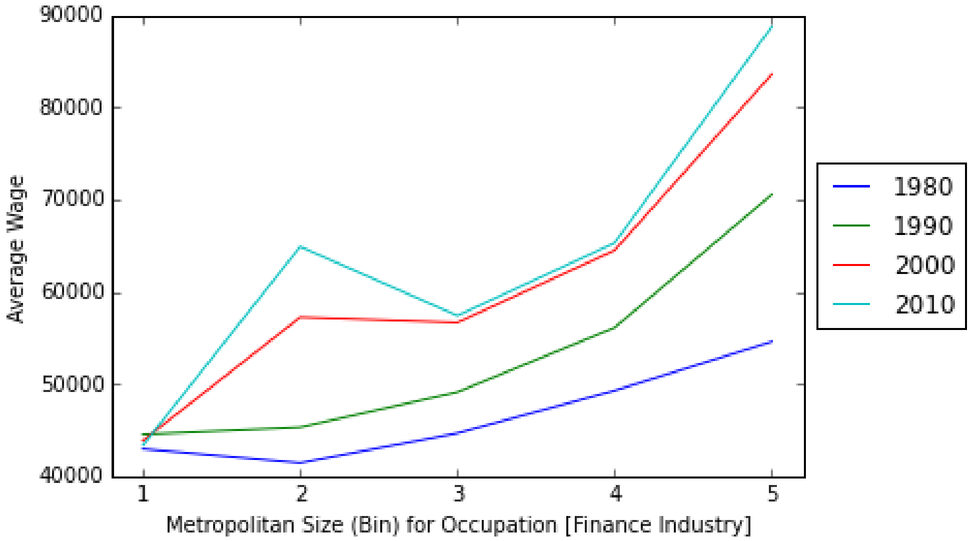

Figure 1. Relationship between binned city size and average wage for finance industry

wage and city size. Figure 2 demonstrates that in each decade since 1980 the average wage of IT industry rises with the city population.

Although the rising trend is not as significant as shown in finance industry, the cities in the top bin still have a big increase in wage. This means that a few largest cities can offer much higher average wages compared with large cities that are assigned to the city size bins 3 and 4.

The above two occupations show the similar trend in wage change, which is the average occupation wage goes up with the city population over time and in particular, there is a wage spike existing in the top large cities. Actually, finance and IT industries are not two exceptions. High-skill and high-technology industries are mostly clustered since they need to take advantage of benefits from agglomeration economies, such as knowledge spillovers, which require proximity.

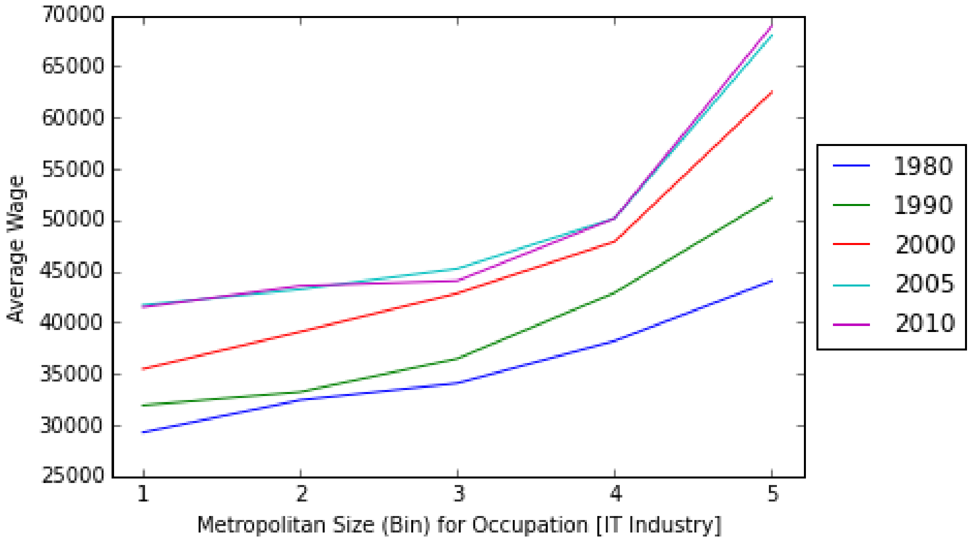

Figure 2. Relationship between binned city size and average wage for IT industry

In order to discover how agglomeration economies affect the relationship between wage and city size, I choose another two knowledge-based occupations for the analysis.

Figure 3 shows the trend on average wage and city size across all years for lawyers. As seen in the figure, agglomeration economies fan out the average wage in the largest cities for lawyers. Figure 4 presents the relations between wage and city size for primary school teachers. The positive externalities in large cities, for example the unionization, explain the positive influence of city size on wage for teachers. In summary, wages for high-skilled or knowledge-based occupations are higher in the larger cities.

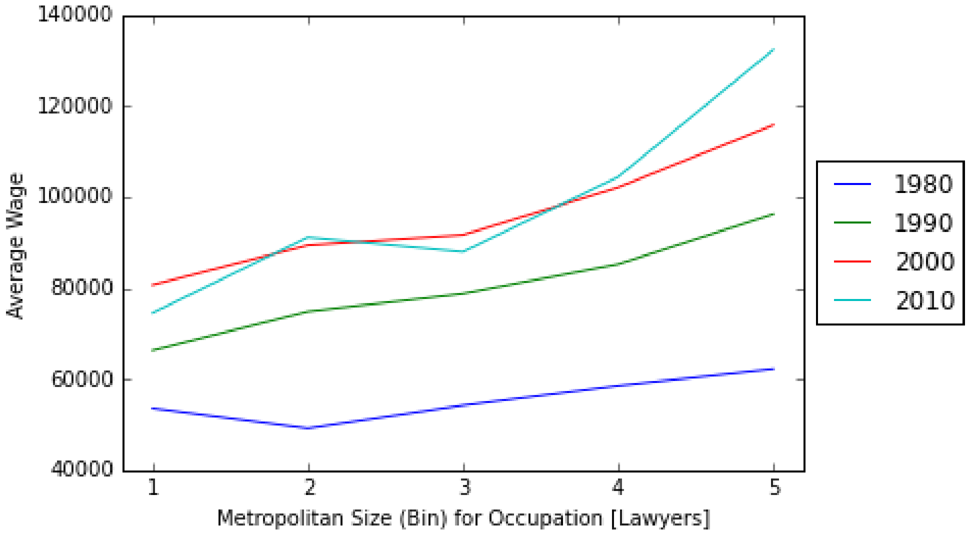

Figure 3. Relationship between binned city size and average wage for lawyers

2.6.2 Manufacturing Related Occupations

From agglomeration economy theories, not all occupations require large markets since different occupations carry various characteristics and not all of them need to be clustered and exhibit agglomeration economies. On the one hand, large cities are more diversified thus knowledge based and service oriented occupations can easily cluster and boom; on the other hand, smaller cities specialize and focus on producing and manufacturing more specific products. Thus, it is an important topic to investigate the relations of wage and city size for manufacturing related occupations.

I first show the results on a typical manufacturing related occupation, assembler, in Figure 5. As opposed to the positive effects of city size on wages, the average wage of assemblers does not increase with the city population as I have shown for previous occupations. Specially, the wage of assemblers of

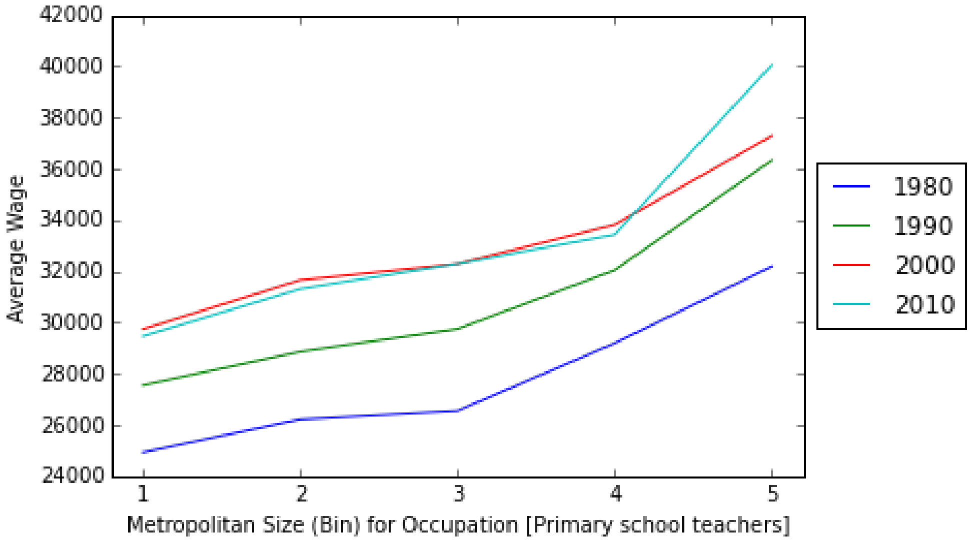

Figure 4. Relationship between binned city size and average wage for teachers

electrical equipment in larger cities, namely city size bin 4 and 5, is the lowest compared to the wages of assemblers in smaller cities. Meanwhile, the wages at mid-sized cities are the highest across all years. This phenomenon is quite distinct from that of knowledge based and service oriented occupations. This is because workers from traditional manufacturing occupations do not necessarily cluster in big cities, whereas smaller cities specialize in specific manufacturing related areas.

Figure 6 demonstrates the effects of city size on the average wage for another manufacturing related occupation, machine operators. The result again reflects that the wages on the Y-axis generally present a downward trend along the size of cities on the X-axis. Although this pattern is different from the ones from service oriented and knowledge based occupations, it follows the previous phenomenon in Figure 5. The characteristics of machine operators determine that these occupations do not need to be spatially

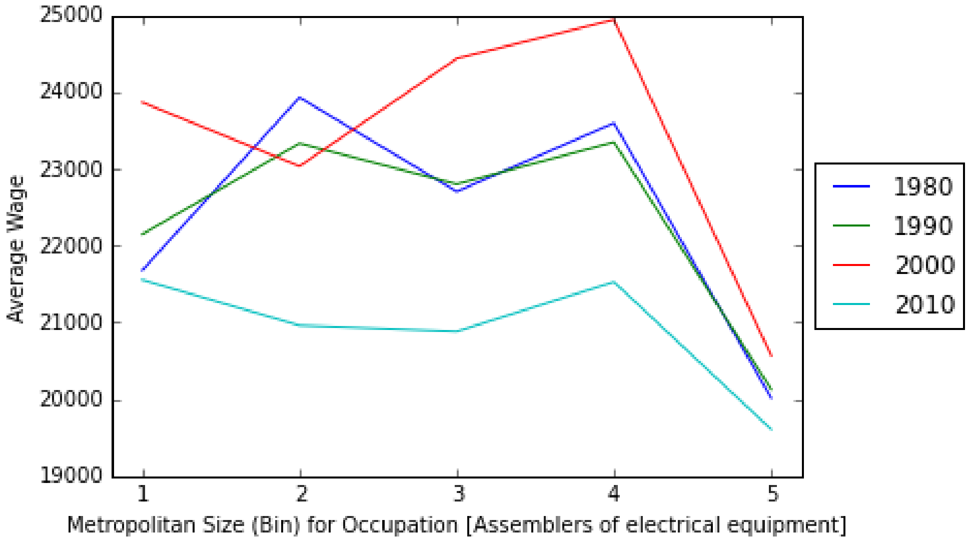

Figure 5. Relationship between binned city size and average wage for assemblers

concentrated in order to leverage the benefits of agglomeration economies. These figures again indicate that agglomeration economies are not significant for manufacturing related occupations as smaller cities tend to be industrially specialized though they are not as diversified as big cities.

2.6.3 Other Occupations

Besides the typical occupations mentioned above, all other occupations respectively have different responses to the city size. Wages of some occupations rise with the city population at different increasing rates. Some have significant increasing patterns, whereas increasing patterns for some are smooth. Occupations with decreasing wage patterns regarding the city size have the same story. Different occupations show various decreasing rates given the nature of each occupation. There are also some occupations which do not demonstrate specific patterns about the wage and city size. Table XIV

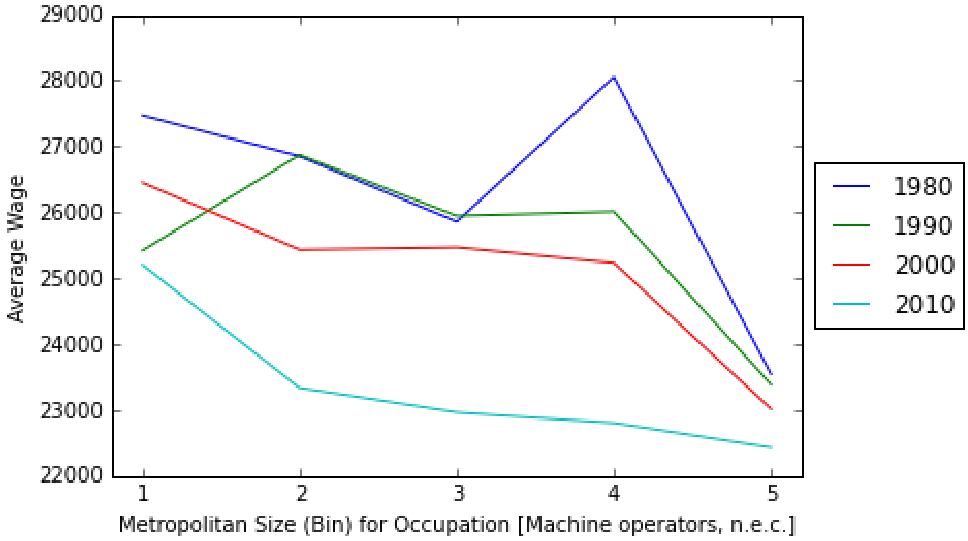

Figure 6. Relationship between binned city size and average wage for machine operators

and Table XV in the Appendix summarize the average wages for the most popular 50 occupations in the U.S. with the city size over 4 decades.

2.6.4 Summary

As seen from the above analyses, there are two categories of occupation which show clear relationship between city size and average wage. High-skilled and knowledge-based occupations demonstrate the increasing pattern of average wage with city size. These occupations take advantage of benefits from agglomeration economies and agglomeration economies fan out the average wage gap of these occupations in the largest cities. Therefore, occupations in this category are paid a lot more in the largest cities.

In contrast with occupations in the first category, traditional occupations, such as manufacturing related occupations show the opposite trend of average wage with respect to city size. The wages of this category do not increase with the city population. In fact, wages in larger cities for these occupations are lower.

2.7 Conclusion

According to the theory of agglomeration economy, some occupations and industries spatially cluster together in order to take advantage of the benefits of agglomeration economies, such as higher productivity, more efficient communication, positive externalities, etc. As the agglomeration economy develops, higher productivities also usually raise the wages. Given the wage gaps between big cities and small cities and various occupations present different patterns based on agglomeration economy, it is interesting to explore how the city size affects the wages of different occupations. In this chapter, I study the problem about the relationship between city size and wages for different occupations.

I collect personal level data from the year 1980 to 2010 and conduct a descriptive analysis on wage and city size for individual occupations. The descriptive analysis clearly shows various occupations represent drastically different characteristics with regard to city size and wage. This further demonstrates that the relationship between city size and wages has to be studied for each specific occupation instead of using all occupations as a whole.

CHAPTER 3

EFFECTS OF CITY SIZE ON WAGE

From the previous chapter, I have demonstrated that different types of occupations show various patterns on the relationships between city size and wage. One important question that is still not solved yet is how much effect of city size is imposed on wage while considering the differences among occupations. In addition, given the importance of occupations on studying city size and wage, it is interesting to analyze what characteristics of occupations are significant with regard to wage gaps. In this chapter, I will introduce two models that measure the effect of city size on wage by incorporating the impact of individual occupations, and also study the occupational attributes related to wage gaps introduced by city size.

3.1 Introduction

In this chapter, I collect data on city size and wages with additional factors that represent the skills, education levels, age and occupations of the sampled individuals. Given the attributes on occupations, I conduct a comprehensive analysis to discover the effects of city size on wages for top 50 occupations in the United States. A list of descriptive analyses demonstrates distinct impacts of city size for a wide range of occupations and agglomeration economy theories are introduced to explain the reasons of these critical distinctions and effects. To further investigate the effects of city size on wages for various occupations, I give a formal definition of the given problem and counterfactual experiments are established to evaluate the effects of city size for occupations.

To further study the effects of city size in a quantitative way in chapter 3, I construct two counterfactual experiments using nonparametric probabilistic methods. The first counterfactual experiment measures the impact of the city size distribution changes across different geographic locations on wage for a given occupation. The second counterfactual experiment directly explores the full effect of city size on wages for different occupations. Specifically, this is achieved by establishing a link between cities and corresponding rural areas. The results show that for some occupations, the wages would have been up to 28% if city size does not change over time, whereas some traditional occupations show little difference on wages between small cities and large cities or even negative correlations.

As it is shown above that different occupations have various responses in wages to the city size changes, it is interesting to explore what factors are related to wage changes. In other words, it is worthwhile to identify what characteristics of the occupation could explain the differences in wages. I construct a regression analysis by setting the wage difference contributed by city size of each occupation as the dependent variable, and a number of factors including education, age, gender, industry and occupational standing are introduced as the independent variables. A regression model is calculated to establish the linear relationship between the occupation-related factors and wage differences introduced by city wage. The model indicates that education is the only factor that is statistically significant with regard to the dependent variable, namely wage differences, although all factors show positive effects in various degrees based on their coefficients. This result further suggests that agglomeration economy is prominent for high skilled workers and skill intensive occupations tend to cluster in large cities with high wage premium.

3.2 Notations and Wage Modelling

In order to quantitatively understand the effect of city size on wage over time for different occupations, I employ a series of nonparametric methods to construct counterfactual experiments. Specifically, the wage is modelled as a likelihood function and it is decomposed into a number of factors corresponding to city size, education level, occupation and age group, thus the impacts of different factors can be accurately controlled and quantified.

I first introduce the concepts and definitions that I will use in the following sections. Nonparametric methods and wage decomposition have been widely used to study the effects of wage in previous research Lemieux (2006) and Baum-Snow and Pavan (2012). For each individual, we use w_i to denote the wage of a person i. A probability $P_{t,o}(a, e)$ is introduced to represent the joint probability distribution of age a and education level e with occupation o at time t. For the age factor, as it is a continuous variable, I specifically divide the data set into 9 age groups with 10 years as the group range. Namely, the age group is [10, 20), [20, 30), ... [90, 100). In addition, I use five education levels to group the data samples based on the highest education that the individual has completed. The following Table I shows the education group that I used in the data:

To consider location and individual skills including age and education, I analyze these three variables and decompose their joint probabilities using the following formulation based on the joint probability theory:

$$P_{t,o}(a, e, b) = P_{t,o}(a, e)P_{t,o}(b|a, e) \tag{3.1}$$

1	N/A or no schooling
2	Nursery school to grade 8
3	Grade 9 to Grade 12
4	1 to 4 years of college
5	5+ years of college

TABLE I

EDUCATION LEVEL CATEGORIZATION

The term $P_{t,o}(a, e, b)$ denotes the joint probability of a given age group, education level and city size bin for an occupation o and time t. This joint probability can be explained by a product of two factors. The object $P_{t,o}(b|a, e)$ describes the conditional probability of city bin b for an individual in age a and education level e with occupation o, whereas $P_{t,o}(a, e)$ represents the joint probability of age a and education level e. Here, these two factors are considered as one random variable in the conditional probability formulation.

Given the conditional probability distribution above, the probability of wage for a given time t can be represented as a combination of factors as:

$$P_{t,o}(w) = \int P_{t,o}(a, e) P_{t,o}(b|a, e) P_{t,o}(w|b, a, e) dbdade \tag{3.2}$$

In this equation, $P_{t,o}(w|b, a, e)$ is the wage distribution for age a, education group e and occupation o in location b at time t. One should note that the overall wage of an occupation o is decomposed into a number of related factors including age, education and city size by Equation 3.2, which enables quantitative studies on these factors and the effects of city size can be accurately evaluated via counterfactual experiments. Given the above annotations, the mean value of wage can be represented as:

$$W_{t,o} = \int P_{t,o}(w)wdw \tag{3.3}$$

In the following, I introduce two counterfactual experiments that allow measurements of different effects of city size on wages.

3.3 Counterfactual Experiments

In the previous subsection, I use the nonparametric method to model the wage distribution. In order to further study the effect of city size on the wage for different occupations, I construct a number of counterfactual experiments to measure what the wage would be if the city size growth pattern is controlled over time.

From Equation 3.2, it is clear that the wage for a given year and occupation is affected by city wage in the following two components:

- $P_{t,o}(b|a,e)$: illustrating the distribution of city size bin under the condition of age and education. The value of this term changes along with the city population distribution over time.
- $P_{t,o}(w|b,a,e)$: illustrating the distribution of wage conditioning on city size, age and education. This term indicates that the change of city population also influences the distribution of wage.

As these two terms representing different aspects of the impact of city size on wage, it is interesting to study which aspect has dominating effects that lead to the wage difference. Therefore, I construct two separate counterfactual experiments to analyze these two components on city size and quantify the relative effects of these two terms on the wage change.

3.3.1 Counterfactual Analysis 1: Constant Occupation Distribution Across Cities Since 1980

I first study the first perspective, which is the effects of evolving occupation distribution across different locations over time. The first counterfactual experiment is constructed by the assumption that the city size distribution of a given occupation would not have changed since 1980 which is the earliest year in the data set. The difference between the results of this counterfactual experiment and the real wage reflects the impact of the city size distribution changes across different geographic locations on wage for a given occupation. As shown in the previous subsection, Equation 3.2 models the marginal wage probability distribution by integrating on age, education and city size. In this nonparametric formulation, the probability distribution $P_{t,o}(b|a,e)$ represents the distribution of city size bin under the condition of age and education. By replacing this probability distribution, I propose a counterfactual distribution as:

$$h^1_{t,o}(w) = \int P_{t,o}(a,e)P_{1980,o}(b|a,e)P_{t,o}(w|b,a,e)dbdade \tag{3.4}$$

In Equation 3.4, the conditional probability of city size bin is always set to be the value in the year of 1980, which corresponds to the first year in the data set. Therefore, the expected wage of the counterfactual analysis by assuming the city size distribution for a given occupation o would have not changed is:

$$H_{t,o}^1 = \int h_{t,o}^1(w)wdw \tag{3.5}$$

In addition, the reduction $W_{t,o} - H_{t,o}^1$ demonstrates the effect of evolving city size for the given occupation o at time t. Although the city size distribution is not changed since 1980 based on the Equation 3.4, it does not restrict the changes of overall population. In other words, the city size still increases over time whereas the relative distribution across locations keeps stable.

3.3.2 Empirical Results of Counterfactual Experiment 1

In this section, reductions for different occupations between the real wage distribution and counterfactual experiment wage distribution are presented.

Year	$H_{t,o}^1$	$W_{t,o}$	$W_{t,o} - H_{t,o}^1$	Percentage
1980	-	48223.59	-	-
1990	56440.15	56827.05	386.90	0.68
2000	65897.95	66865.86	967.91	1.44
2010	68897.34	69902.92	1005.58	1.43

TABLE II

RESULTS ON COUNTERFACTUAL EXPERIMENT 1 FOR FINANCE INDUSTRY

Table II shows that the average wages of finance industry in counterfactual experiment for different time periods are lower than the real ones. The wage reductions in counterfactual experiment are due to the unchanged city size distribution. In other words, if the city size distribution keeps stable since 1980, the average wage of finance industry in the following decades would have increased less. Specifically, the wages of 1990, 2000 and 2010 would have been from 0.68% to 1.44% less than the actual wages due to the effect of evolving city size distribution that is likely caused by agglomeration economy.

Table III presents the comparisons between the counterfactual experiment and the ground truth for IT industry. Similar to the finance industry, the city size distribution change contributes to the increasing average wage for the IT industry. However, the impact of city size distribution change on the average wage for IT is less compared with that of finance related occupations. In 1990, the contribution of city size distribution change to the increasing average wage for IT is only 0.48%.

Year	$H^1_{t,o}$	$W_{t,o}$	$W_{t,o} - H^1_{t,o}$	Percentage
1980	-	36202.34	-	-
1990	40991.25	41189.19	197.94	0.48
2000	51119.24	51804.34	685.10	1.32
2010	54556.70	55706.59	1149.89	2.06

TABLE III

RESULTS ON COUNTERFACTUAL EXPERIMENT 1 FOR IT INDUSTRY

The comparison results between the counterfactual experiment and the real wages for waiters are presented in Table IV. The effect of city size distribution change on wage for waiters is larger than that of the IT industry. More importantly, the effect is not always positive. In 2000, the average wage in the counterfactual experiment is higher than the one in the real case, and hence the effect of city size distribution change on the average wage is negative. The negative value means that the average wage for waiters could have been higher in 2010 if the relative city size distribution keeps unchanged since 1980. In other words, the stable city size distribution lowers the average wage for waiters in 2000.

Year	$H^1_{t,o}$	$W_{t,o}$	$W_{t,o} - H^1_{t,o}$	Percentage
1980	-	8034.83	-	-
1990	9489.02	9534.80	45.78	0.48
2000	10348.06	10401.69	53.62	0.51
2010	9032.96	9170.50	137.54	1.49

TABLE IV

RESULTS ON COUNTERFACTUAL EXPERIMENT 1 FOR WAITERS

For the occupations whose average wages decrease with the city population, city size distribution change still plays the same role. Table V shows the effect of city size distribution change for assemblers. As seen in the table, the effect is not significant over time. From 2000 to 2010, city size distribution change pushes up the average wage for assemblers by 0.69%.

Year	$H^1_{t,o}$	$W_{t,o}$	$W_{t,o} - H^1_{t,o}$	Percentage
1980	-	22604.48	-	-
1990	22471.48	22518.62	47.13	0.21
2000	23509.93	23630.00	120.07	0.50
2010	20808.63	20953.36	144.73	0.69

TABLE V

RESULTS ON COUNTERFACTUAL EXPERIMENT 1 FOR ASSEMBLERS

3.3.3 Counterfactual Analysis 2: Fully Removing City Size Effects from Wages

In the first counterfactual experiment, I examined the effect of city size distribution changes on the average wage for different occupations. In order to further understand the relationship between city size and wages of different occupations, I construct the second counterfactual experiment to directly explore how the city size affects wages for different occupations.

As shown in Equation 3.4, the conditional city wage distribution $P_{t,o}(b|a,e)$ is replaced by $P_{1980,o}(b|a,e)$ such that the population distribution across cities for any given age and education in the following years are always the same as the one in 1980. Although Equation 3.4 properly models the change of city wage due to the migration of workers over time, the third term $P_{t,o}(w|b,a,e)$ that represents the conditional distribution of wage for labors locating in a city bin b with education e and age a is still affected by the city size. In order to fully understand the effect of city size on wages for different occupations, I construct the second counterfactual experiment by replacing $P_{t,o}(w|b,a,e)$ as well as the conditional probability $P_{t,o}(b|a,e)$ that I modified in the first counterfactual experiment.

Inspired by previous work Athey and Imbens (2002) on difference-in-difference models and Baum-Snow and Pavan (2012) on wage inequality analysis, I apply Changes-In-Changes (CIC) model to modifying the wage conditional probability distribution $P_{t,o}(w|b, a, e)$ because the CIC model is robust and effective on non-parametric modeling and transformation. The CIC model is adapted in such a way that a relationship is established between each conditional city wage percentile and a conditional rural wage percentile. Specifically, for each percentile of wages given a city bin b, education level e and age group a, I identify a closest percentile of wage in the rural areas under the same education and age conditions in 1980. Namely, a group of similar labors in the rural areas is linked with city workers belonging to each percentile of conditional wages. This type of relationship is maintained throughout the later years. For example, the 4th percentile of wages in a city size bin b, education level e and age group a corresponds to the 2nd percentile of wages in rural areas with education level e and age group a in the year of 1980. Then, a relationship between the 4th percentile of city wages based on b, e, a and the 2nd percentile of rural wages given e and a is established. For later years, the modified 4th percentile of city wages is derived from the 2nd percentile of rural wages for that given year. Similar relationship and modification are applied to all percentiles of city wages. To sum up, I use rural areas as the control group and the city wages across bins are treatment groups.

I denote the corresponding rural wage distribution as $R_{t,o}(w|b, a, e)$ after applying the CIC model as shown above. The second counterfactual experiment can be modeled as:

$$h^2_{t,o}(w) = \int P_{t,o}(a, e)P_{1980,o}(b|a, e)R_{t,o}(w|b, a, e)dbdade \tag{3.6}$$

From Equation 3.6, it is clear that the effect of city size is eliminated via two transformations:

- Replacing the city size distribution with the one in 1980;
- Replacing the wage distribution with the corresponding rural wage distribution derived from the CIC model.

Therefore, the effects of city size on wage introduced in Equation 3.2 are fully removed. Compared with the first counterfactual analysis that only assumes the city size distribution would not have changed since 1980, the formulation in Equation 3.6 considers the second term and third term in Equation 3.2 that are both related to city size and can influence the wage. The difference between Equation 3.2 and Equation 3.6 demonstrates the full effect of city size on wages.

3.3.4 Empirical Results of Counterfactual Experiment 2

Since Equation 3.6 provides a more complete theoretical foundation on modeling the effect of city size compared with the equation used in the first counterfactual example, it is interesting to apply Equation 3.6 to real world data and examine the differences between the two established counterfactual experiments.

Similar to the previous analysis, the mean wage by applying counterfactual experiement 2 that fully captures the effect of city size is presented as:

$$H_{t,o}^{2} = \int h_{t,o}^{2}(w) w dw \tag{3.7}$$

Table VI shows the results under the second counterfactual experiment in different years for the workers in the finance industry. Instead of having slight differences (up to 1.44%) as shown in the first counterfactually experiment for finance, the percentage values across all three years are significant, namely 13.02%, 22.07% and 18.79% for the years of 1990, 2000 and 2010, respectively. This phenomenon indicates that around one quarter of the wage is rooted from the effects of city size for workers in finance. The result in this table is fundamentally different from the result in the previous section using counterfactual experiment 1. This is due to the fact that the second counterfactual experiment can fully eliminate the effects of city size by establishing relationships between city and corresponding rural areas, whereas the first counterfactual experiment only considers the effect of city size distribution changes.

Year	$H^2_{t,o}$	$W_{t,o}$	$W_{t,o} - H^2_{t,o}$	Percentage
1980	-	48223.59	-	-
1990	49407.91	56809.01	7401.10	13.02
2000	52096.94	66847.95	14751.01	22.07
2010	56758.46	69894.61	13136.15	18.79

TABLE VI

RESULTS ON COUNTERFACTUAL EXPERIMENT 2 FOR FINANCE

Table VII presents the effect of the city size on the wage in the IT industry. The overall city wages would have been 8.75%, 20.92% and 20.90% lower if there were no effect of city size in 1990, 2000 and 2010, respectively. This result is also much more significant than that of the first counterfactual experiment, although the absolute percentage differences for the IT industry is relatively lower than the values of the finance industry. The city size factor contributes more and more to the wage from 1990 to 2010, growing from less than 10% to almost one quarter. This demonstrates that city size has more impact on the wage for high-skilled occupations such as IT and finance in recent years as cities become more developed over time.

Year	$H^2_{t,o}$	$W_{t,o}$	$W_{t,o} - H^2_{t,o}$	Percentage
1980	-	36202.34	-	-
1990	37575.04	41176.65	3601.61	8.75
2000	40947.32	51779.28	10831.95	20.92
2010	44044.17	55681.56	11637.38	20.90

TABLE VII

RESULTS ON COUNTERFACTUAL EXPERIMENT 2 FOR IT

After applying the second counterfactual example model to two skill intensive occupations, I further calculate the wage difference for a popular service oriented occupation, waiter/waitress, and present the results in Table VIII . It is clear that waiters earn higher wages at bigger cities and the effects of city size on wages are between 16.84% and 22.46% for the three decades after the year of 1980. In other words, for a typical service oriented occupation such as waiter, the effect of city size takes up one fifth of the overall wage, namely the wage would have been around 20% lower if the impact of city size is fully removed. Furthermore, the pattern on waiters is consistent with the findings on skill intensive occupations including finance and IT. This further demonstrates that the wages of both skill intensive occupations and service oriented occupations are largely affected by the city size factor and agglomeration economy plays an important role for these occupations.

Year	$H_{t,o}^2$	$W_{t,o}$	$W_{t,o} - H_{t,o}^2$	Percentage
1980	-	8034.83	-	-
1990	7956.78	9531.34	1574.55	16.52
2000	8647.60	10399.29	1751.68	16.84
2010	7109.27	9168.11	2058.84	22.46

TABLE VIII

RESULTS ON COUNTERFACTUAL EXPERIMENT 2 FOR WAITERS

The difference about the effect of city size on the occupation assembler is shown in the following table IX. In sharp contrast to previous results, city size does not have significant effects on the wages of the assemblers, only 5.32% in 2000. In the years of 1990 and 2010, the wages of the assemblers are even negatively affected by the city size, which indicates the wages of assemblers in cities would have been higher if there was no city size effect. The main reason is assembler is one of the typical traditional occupations and it shows very different patterns with skill intensive occupations and service oriented occupations. Based on agglomeration economy theories, traditional occupations tend to use more lands and more likely to cluster and concentrate in small cities and rural areas. Our model clearly quantifies the impact of city size on wages for assemblers that is negligible if not negative compared with service oriented occupations or skill intensive occupations, e.g., waiters, IT workers and finance workers.

Year	$H^2_{t,o}$	$W_{t,o}$	$W_{t,o} - H^2_{t,o}$	Percentage
1980	-	22604.48	-	-
1990	21185.495	22512.45	1326.95	5.89
2000	22361.99	23618.83	1256.84	5.32
2010	21434.26	20942.57	-491.69	-2.34

TABLE IX

RESULTS ON COUNTERFACTUAL EXPERIMENT 2 FOR ASSEMBLERS

3.4 Understanding Wage Gaps of Large Cities

In the previous section, we quantitatively evaluated the effects of city size on wages for various occupations. Larger cities may or may not put a wage premium on workers depending on their occupations. As various occupations reveal significantly different wage gaps introduced by city size, this further proves that traditional approaches that investigate effects of wage and city size by treating all occupations as a whole is not ideal. In addition, there are many factors of occupations that could lead to the wage difference divergence affected by city sizes. Since the wage differences have been quantitatively measured in the last section, it makes sense to identify the key components related to occupations that are the result of city size effects on wages for various occupations.

I extract a number of characteristics of occupations and construct a regression model to evaluate the wage gaps introduced by city size effects:

$$\begin{aligned}\Delta w_o = \beta_0 + \beta_1 \times age_o + \beta_2 \times gender_o + \beta_3 \times edu_o + \beta_4 \times industry_o \\ + \beta_5 \times prestige_o + \beta_6 \times dummy_o + \epsilon_o\end{aligned} \tag{3.8}$$

where Δw_o is the wage gaps introduced by city size for occupation o, namely the difference between $H^2_{t,o}$ in Equation 3.7 and the real wage value; β_0 is the intercept of the regression model; age_o, $gender_o$, edu_o, $industry_o$ and $prestige_o$ are the aggregated value of the individual age, gender, education, industry and prestige for occupation o, respectively; $dummy_o$ is a dummy variable derived from occupation o. β_1 to β_6 correspond to these terms coefficients; and ϵ_o represents the error term.

Specifically, we use the average function to aggregate the values for terms including age, gender, education and prestige. For the factor of industry, we calculate how many industries have the given occupation o. It is worth to note that the prestige term represents the social standing of a given occupation o. I use a widely used measure Nam-Powers-Boyd Occupational Status Score (NPBOSS) (Nam and Boyd (2004)) as the prestige term. It represents the occupational status and occupational standing.

In addition to study the impact of agglomeration economy on occupation, I create a binary indicator variable that represents if a specific occupation is attracted and benefited by agglomeration economy. The 1990 occupational schemes from IPUMS (Ruggles et al. (2015)) contains 389 fine-grained categories and they are grouped into 7 aggregated meta-categories as shown in Table XI. The dummy variable $dummy_o$ is 1 if the occupation o belongs to meta-categories either *Managerial and Profes-*

```
==============================================================================
Dep. Variable:                      Y   R-squared:                       0.476
Model:                            OLS   Adj. R-squared:                  0.399
Method:                 Least Squares   F-statistic:                     6.211
Prob (F-statistic):          0.000106   Log-Likelihood:                -465.49
AIC:                            945.0   BIC:                             958.1
Df Model:                           6

==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept  -1.245e+04   7139.270     -1.744      0.089     -2.69e+04  1968.216
age           65.0088    162.631      0.400      0.691      -263.430   393.448
gender      -385.9350   2677.451     -0.144      0.886     -5793.157  5021.287
edu          967.2983    521.610      1.854      0.007       -86.115  2020.711
industry       9.6161      9.927      0.969      0.338       -10.432    29.664
prestige       8.4560      5.171      1.635      0.110        -1.987    18.900
dummy       5439.1923   1895.497      2.870      0.018      1611.159  9267.225
==============================================================================
```

Figure 7. Regression result on regression model in Equation 3.8

sional or *Technical, Sales, and Administrative*, and 0 otherwise. Therefore, the derived dummy variable indicates if the occupation is strongly affected by agglomeration economies.

The regression model in Equation 3.8 allows for an analysis on wage difference introduced by city size for different occupations. The key characteristics of occupations including overall age, gender, education levels, industry coverage, social standings and agglomeration economy effects are all presented in the model. I use the data of the year 2010 that represents the latest city effects on wages and calculate the wage difference Δw_o. The values of Δw_o for the top 50 occupations in the United States measured by popularity are shown in Table X.

Figure 7 shows the result of fitting by using the regression model in Equation 3.8. A R^2 value at 0.476 indicates that about half of the variability with regard to wage gap can be explained by the proposed regression model. The coefficients of all terms except for gender are positive, which means these attributes have positive impacts on wage difference. Furthermore, among six factors associated with occupation, four attributes are not statistically significant since their p-values are much greater than the common alpha level of 0.05 (0.691 for age, 0.886 for gender, 0.338 for industry and 0.110 for prestige). Interestingly, education is the most significant attribute that is statistically significant with regard to wage gaps introduced by city sizes, since its p-value is 0.007. The result clearly indicates that requiring higher education levels and potentially needing more skills embedded in occupations are associated with higher wage premiums in larger cities. It is also clear that the dummy variable representing agglomeration economy effects is also statistically significant with a p-value of 0.018. This further demonstrates that different agglomeration economy effects on occupations has a direct impact on the dependent variable Δw_o. In the meantime, factors including age, gender, social standing and industry coverage are not significant with regard to wage gaps introduced by city size.

The conclusion drawn from the regression result is consistent with the descriptive analysis carried out in Section 5 based on theories of agglomeration economy. Job skill is a critical component on determining wage gaps, and larger cities are more attractive to workers with higher level of educations and skills. Furthermore, the nature of involved work for different occupations may or may not show the phenomenon of agglomeration economy. This is also closely related to the wage gaps led by city size. Therefore, skill intensive occupations tend to cluster and centralize in large cities, and the wage premium rooted from city size is significantly higher.

3.5 Conclusions and Contributions

I use nonparametric methods to set up two counterfactual experiments in order to analyze the effect of city size on occupation wages from two perspectives. The first counterfactual experiment focuses on the effects of changes in the city size distribution of occupations. Based on the results, I claim that the impact of city size distribution changes over years is trivial with up to 2%. In the second counterfactual experiment, I construct a probabilistic model that captures the effects of city size distribution change and the wage distribution conditioned on city size. To remove the effect of city size on wage distribution, I propose a model that establishes the links between each city and its corresponding rural area. Such link relationship is utilized to treat rural areas as control group such that I can quantify the city size effect on wage distribution. The results show that the impact of city size on wage has grown substantially for the occupations with agglomeration economies, for example, skill intensive occupations and knowledge-based occupations.

While the effect of city size on wages for individual occupation has been measured, I further study what attributes from occupations lead to the wage gaps introduced by city size. Through a regression model, I confirm that education and agglomeration effects on occupation are the two factors that are statistically significant and have positive effects on wage premium. I conclude that high education background and skill intensive occupations have stronger tendency to follow agglomeration economy, thus are more likely to carry high wage premiums while working in big cities.

Occupation	Δw_o	Occupation	Δw_o
Managers and administrators, n.e.c.	9905.40	Salespersons, n.e.c.	13746.86
Secretaries	4608.23	Janitors	-959.16
Truck, delivery, and tractor drivers	-3159.35	Cashiers	124.60
General office clerks	2614.16	Primary school teachers	4540.74
Cooks, variously defined	1887.84	Machine operators, n.e.c.	-837.56
Nursing aides, orderlies, and attendants	30.90	Waiter/waitress	2831.28
Bookkeepers and accounting and auditing clerks	5456.74	Assemblers of electrical equipment	-1222.41
Registered nurses	19258.94	Laborers outside construction	-2335.77
Military	9533.21	Production supervisors or foremen	1095.94
Supervisors and proprietors of sales jobs	3290.16	Accountants and auditors	12017.02
Housekeepers, maids, butlers, stewards, and lodging quarters cleaners	226.10	Office supervisors	2984.37
Typists	4851.58	Construction laborers	661.47
Guards, watchmen, doorkeepers	1319.85	Textile sewing machine operators	1887.98
Misc food prep workers	1676.05	Subject instructors (HS/college)	-2302.79
Managers and specialists in marketing, advertising, and public relations	10359.26	Stock and inventory clerks	-3506.66
Child care workers	2197.55	Receptionists	2760.02
Shipping and receiving clerks	-3972.75	Carpenters	-868.86
Social workers	2303.24	Lawyers	28475.70
Automobile mechanics	-1468.17	Packers and packagers by hand	-1443.83
Secondary school teachers	4454.34	Customer service reps, investigators and adjusters, except insurance	-3347.84
Production checkers and inspectors	2.00	Retail sales clerks	3798.90
Administrative support jobs, n.e.c.	1932.93	Welders and metal cutters	-2135.76
Stock handlers	7.41	Data entry keyers	605.63
Teachers , n.e.c.	-1988.38	Police, detectives, and private investigators	6141.46
Physicians	6867.26	Personnel, HR, training, and labor relations specialists	7406.59

TABLE X

VALUES OF WAGE GAPS

Occupational Meta-category	Occupation Code Range
Managerial and Professional	000 - 200
Technical, Sales, and Administrative	201 - 400
Service	401 - 470
Farming, Forestry, and Fishing	471 - 500
Precision Production, Craft, and Repairers	501 - 700
Operatives and Laborers	701 - 900
Non-occupational responses	900 - 999

TABLE XI

OCCUPATIONAL META-CATEGORIES

CHAPTER 4

CITY SIZE AND MIGRATION

4.1 Introduction

Migration is an interesting topic to study since it is important to understand the economic trend on population change, worker movement and regional government policy. Many researchers have studied migration related to regional economies from a number of aspects. Long (1973) examined the relationship between the education and migration pattern. Walker et al. (1992) investigated how internal labor migration is affected by immigration in the United States. Rappaport (2007) claimed that most U.S. residents choose areas with nice weather as their migration destinations. However, most research either did not consider the role of city size or only focused on the aspect of education. In reality, city size is an important direction to study migration with regard to local economy. Some natural questions such as ”Are people more likely to move to large cities?” and ”Why some workers tend to migrate to denser areas whereas some prefer smaller cities?” are critical to understand the fundamentals of migration.

In the meantime, different people make various decisions on migration. The nature of work, or more precisely the occupation engaged in, highly affects migration patterns. Workers engaged in different occupations with various backgrounds may have divergent patterns regarding to migration. Thus, to have a comprehensive understanding on migration, it is necessary to analyze migration by considering both city size and engaged occupation.

In this study, I obtain individual level data on metropolitan areas including detailed information on migration for seven years between 2005 and 2011 from Integrated Public Use Microdata Series (IPUMS). Based on the data, I first conduct the regression analysis on city population and number of in-migrants across years. The results show that city with bigger population has higher number of in-migrants, and the regression coefficient is equal to 1. This indicates larger metropolitan areas attract more in-migrants, but the proportion of in-migrations in the metropolitan areas is consistent regardless of the metropolitan population. This finding is valid in the sampled seven years from 2005 to 2011.

In order to further study the relationship between city size and migration, including the effects of city size on different category and type of migration, I analyze the problem by occupations. According to the agglomeration economy theory, some occupations have need to spatially cluster together in large areas to take advantage of benefits from agglomeration economy, but some don't. Therefore, individual's occupation plays an important role in making their migration choice, such as moving to larger areas or staying in smaller cities. It is critical to consider the role of occupation while investigating the effects of city size on migration. The value of coefficient α in the regression model of city size and migration by occupation indicates the sensitivity and preference of an occupation to the big cities. If α is greater than 1, large cities are more attractive for individuals with this occupation and these people tend to migrate to large cities.

Since the coefficient α differs largely across occupations, it is crucial to understand what factors lead to the divergence on migration, especially with regard to the tendency of moving to large metropolitan areas. I propose a novel regression model that studies the relationship between the tendency of migration α and a number of core attributes associated with occupation, including age, education, industry coverage, occupational prestige, wage, average residential population and unemployment rate. I demonstrate that education level, industry coverage and average residential population are statistically significant on migrating to large cities.

The following content is organized as follows. Section 4.2 is the literature review; Section 4.3 presents the data that I used in this study; Regression analysis between city size and migration across years and further analysis involving in occupations are discussed in Section 4.4; Section 4.5 demonstrates what factors associated with occupations characterize the migration patterns; Section 4.6 is the conclusion.

4.2 Literature Review

Bowles (1970) argued that migration of workers is actually an investment problem. The individuals consideration about the costs and benefits of moving explains the pattern of the migration. Individuals who gain the most benefits from moving would cause higher net out-migration rates. Based on these results, the author further concluded that the derivative of the net migration rate with regards to the present value of the income increase from migration has positive relationship with the years of schooling. In other words, the effect of income increase on the probability of moving is enhanced by the level of education, because more educated people easily adapt to economic disequilibria. In addition, the author pointed out that whites are more sensitive to the income gain from migration than blacks.

Long (1973) examined the relationship between the education and migration pattern. It was found that education is an important factor for the long-distance migration but it cannot well predict the short-distance moving. The results suggested that the frequency of migration between states for the men with graduate degree in the age group of 25-29 during a year is as three times as the chance of moving for the men who did not finish high school. The author further argued that education plays a more important role in migration than occupation. When controlling for the age and occupation, the effect of education on migration is still positive. The finding also indicated that the characteristics of individuals better determine the migration process compared with the natures of occupations. However, when predicting the migration pattern, education has limitations on determining who migrates in the country, since the number of persons going to the college increases and college graduates group has less distinctive migration trends.

Borjas et al. (1992) used the self-selection model to analyze the internal migration flows based on data from the National Longitudinal Surveys of Youth. The empirical results indicated that differences in the returns to skills among states largely determine the size, direction and skill composition of the internal migration process. Skilled workers migrate to the places which offer high rewards for skills whereas unskilled individuals move to the areas with low skill pays. In other words, individuals are most likely to migrate when their skills are mismatched with the reward system in their current regions, and they are expected to move to the regions where the skill prices are more compatible with their levels of skill. Moreover, the direction of internal migration flows is determined by comparative advantage of the region. If a region offers skilled workers a greater wage dispersion than their current place, workers will have incentives to migrate to there.

Walker et al. (1992) developed a mobility model to investigate how immigration affects internal labor migration in the United States. The model incorporates occupation of worker, production and economic restructuring in the analysis of immigration and internal movements. For the local blue-collar workers, their jobs have been spatially replaced by immigration. For the white-collar workers, their choices about migration destination are affected by the process of capital accumulation and economic restructuring in the mobility system. All these factors also influence the immigration pattern. The author then concluded that previous estimates of immigration effects on the native labor market are underestimated when measured at the SMSA level, because they ignore the competition between the native occupational classes and immigrants.

Rosenbloom and Sundstrom (2004) used data from the Integrated Public Use Microdata Series of the U.S. Census (IPUMS) to describe and examine trends in internal migration over the past 150 years. The authors proposed two alternative measures to calculate the interstate migration. The first measure is based on individual's state of birth. The second one considers families with young children and treat it to move if the current state different from the birth state of one of the family's young kids. The authors claimed that the second measure provides more accurate estimation on the timing of interstate moves compared with the first one. However, both measures suggest that the trend of migration propensities is U-shaped since 1850. It is falling until year 1900 and gradually going up over much of the 20th century. In addition, the authors investigate variation in the migration propensity by several factors, such as age, sex, education, etc. They argued that rising educational attainment is the main reason to increase the internal migration after 1900.

Quinn and Rubb (2005) claimed that education-occupation matching is an important factor which influences the migration decision. Based on the data from Mexico, they incorporated education-occupation matching into the migration decision to set up the model. The results from the model indicated that individuals with higher education level than required by occupations have more incentives to migrate, while positively mismatched between education and occupation workers tend to stay. They further suggest that positively mismatched workers are more likely to migrate domestically rather than internationally, when migration is given. In addition, the authors pointed out that the relationship between education and migration can be positive or negative since education-occupation mismatches at different education levels are different.

Rappaport (2007) concluded that areas with nice weather are the migration destinations chosen by most U.S. residents. The possible explanations are provided as air conditioning, change in the business composition of employment, elderly migration and income increase. However, the regression results of population growth on weather indicated that the drive of these weather-related migrations is that people treat nice weather as a consumption amenity and its valuation is increasing. The reason of the increased valuation of consumption amenities is that per capita income is rising. Therefore, with the per capita income increasing, the places with nice weather will continue attracting more migrations.

Chen and Rosenthal (2008) investigated the intention and drive of household migration and the potential of which cities and regions might thrive in the near future. They used the US census data from year 1970 to 2000 to examine the individual household migration decisions. They found that households prefer to move to the warm coastal places and non-metropolitan areas whereas firms tend to choose core or large cities. In addition, they further pointed out that highly educated households migrate toward areas with better business environments, but people after 50 years old especially retirees choose places with higher quality consumer amenities to migrate. Hence, locations with attractive consumer amenities or high valued business environments are likely to thrive.

O'Hagan and others (2014) examined in-migrants and return migrants to Canadas Census Metropolitan Areas and Census Agglomerations and compared these migration patterns of small, medium and large cities in Northern Ontario. The author claimed that small cities mainly attract intra-regional in-migrants and large cities are chosen by inter-provincial in-migrants and immigrants. For return migrants, the results showed that large cities are more attractive to inter-regional return migrants than small cities. Compared with intra-regional return migrants, inter-regional return migrants have less chances to go back to large cities. The paper also reveals large cities have comparative advantages to attract inter-regional migrants with higher skills and education compared to small cities, however, these advantages are not significant between large cities and medium sized cities. This paper gives a clear study on the significance of medium sized cities on appealing to inter-regional migrations with high skills and education compared to their rival large cities in the current knowledge economy.

4.3 Data Description

4.3.1 Problem Statement

The research on city size has attracted interests of economic researchers for a long period of time. City sizes could vary drastically, from tiny cities to very large cities with millions population. Many factors such as location, industry, weather, government policy and history can all affect the size of a city. Arnott (1979) studied the optimal city size under a spatial economy setting by considering social welfare, utility, space and transportation costs and government intervention. In the meantime, city size is dynamic and changing over time, and urban growth leads to the development of city size distribution in almost every economic and social systems. Gabaix and Ioannides (2004) reviewed a number of theories including Zipf's law, Gibrat's Law for means and Gibrat's law for variances related to urban growth and urban evolution. However, there is little research on establishing a clear relationship between city size and migration. Naturally, the movement of people from one place to another creates the urban evolution and city size dynamics, usually when people are changing permanent or semi-permanent residence.

Such migration of human behavior can either be chosen by people as voluntarily or be forced to move (involuntary migration). Many factors may influence people's decisions, including environments, political factors, global and regional economy and cultural effects. There are multiple ways on measuring migration:

- in-migration: people moving into one place from another
- out-migration: people moving out of one place to another
- gross migration: total number of in-migrants and out-migrants
- net migration: the difference between in-migration and out-migration

This study investigates the relationship between city size and migration from an economic point of view. Thus, I will focus on in-migration of urban cities in this paper. Mueser and White (1989) has shown that there is a positive association between in-migration and out-migration across locations. Such strong correlation means the findings and observations will not change significantly even if another migration measurement were to be used.

In this study, I collect individual level data on metropolitan areas with detailed information on migration status. Then I build a quantitive regression model between in-migration and city size across years, and uncover if there is a linear relation between these two measures which is consistent across a number of years. This will help demonstrate if larger cities are more attractive to migrants compared with medium and small cities given the advantages in larger cities. As migrants make moving decisions on many factors and occupation is one of the most important ones, I further conduct a comprehensive analysis to discover the effects of city size on migration for top 50 occupations in the United States. With the help of the theories of agglomeration economy, I will explain the distinct patterns and impacts of city size on migration for various occupations. Furthermore, I give an empirical study on which characteristics associated with occupation are significant given the nature of migration.

4.3.2 Data

I obtained seven years of individual level data from the year of 2005 to 2011 from Integrated Public Use Microdata Series (IPUMS). Only the individuals from metropolitan areas are selected for analysis in this paper. The data is extracted from American Community Survey (ACS) in corresponding years with a density at 1.0%.

As I use in-migration as the measurement on migration, the sampled individual can be an in-migrant if this person either moved between states or abroad one year ago to a specific metropolitan area. Thus, each sampled individual can be accurately classified as in-migrant or non-in-migrant in this study. A number of other factors related to city and occupation are also extracted. One should note that the Census Bureau reorganized the occupational classification system across different years, and the occupation definitions change with each census year. To make the obtained data set consistent and accurate with a robust occupational classification system, I use the 1990 Census Bureau occupational classification scheme as the main classification system despite the original meaning of occupational codes changes with each census year. All data samples are updated with the occupational information from the 1990 classification scheme to ensure the data quality of the following analysis and study.

4.4 Regression Analysis on Migration and City Size

4.4.1 Migration and City Size across Years

Regional economy determines a lot of factors associated with local characteristics, including environment, population, industries, education, welfare, etc. In addition, regional economy is highly dynamic and may change over time dramatically given the changes from both inside and outside. One critical yet not fully studied aspect is in-migration related to regional and local economy. Many attributes affect people's decisions to stay, leave and move back while the local city condition can have large impacts on the decisions. One common belief is that larger cities have more in-migrations than smaller cities, whereas it's not clear if this is because larger cities are more attractive to in-migrations or larger cities simply have more population that lead to more in-migrations in quantitative terms. Furthermore, the purposes of migrations are diverse and previous research in migration has shown that cities with different sizes attract workers with various backgrounds and education. Thus, it is interesting to investigate the relationship between in-migrations and city population and examine if such relationship is consistent across different years.

I use the following Equation 4.1 to conduct the regression analysis on the number of in-migrants and city population

$$log(MIG) = \beta + \alpha log(POP) \tag{4.1}$$

where MIG represents the number of people moving into the city, POP stands for the city population, and α is the regression coefficient.

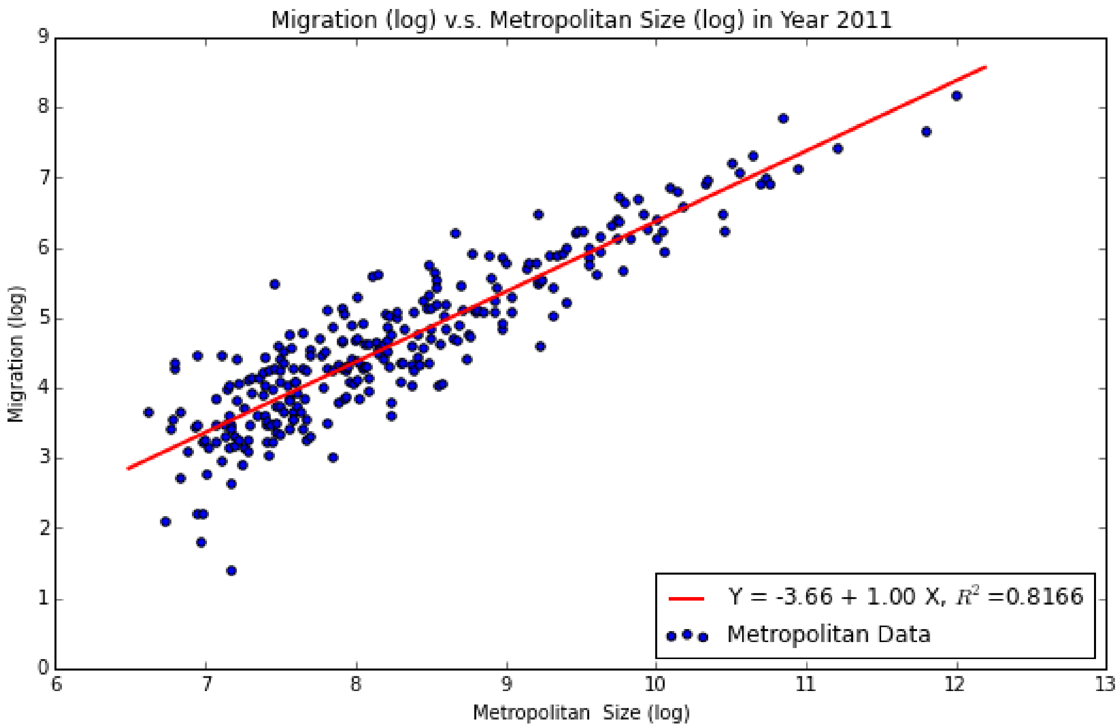

Figure 8. Regression result between in-migration and city size in 2011

Figure 8 shows the regression result in year 2011. The scattered blue points represent metropolitans with their corresponding population and in-migration size. One should note that I use log scale on both x axis and y axis. The red straight line is the fitted result from the above regression model. Consistent with the previous hypothesis, in-migration and city size demonstrate a positive relation. Not surprisingly, the larger region attracts more people to move in. As the coefficient α is equal to 1, one unit increase in log metropolitan population leads to one unit increase in the log of in-migration. The regression model fitted the underlying data relatively well with a R^2 at 0.8166.

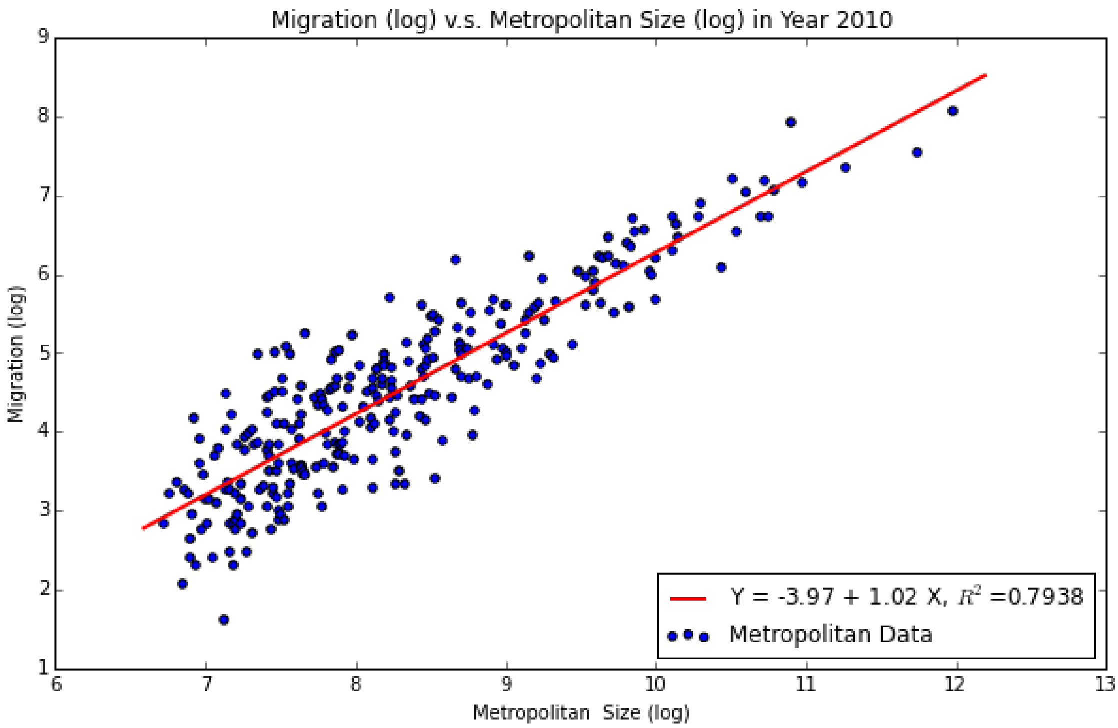

Figure 9. Regression result between in-migration and city size in 2010

This positive relationship between in-migration and city population with coefficient almost equal to 1 has been held for past few years. The following figures respectively demonstrate how the city population affects the number of in-migrants from year 2005 to year 2010. As seen in these figures, the coefficient α is nearly 1 and R^2 value is equal to 0.8 or so indicating the data is well described by the model.

It is worth noticing that the regression results in these 7 years all show that the coefficient α is positive, namely larger metropolitan population indicates more in-migrations. More specifically, the coefficient being 1 represents the proportion of in-migrations in the metropolitan areas is consistent

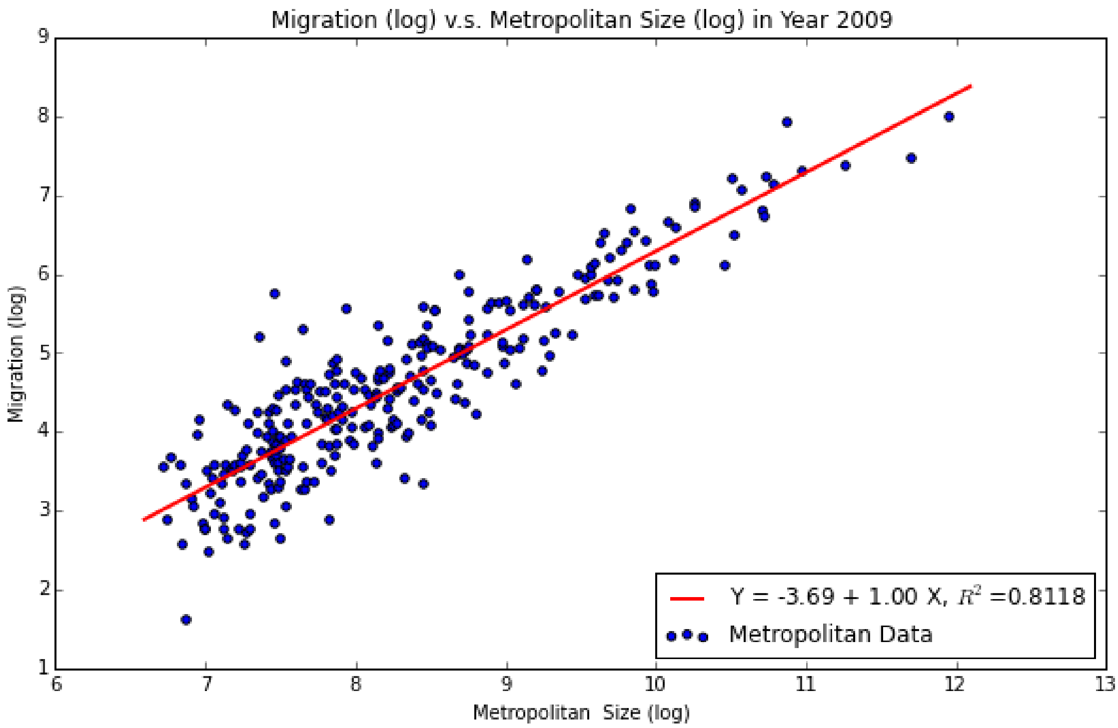

Figure 10. Regression result between in-migration and city size in 2009

regardless of the size of the metropolitan area, although the absolute number of in-migrants is bigger in larger metropolitan area. In other words, more in-migrations are found in larger cities but this does not mean larger areas have higher ability of attracting in-migrants compared with smaller cities.

4.4.2 Migration and City Size by Occupations

In the last section, I examined the effect of city population on in-migration and found that large city size does not necessarily indicate higher attractions for migrants, although the number of in-migrants is bigger in larger cities compared to small cities. However, this conclusion about the relationship between city size and migration does not specify the category and type of migration. According to previous

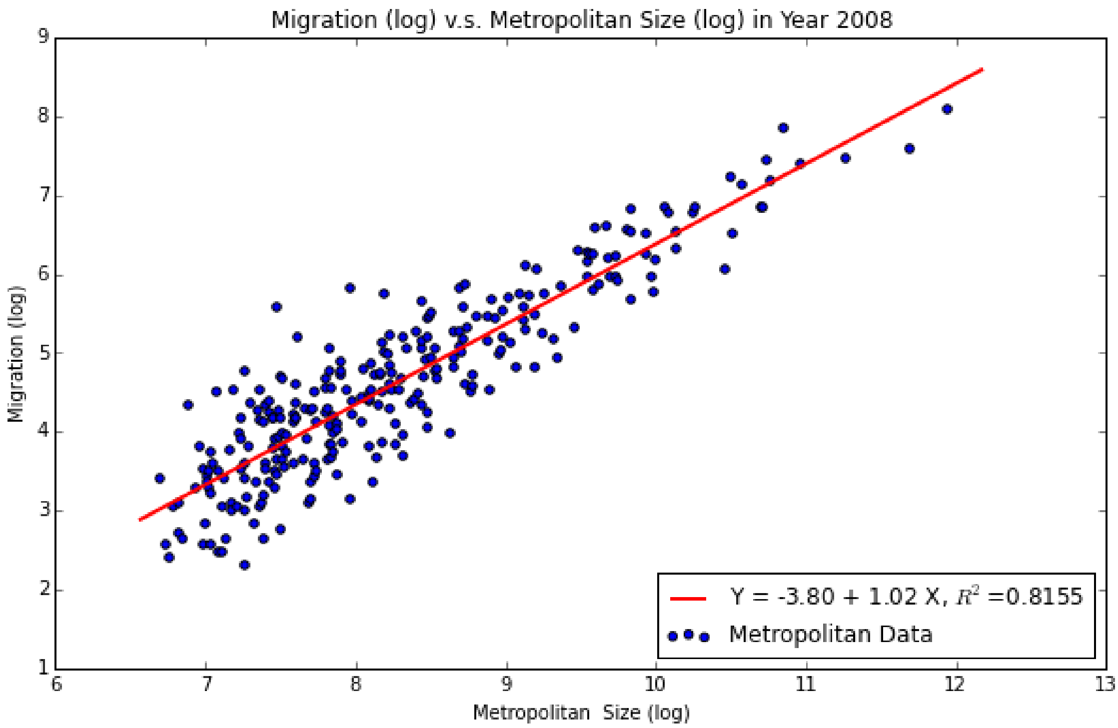

Figure 11. Regression result between in-migration and city size in 2008

research Chiswick and Miller (1994), Khan (1997), Rao (2010), Dustmann et al. (2011), individuals with better education background and higher level of skills tend to migrate to big cities. In this paper, I propose that individuals' occupations largely determine their migration choices - staying in smaller cities or moving to larger areas.

According to the theory of agglomeration economy, some occupations need to be spatially clustered together in big regions in order to take advantage of benefits, such as more efficient communication, better information share and so on. But for some other occupations, it is better not to be located in large cities. While in more competitive environment from big cities, these occupations are not given better

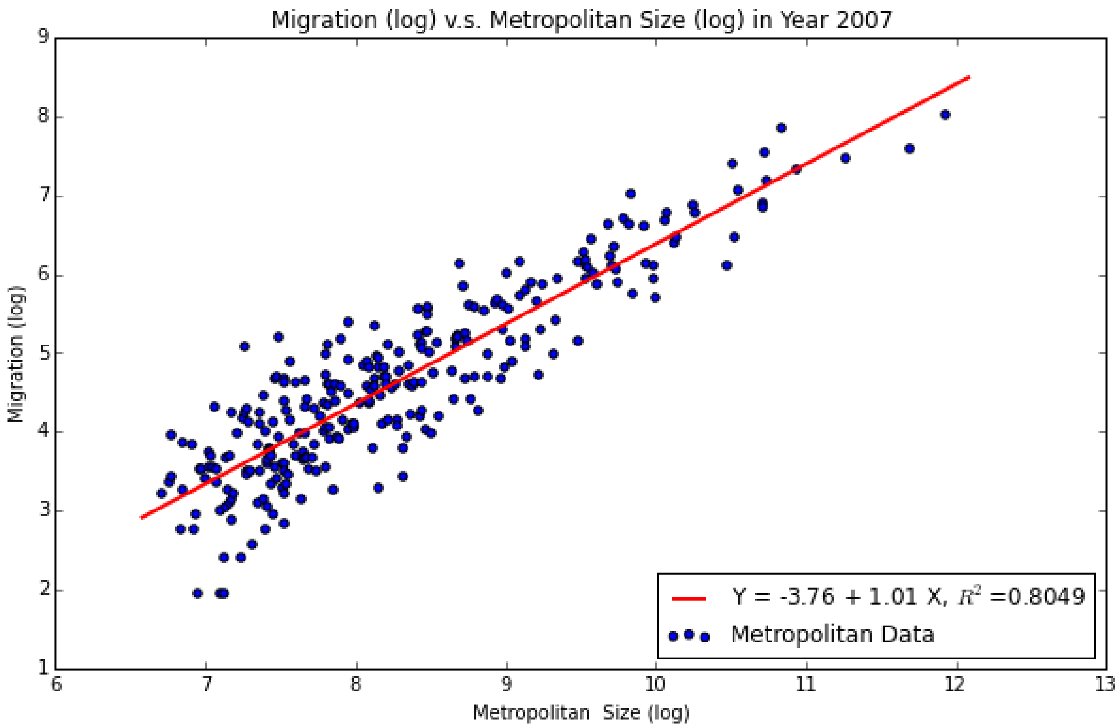

Figure 12. Regression result between in-migration and city size in 2007

opportunities and get highly paid. Individuals with these occupations will not choose to migrate to big areas. In the meantime, the spatial constraint of the nature of the work may also push the jobs away from big cities. Analyzing migration by treating all occupations as a whole overlooks a lot of valuable information and its conclusion might be incomplete or even misleading. Therefore, it is interesting and critical to investigate what occupations are more easily to be attracted to large cities and what occupations would rather stay.

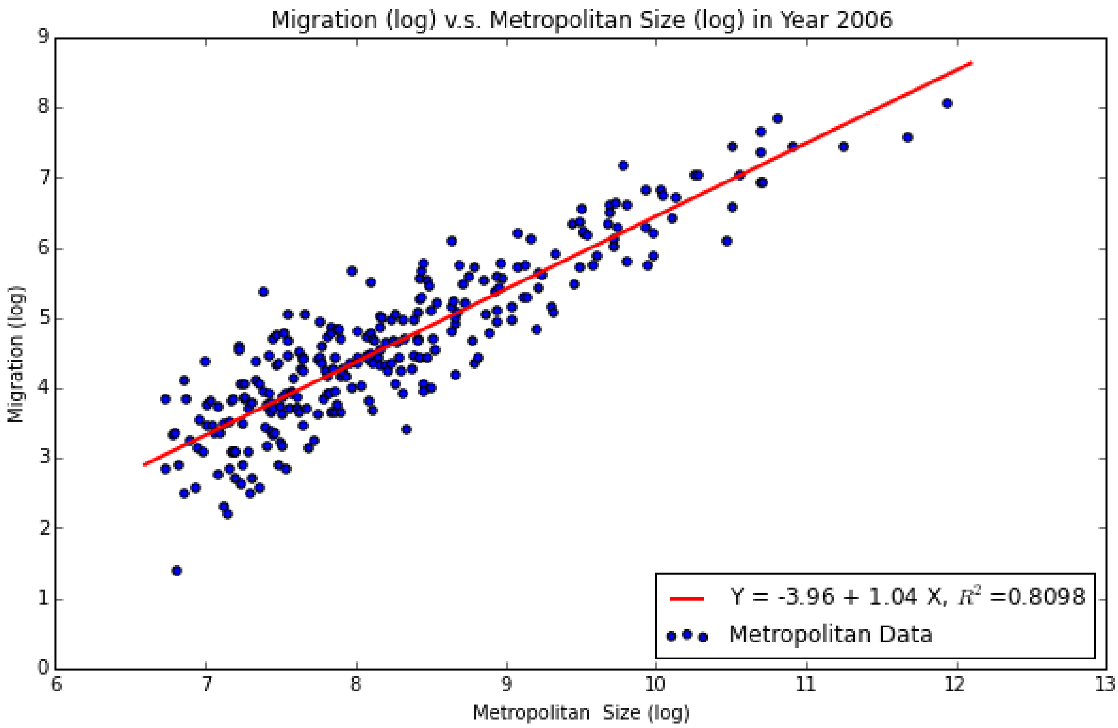

Figure 13. Regression result between in-migration and city size in 2006

4.4.2.1 Occupations Attracted to Large Cities

Based on Equation 4.1, I apply the regression model to the sampled data for each occupation and study the patterns associated with top 50 different occupations specifically. To better present the data, metroplitans are aggregated based on its size into 6 buckets, correponding to the log metropolitan size from 6 to 11. The average log migration poulation for each metropolitan size bucket is calculated as the dependent variable in the regression model. From previous observations, the overall coefficient α is equal to 1 across years, which demonstrates the migration is not strongly correlated with metropolitan population without considering occupation specific groups. In this subsection, running an empirical

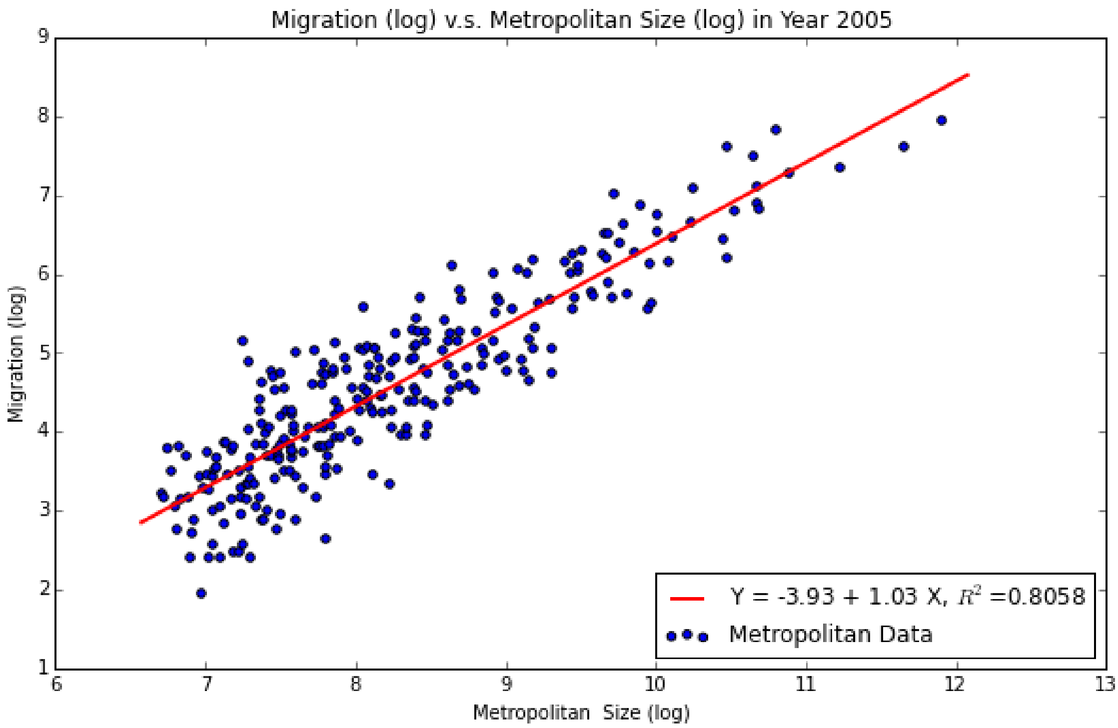

Figure 14. Regression result between in-migration and city size in 2005

study for each occupation and their regression analysis results will reveal the association between migration and occupation. In the previous regression model, the value of α suggests the sensitivity and preference of an occupation to the big cities. While α is greater than 1, large cities have bigger power to attract the individuals with this occupation to move in. In other words, smaller cities are less likely to attract in-migrants under such condition since their in-migration rates are less than those of big cities.

Figure 15 demonstrates the regression result between metropolitan population and migration for occupation *financial managers* in year 2010. As an occupation in the finance industry, *financial managers* have a positive relationship between metropolitan size and migration with $\alpha = 1.18$. This means that the large city size is very attracting for this occupation and financial managers have preference to migrate to bigger cities. In other words, compared with small cities, larger cities attract more financial managers.

This result is consistent with the theory of agglomeration economy. As a typical agglomeration economy, occupations in this industry need to move to big areas to cluster together in order to take advantage of all benefits from the agglomeration economy. Large cities provide more job opportunities to individuals with these occupations, and individuals also have better environment to get and share newest information. Assistances and supports from relevant occupations are easier to be obtained in large cities.

Similar to the occupations in finance industry, occupations in *marketing, advertising and public relations* also show the positive relationship between city size and migration with $\alpha > 1$. The result is shown in figure 16. Compared to small areas, large cities can provide enough spaces for these relevant occupations to better develop. Other occupations which demonstrate this pattern include *accountants and auditors* (figure 17), *salespersons* (figure 18), etc.

The other typical agglomeration economy is in high technology industry. The figure 19 demonstrates the pattern on migration and city population for *computer software developers*. As seen in the figure, one unit increase in log city population would lead to 1.24 unit increase in log number of in-migrants in metropolitan cities. Like the occupations in finance, computer software developer migrants have more probabilities on moving to large metropolitan cities compared with the patterns of overall migration. This is due to the benefits from the agglomeration economy that attract individuals with these occupations to large cities and cluster the professionals of these high-skilled occupations with higher density at big metropolitan areas.

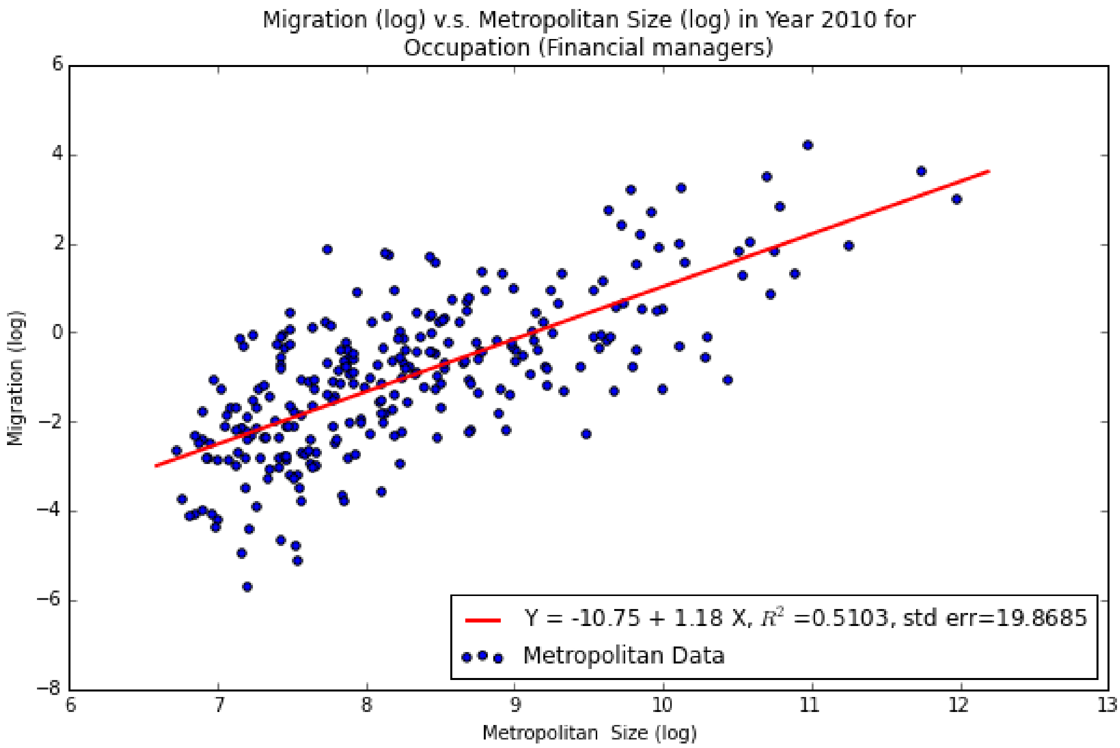

Figure 15. Regression result between in-migration and city size for Financial managers in 2010

Some other high-skilled occupations also tend to migrate to larger cities for more job opportunities and better salaries. In this category, *lawyers* and *physicians* are two good examples. Figure 20 and figure 21 show their migration city population patterns. It is seen from the figures that the α values are respectively 1.31 and 1.08. It is not hard to understand that individuals with these occupations have strong preference to migrate to large metropolitan areas searching for better career development and new advanced knowledge.

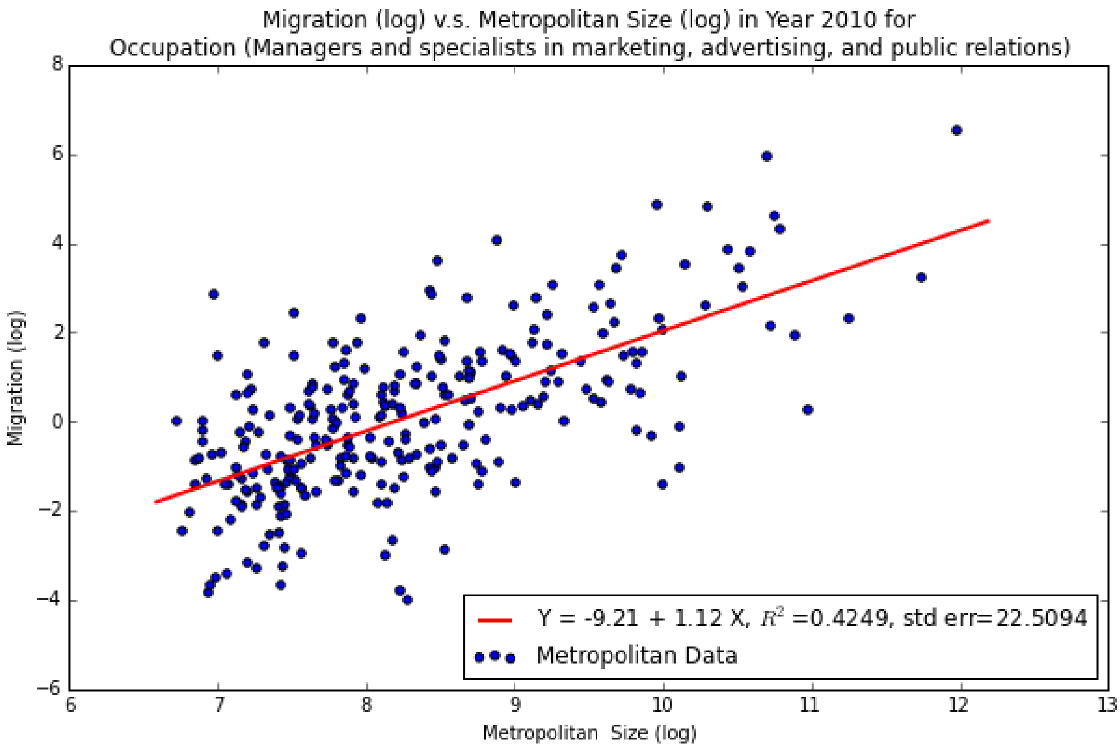

Figure 16. Regression result between in-migration and city size for marketing, advertising, and public relations in 2010

4.4.2.2 Occupations Preferred to Stay in Small Areas

As opposed to above occupations which demonstrate the characteristics of agglomeration economy or advanced skills, more occupations especially conventional occupations are not necessary to get better environment and development in large cities. While having to face more competitions and higher living costs, these occupations would not necessarily get well paid as a return.

Machine operators and *assemblers of electrical equipment* are two typical conventional occupations. Figure 22 and figure 23 illustrate the patterns of migration and city population for these two

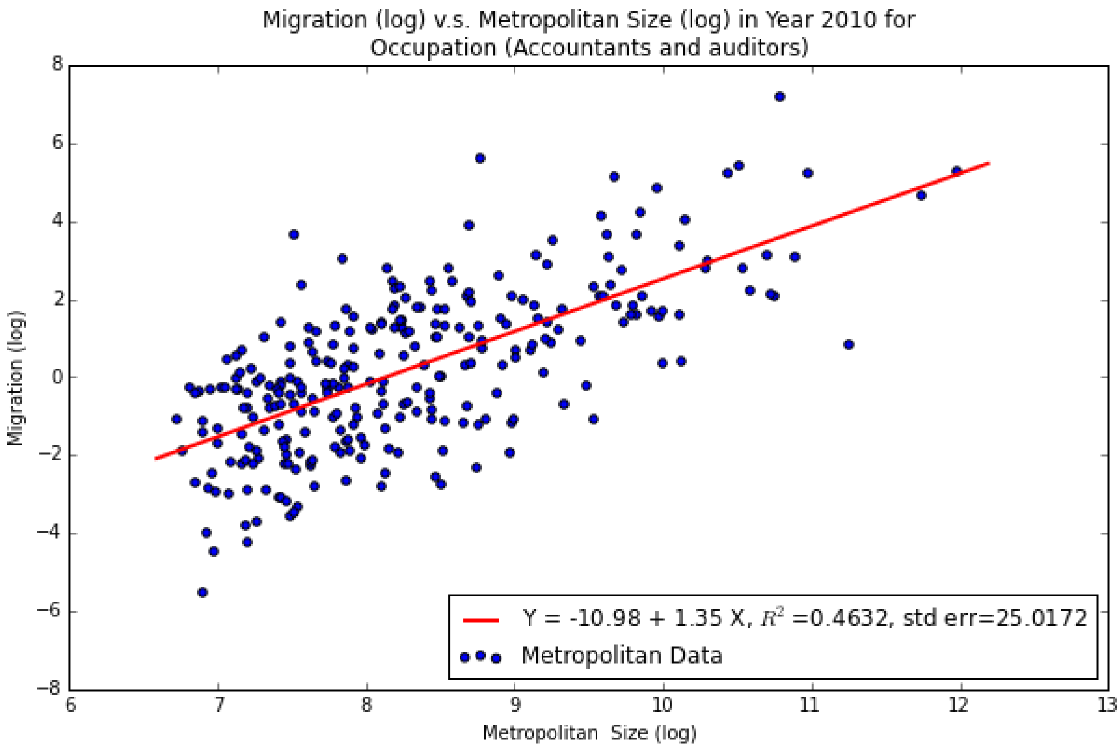

Figure 17. Regression result between in-migration and city size for Accountants and auditors in 2010

occupations. It is shown that α value of *machine operators* is 0.56, much less than 1 and the α value of *assemblers of electrical equipment* is only 0.69. These α values clearly indicate that large areas lose comparative advantages and powers to attract migrants compared with small regions. The explanation is that these occupations do not involve any high technologies, and their characteristics determine the spatial clustering is not necessary and beneficial.

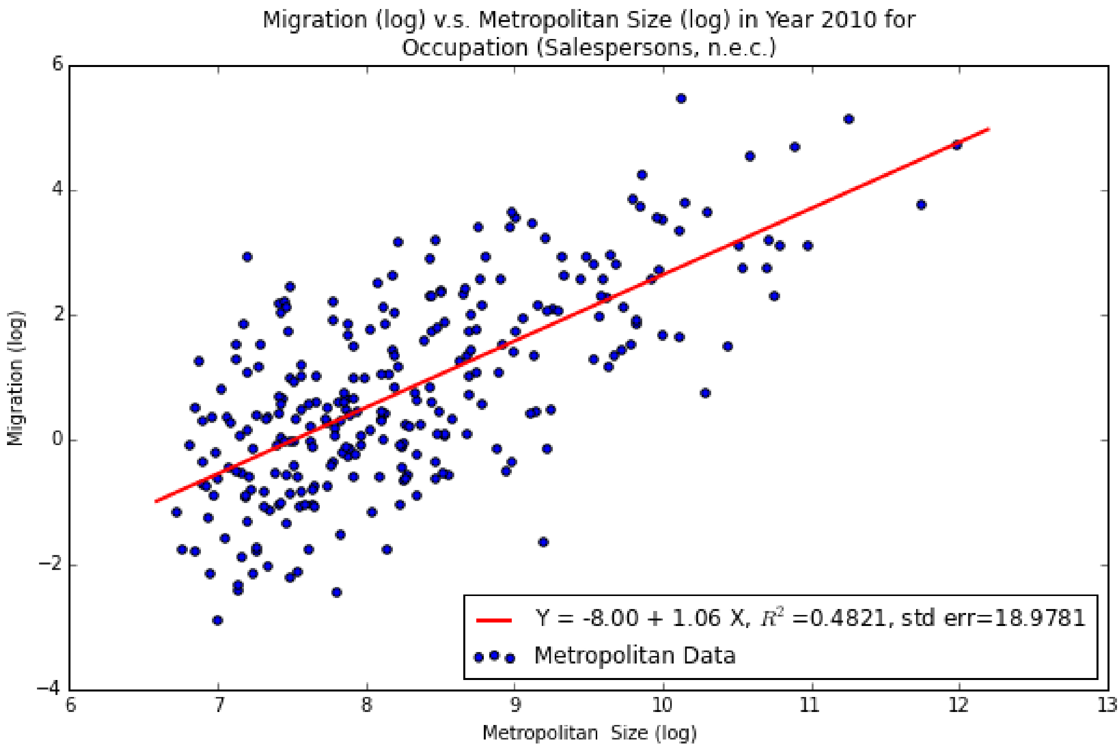

Figure 18. Regression result between in-migration and city size for salespersons in 2010

Truck, delivery and tractor drivers show the similar relationship between migration and city size. It is found in the figure 24 that α value is 0.75, meaning that one unit of increase in log metropolitan size would only cause 0.75 unit of increase in in-migration.

Large metropolitan areas are not attractive for traditional service-oriented occupations as well. Figure 25 demonstrates how number of in-migrants responds to the city population for *cashiers*. Unsurprisingly, the α value is 0.78 less than 1. Same as *cashiers*, *receptionists* display the similar pattern on metropolitan size and migration with α equal to 0.84, which is shown in figure 26. Compared with

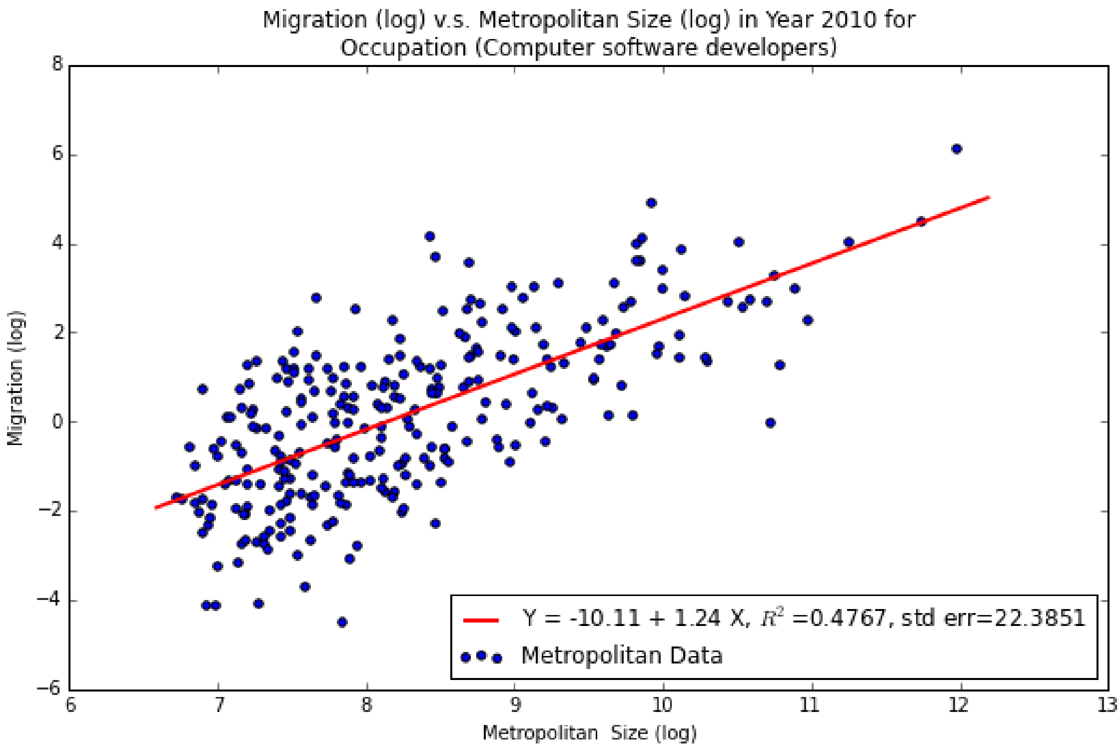

Figure 19. Regression result between in-migration and city size for computer software developers in 2010

small cities, large metropolitan areas do not provide special advantages to receptionists and cashiers, but instead, competitive environment and high living costs in large cities push them out of the regions. Thus, the α values are smaller than 1 and people from these two occupations are less likely to migrate to big cities.

In fact, most conventional occupations are not sensitive to large areas. In the sampled data with 50 occupations, more than a half demonstrates the positive relationship between metropolitan size and migration with α less than 1. Table 4.4.2.2 summarizes main results for these occupations.

Occupation	α Value	R^2
Bookkeepers and auditing clerks	0.88	0.9987
Customer service reps	0.83	0.9644
General office clerks	0.79	0.9204
Stock and inventory clerks	0.88	0.9986
Housekeepers, maids, butlers,etc	0.80	0.9817
Police, detectives and private investigators	0.85	0.9453
Waiter and waitress	0.91	0.9948
Cooks	0.84	0.9990
Waiter's assistant	0.71	0.9611
Misc food prep workers	0.84	0.9948
Nursing aides, orderlies, and attendants	0.88	0.9987
Janitors	0.87	0.9986
Hairdressers and cosmetologists	0.82	0.9049
Supervisors of construction work	0.63	0.7980
Carpenters	0.71	0.9894
Construction laborers	0.70	0.9330
Laborers outside construction	0.76	0.9795
Military	0.50	0.7581
Managers of food-serving and lodging establishments	0.88	0.9791
Registered nurses	0.82	0.9879
Primary school teachers	0.94	0.9966
Supervisors and proprietors of sales jobs	0.89	0.9808

TABLE XII

α VALUES FOR CONVENTIONAL OCCUPATIONS

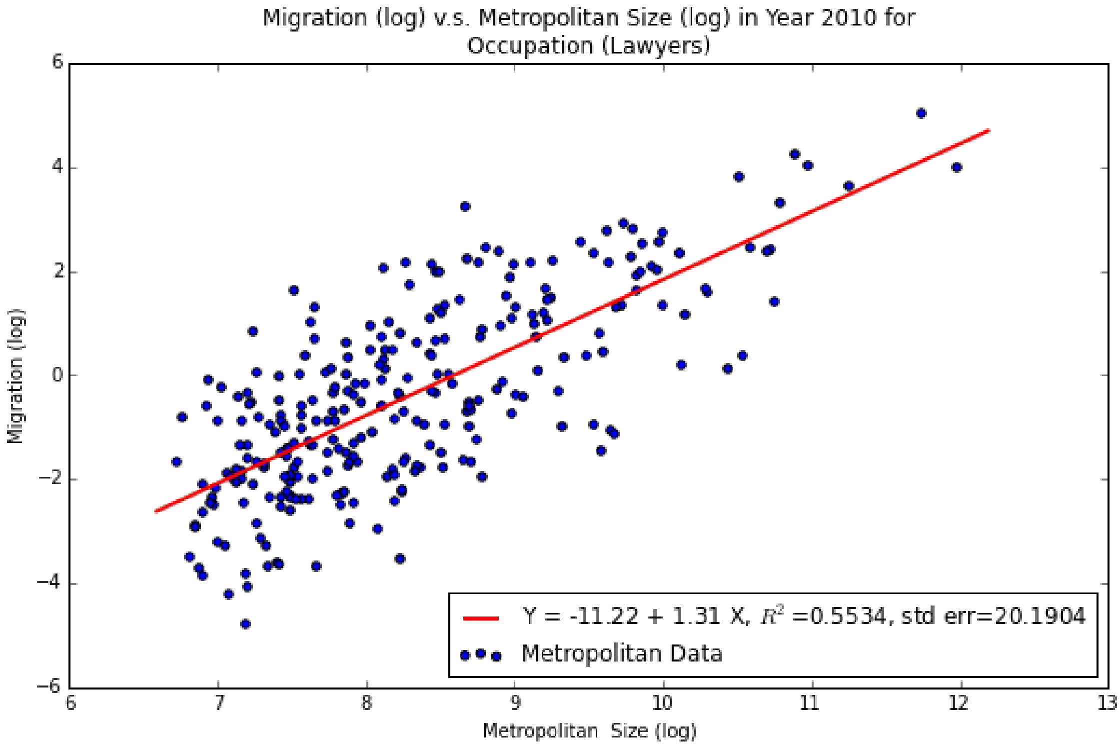

Figure 20. Regression result between in-migration and city size for lawyers in 2010

Since above occupations do not have particular requirements about communication efficiency, fast information share, new advanced technology dependency and many other benefits from large cities, they do not have interests to migrate to big areas. In other words, those typical advantages of big cities cannot particularly attract them to move to large metropolitans. In the meantime, medium or small sized cities could provide professionals from those occupations with equally satisfying conditions. Furthermore, living in large cities has the potential of facing higher living costs, stronger competition and sharing

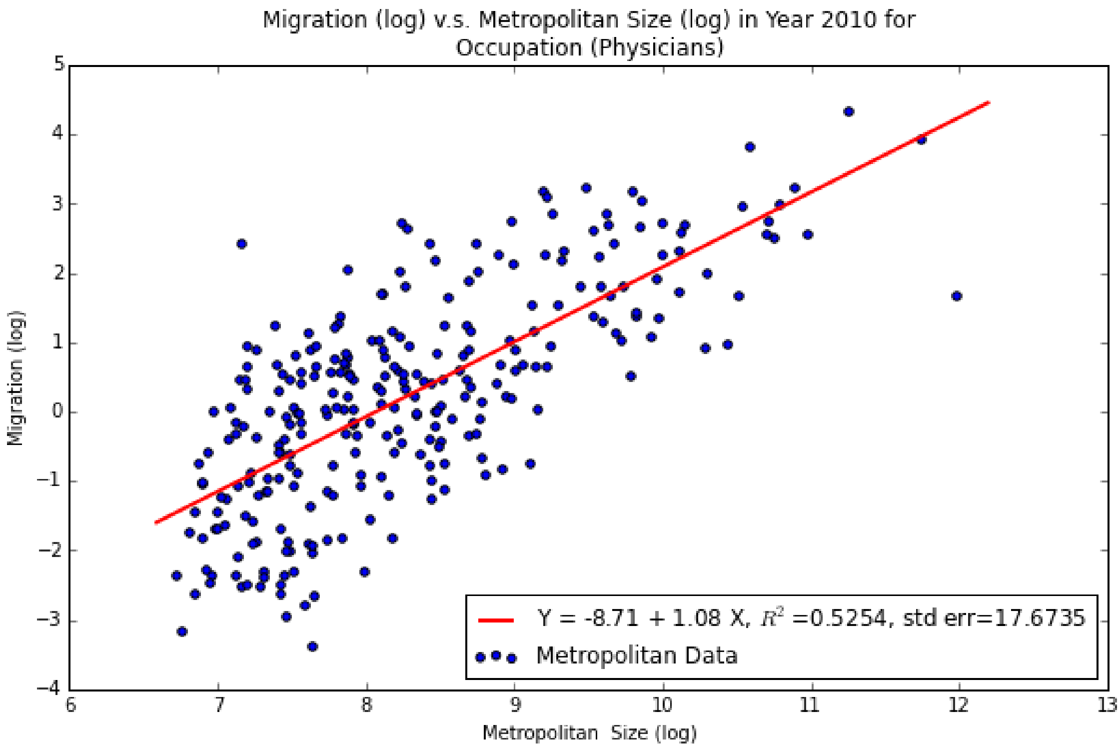

Figure 21. Regression result between in-migration and city size for physicians in 2010

some limited resources with others. Thus, people from conventional occupations do not have strong need to migrate to large metropolitan, whereas they tend to move or stay in smaller cities.

4.5 Why various occupations differ in migration?

In the previous chapter, I have conducted a comprehensive regression analysis on the relationship between migration and city size. From the regression formulation, it is clear that the coefficient α value indicates if the given group is attracted to large metropolitan areas. Specifically, the larger of α, the more likely the group with the given occupation moves to large cities. Through a detailed case study

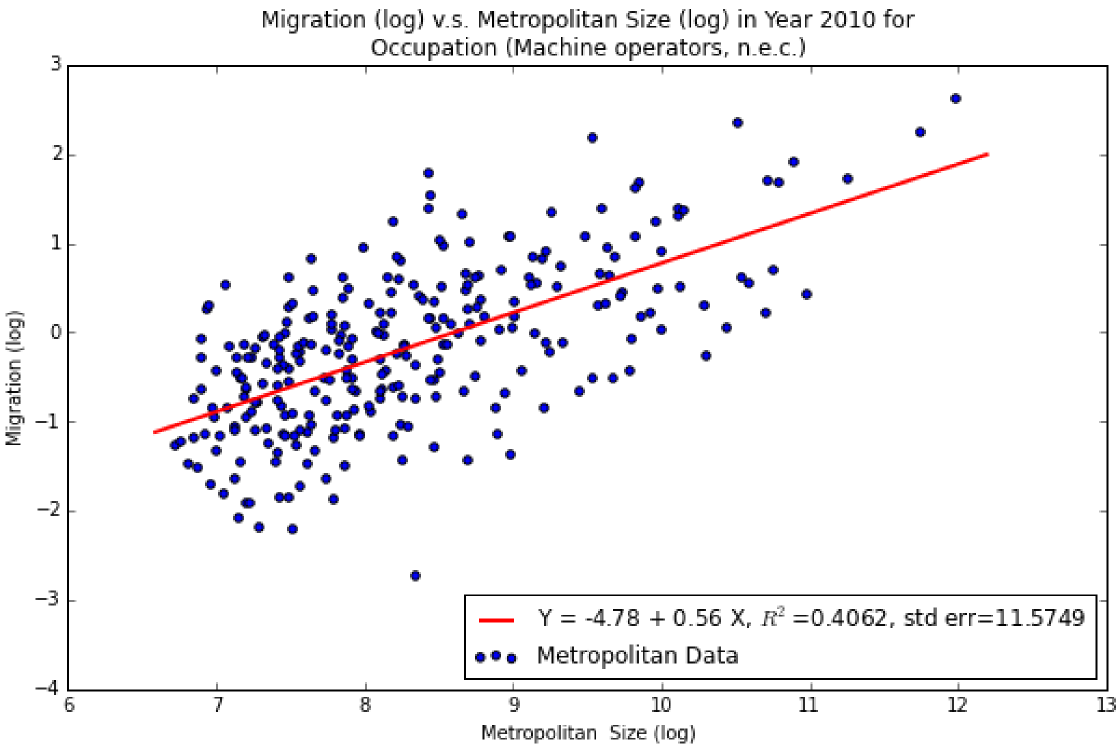

Figure 22. Regression result between in-migration and city size for machine operators in 2010

for top occupations, I discovered that some occupations are more attracted to big cities. The examples include occupations in finance industry and high technology industry. This can be explained using agglomeration economy theories that high-skilled occupations tend to migrate and cluster for better information share and higher communication efficiency. In the meantime, people from some specific occupations tend to move to smaller cities, especially for practioners from conventional occupations and service-oriented occupations. Such disparity shows that different occupations clearly have their own characteristics that lead to various patterns on migration. Despite the detailed study on each occupation

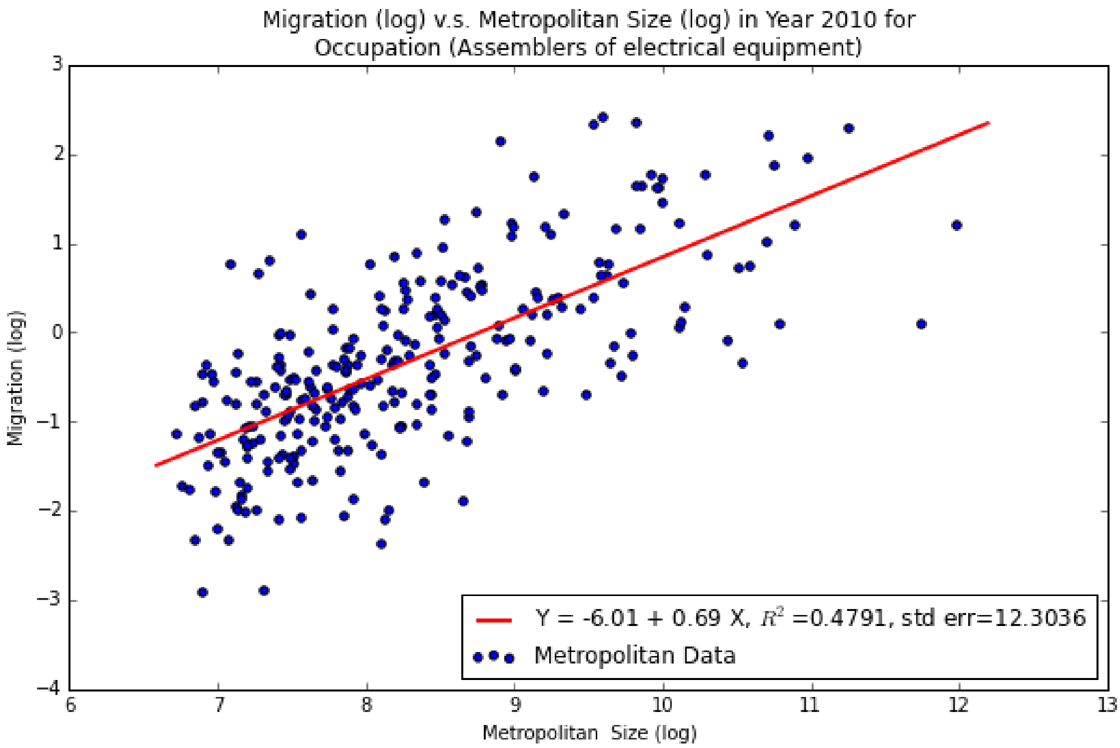

Figure 23. Regression result between in-migration and city size for assemblers of electrical equipment in 2010

that I described in the previous chapter, can we generalize the findings and determine what factors associated with occupations that characterize the migration patterns?

Previous research Chiswick and Miller (1994), Khan (1997), Rao (2010), Dustmann et al. (2011) have shown that education is the indicator that determines the migration pattern. People with higher education levels tend to move to larger and potentially more competitive metropolitan areas, whereas workers with limited education background are more likely to stay or migrate to smaller cities. However, there is few research on the relationship between migration and other factors, especially those factors

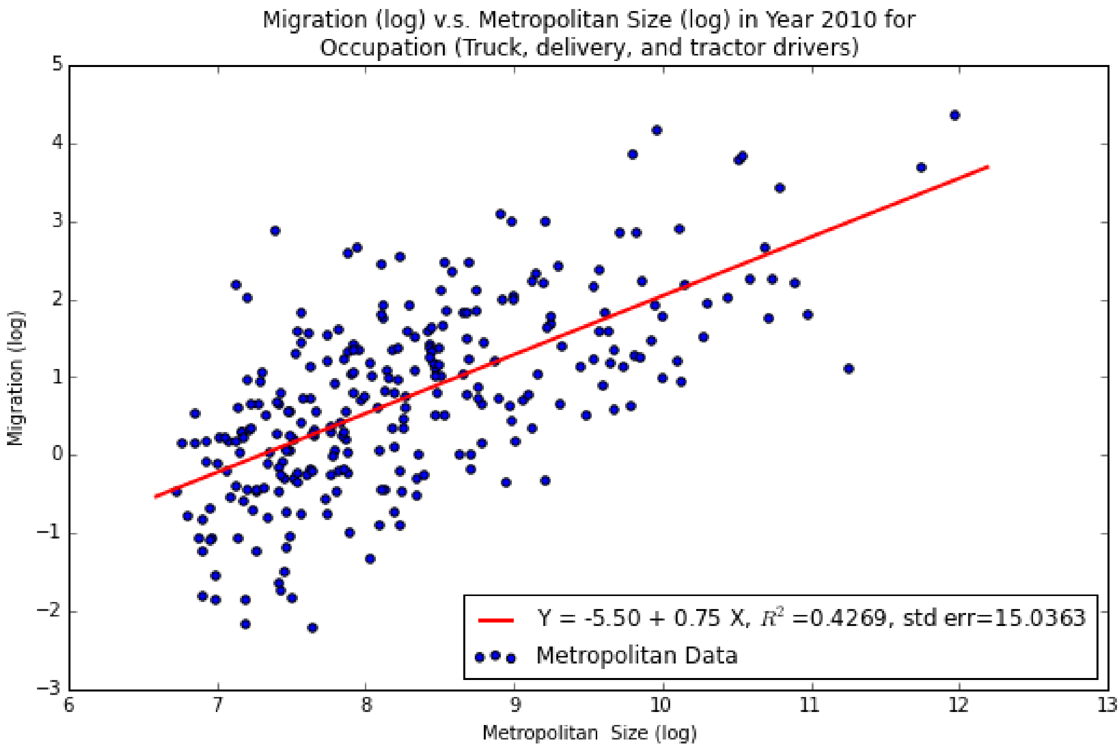

Figure 24. Regression result between in-migration and city size for truck delivery and tractor drivers in 2010

associated with occupations. As various occupations clearly differ largely in terms of migration, it is valuable to evaluate the important factors that affect the migration phenomenon from the perspective of occupation. In this chapter, I extract core attributes associated with occupations and use a quantitative approach to study the statistically significant factors that determine the distinguishing migration patterns of professionals from different occupations. I first describe the attributes that are extracted on the occupation level as follows.

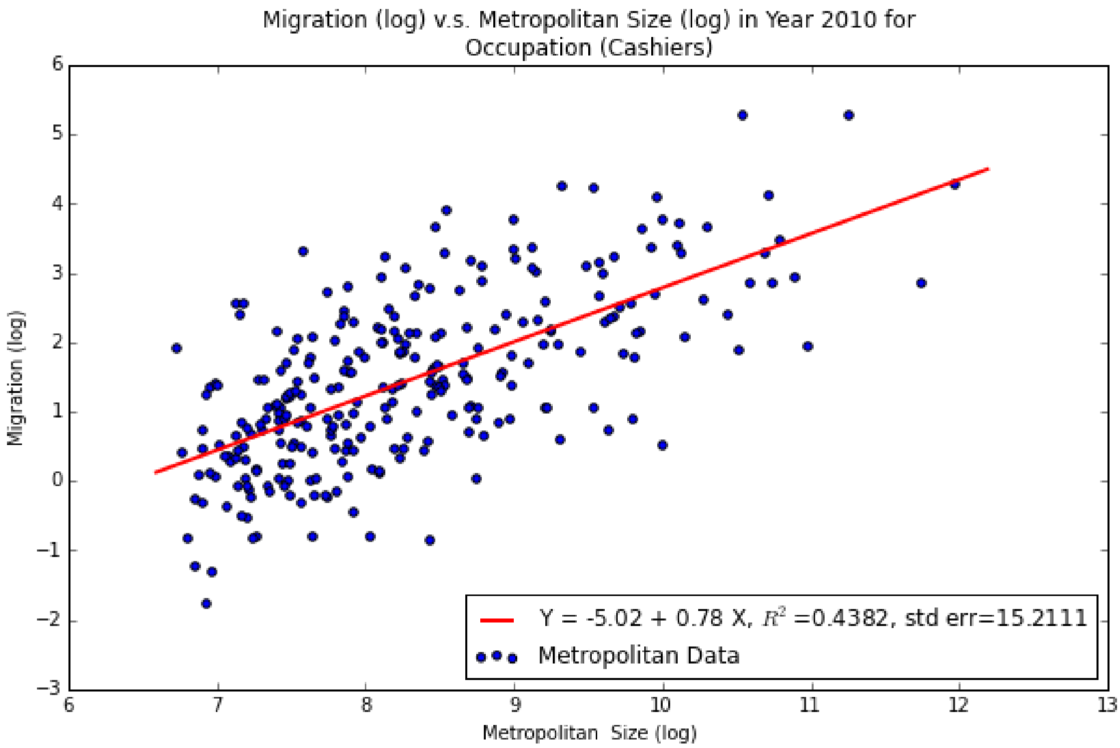

Figure 25. Regression result between in-migration and city size for cashiers in 2010

4.5.1 Average Age

The first factor that I study is the average age of professionals with a specific occupation. Since migration involves permanent movement into new residential areas and different age groups may show different degrees on the tendency of movement, it is natural to measure the average age associated with each occupation. A question we want to answer here is if some specific occupations are more likely to migrate to big cities simply because of their corresponding groups are younger or older.

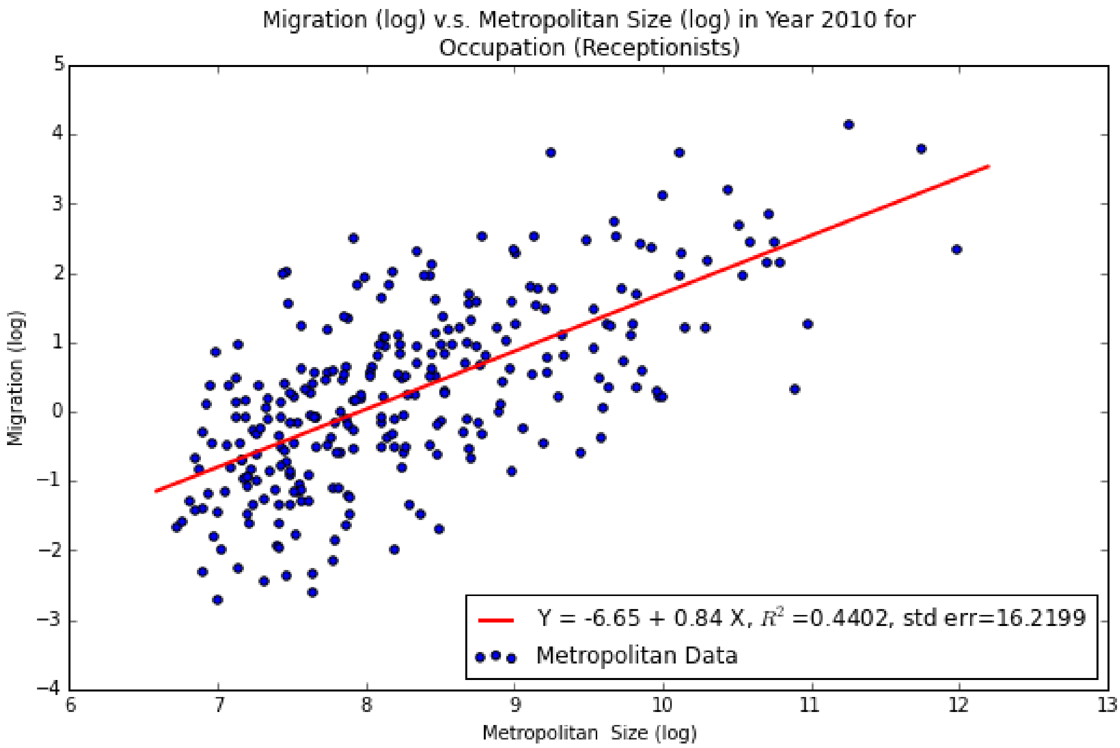

Figure 26. Regression result between in-migration and city size for receptionists in 2010

Figure 27 first presents the distribution of average ages for the most 50 popular occupations in the United States. The majority of the average age lies in the range of 38 to 48, and not surprisingly, the shape of the distribution can be roughly regarded as bell- shaped. The occupation with the smallest average age is *military* with an average age at 29.78 whereas the occupation with the highest average age is *chief executives and public administrators* at 51.63. The relationship between average age and the regression coefficient α is illustrated in Figure 28 and each dot in the figure represents an occupation. From the figure, each occupation does not show a strong correlation between average age and α as the

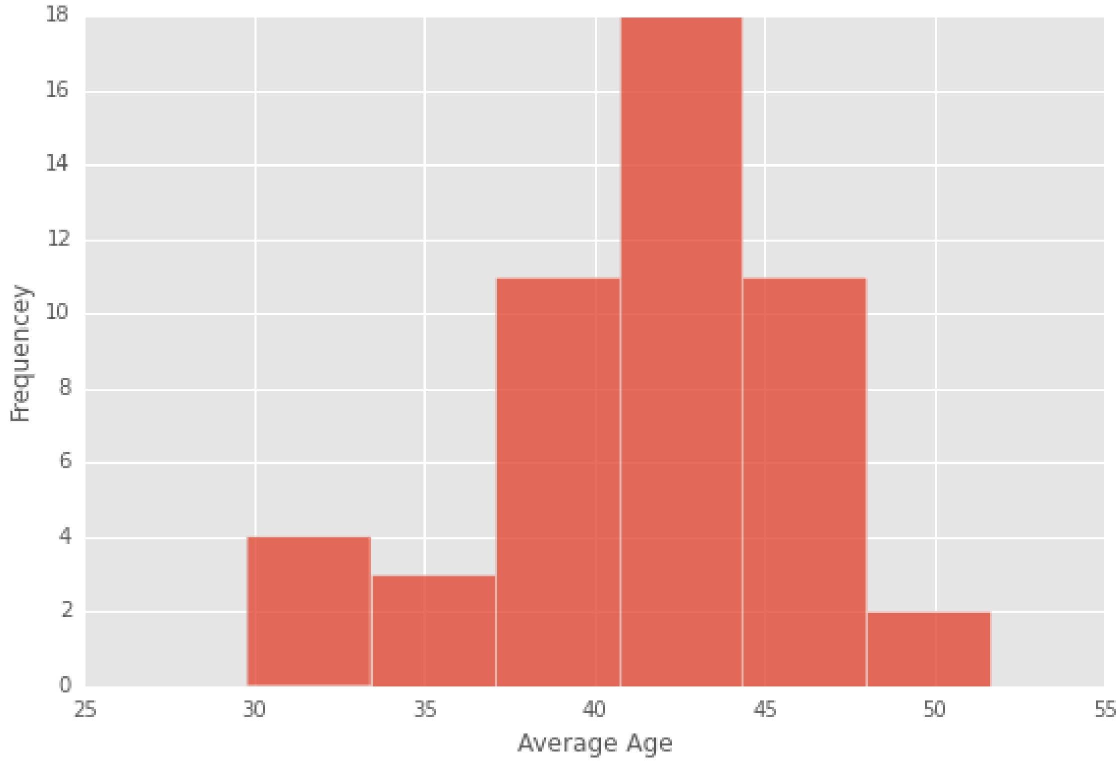

Figure 27. Average Age Distribution Across Top 50 Occupations

data points are seemingly random scattered. Later on I will present a regression model to quantify the effect of average age on migration.

4.5.2 Average Education Level

One important factor associated with occupation is education which has been mainly studied on migration in previous research. Here I extract a numeric value representing average education level for each occupation that is derived from Flood et al. (2015) and a higher value means the workers with the corresponding occupation have an overall higher education level. The education levels and their

Numeric Value	Education Level
00	N/A or no schooling
01	Nursery school to grade 4
02	Grade 5, 6, 7, or 8
03	Grade 9
04	Grade 10
05	Grade 11
06	Grade 12
07	1 year of college
08	2 years of college
09	3 years of college
10	4 years of college
11	5+ years of college

TABLE XIII

EDUCATION LEVELS AND THEIR CORRESPONDING NUMERIC VALUES

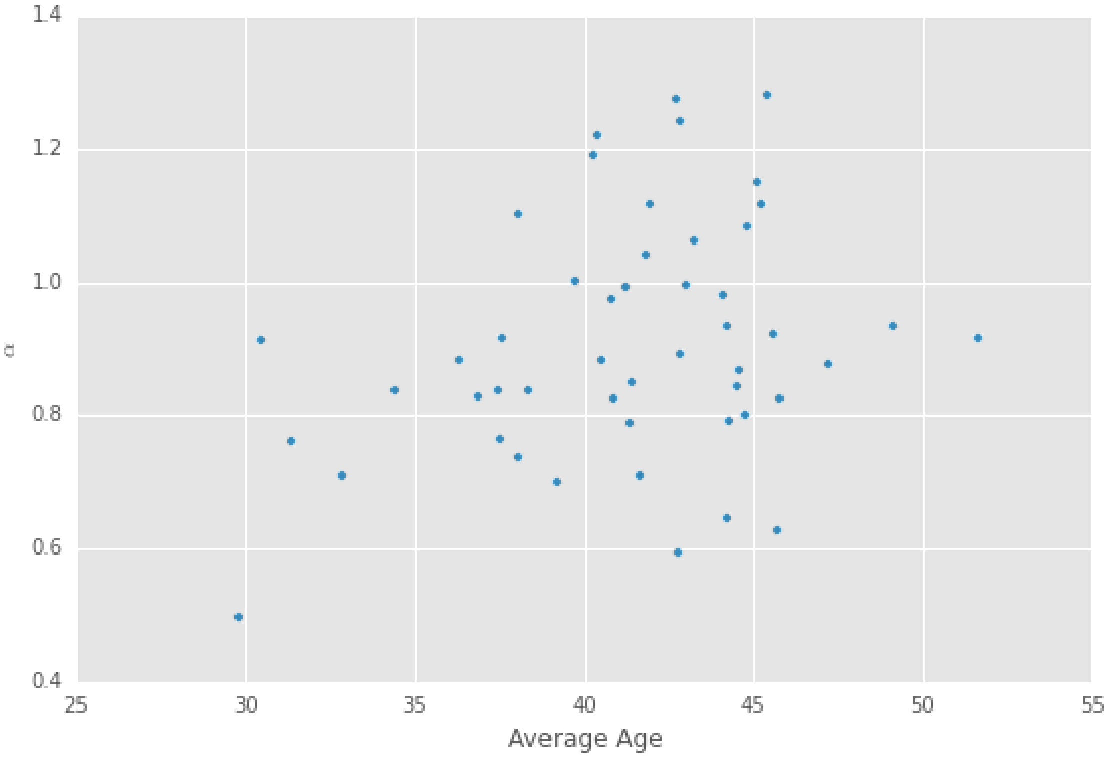

Figure 28. Relationship Between Average Age and Regression Coefficient α

related numeric values are shown in Table 4.5.1. The histogram on average education level is presented in Figure 29. From the figure, one can observe that all occupations have the average education level between 4 and 11. The occupation that has the highest average education level is *Physicians* with a value at 10.96, and the occupation with the lowest average education level score is *Housekeepers, maids, butlers, stewards, and lodging quarters cleaners* at 4.84.

I further show the relationship between average eduction level and regression coefficient α in Figure 30. It is clear that the average eduction level factor has a strong correlation with the regression

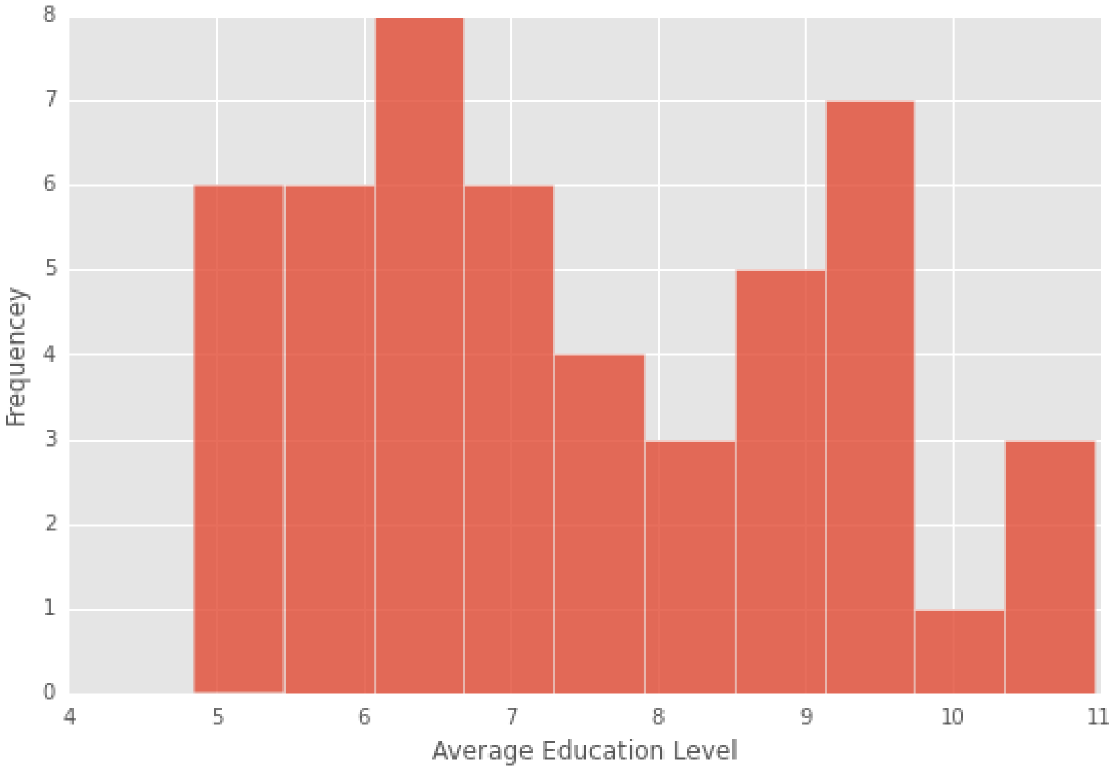

Figure 29. Average Education Level Distribution Across Top 50 Occupations

coefficient α, and this observation is consistent with previous research results. Occupations with higher values on average eduction level also have higher α that indicates these occupations are more likely to migrate to big cities. One outlier shown in the figure is *military*. This occupation has nearly education level 8 but α value is only 0.5. To further establish the relationship between average education level and α, a regression analysis is applied and the result is shown in the figure as the red line. The R^2 of the regression model is 0.4747 which means by using average eduction level alone, the model can explain 47.47% variability of the response variable, namely α. In other words, education indeed is critical on

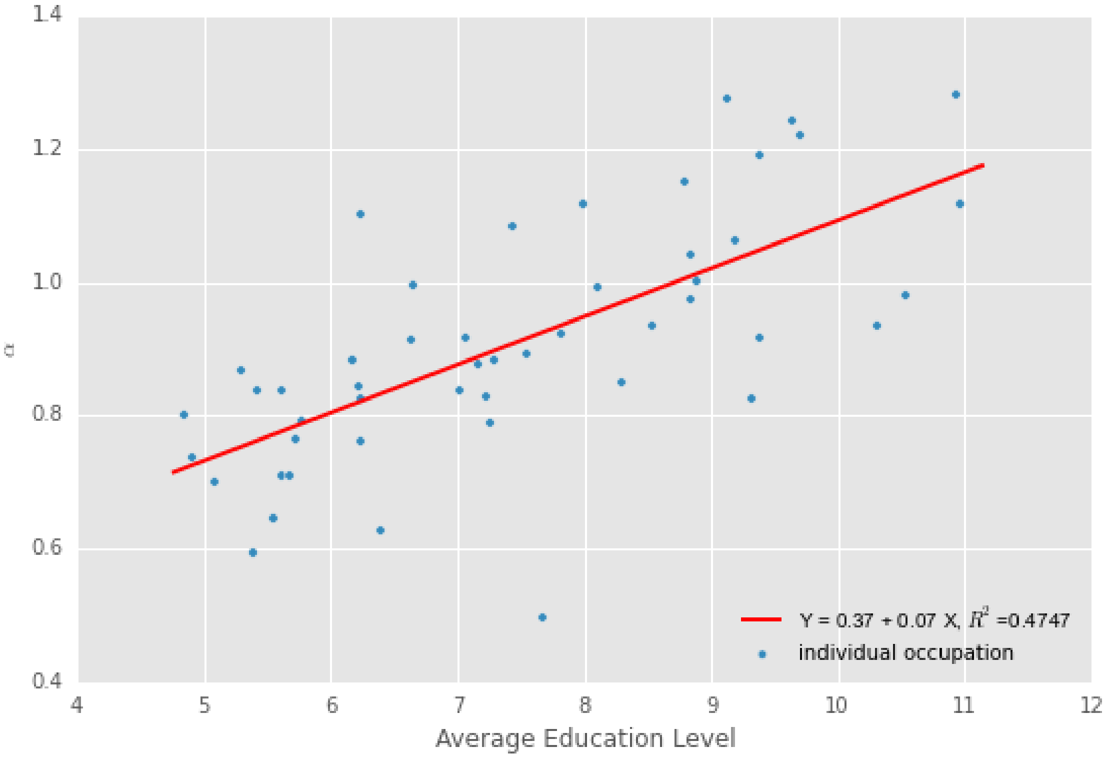

Figure 30. Relationship Between Average Education Level and Regression Coefficient α

determining if a given occupation tends to migrate to large metropolitan areas. However, the question still remains that if education is the only occupation related factor that is important to migration. I will answer this question in the end of this chapter.

4.5.3 Industry Coverage

Another aspect related to occupation that is worth studying is industry coverage. Specifically, given an occupation, I measure how many industries employ workers with this occupation. This measurement demonstrates how widely required of the given occupation. The distribution of industry coverage is

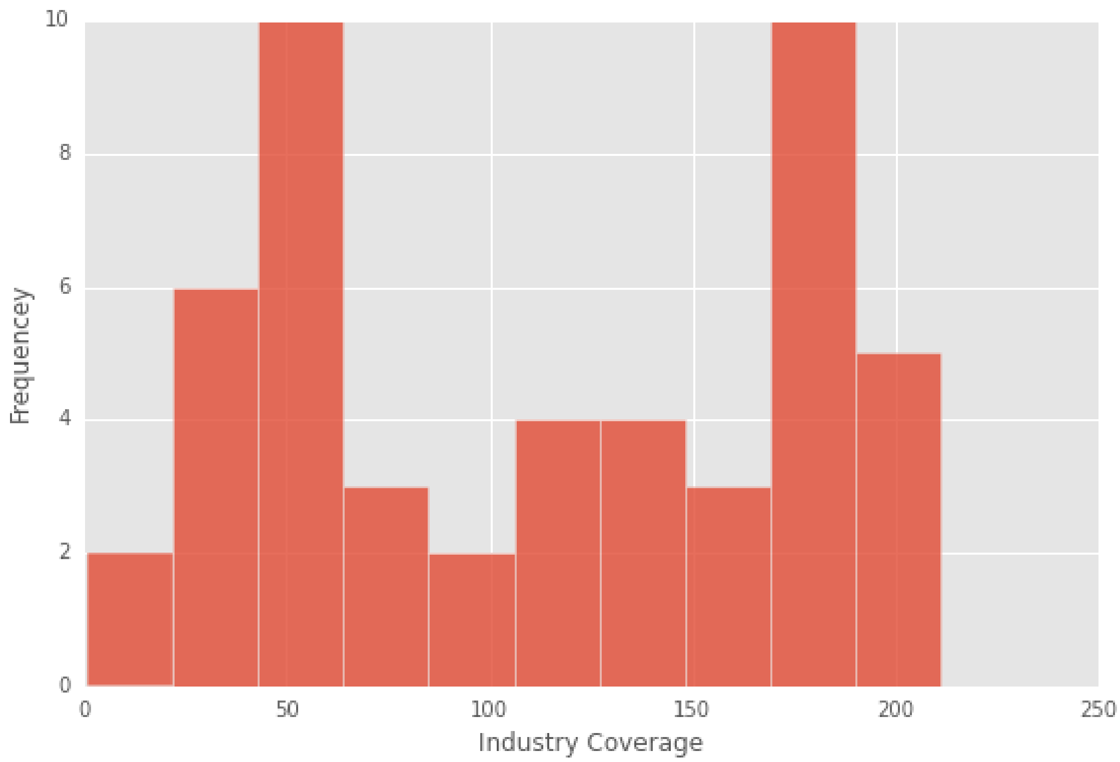

Figure 31. Industry Coverage Distribution Across Top 50 Occupations

illustrated in Figure 31. As the Census Bureau has reorganized its industrial classification system in almost every census administered, a consistent classification scheme is required to conduct such analysis. Thus, the 1990 Census Bureau industrial classification scheme is used to classify industries and there are 225 industries identified in total.

One can note that this measure has wide ranges across different occupations. Some occupations are needed in only a few limited industries but some occupations are widely adopted in hundreds of industries. *Primary school teachers* as an occupation has the least industry coverage by appearing only

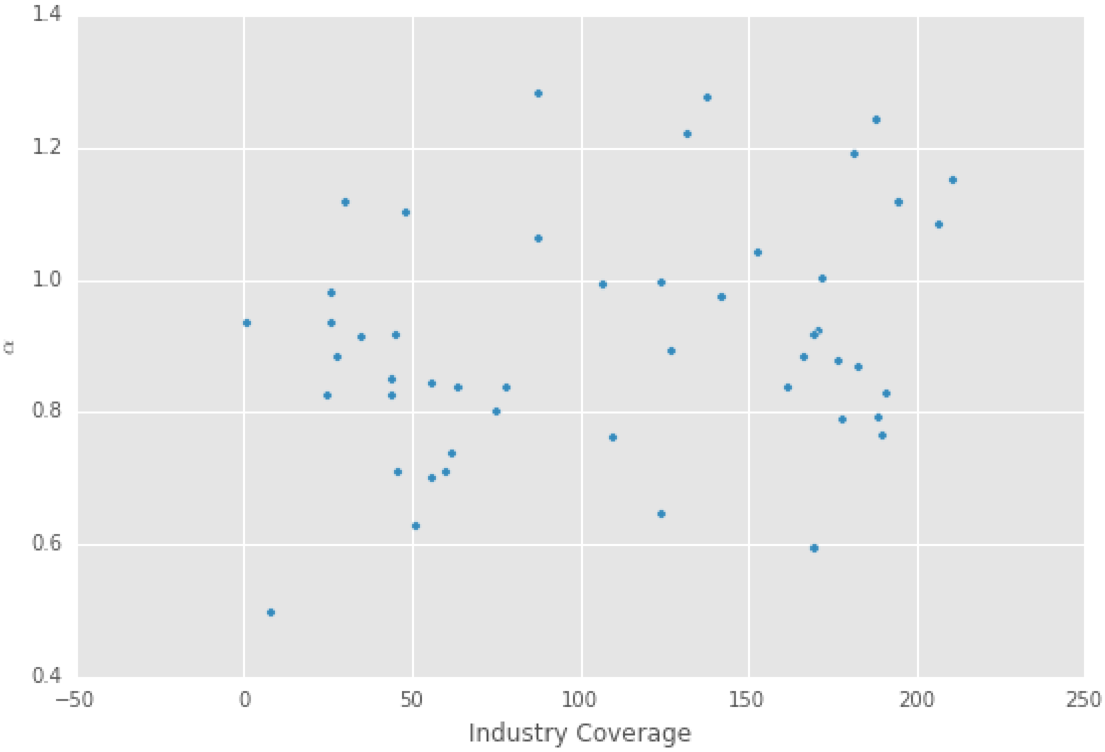

Figure 32. Relationship Between Industry Coverage and Regression Coefficient α

in one industry *elementary and secondary schools*. *Managers and administrators* is the occupation with the largest industry coverage value by showing up in 211 industries. This is reasonable as most (if not all) industries require leadership and management positions such as managers and administrators.

I further show the scatter plot on industry coverage for each occupation to illustrate the relationship between industry coverage and regression coefficient α in Figure 32. There is an observable positive relationship between industry coverage and α as the value of regression coefficient α increases when the industry coverage gets a higher value. This can be explained by the fact that an occupation is more

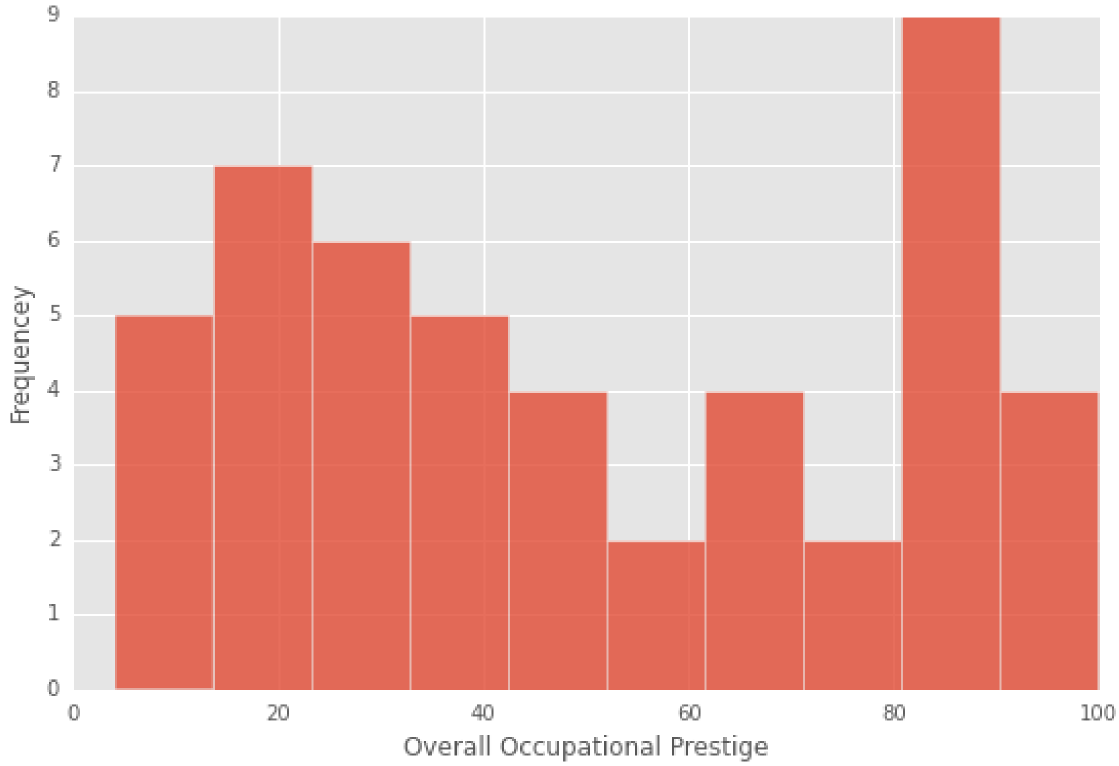

Figure 33. Overall Occupational Prestige Distribution Across Top 50 Occupations

easily to migrate to big cities if it is more widely adopted in many industries. As the industrial need goes up, the workers with the specific occupation have more flexibility on migration and tend to cluster in more dense areas to achieve better economic efficiency.

4.5.4 Overall Occupational Prestige

One important characteristic of occupation is the occupational status. The occupational socioeconomic status represents a level of living for persons in the occupation. I adopt the Nam-Powers-Boyd occupational status score developed in Nam and Boyd (2004) and use the score to reflect the overall

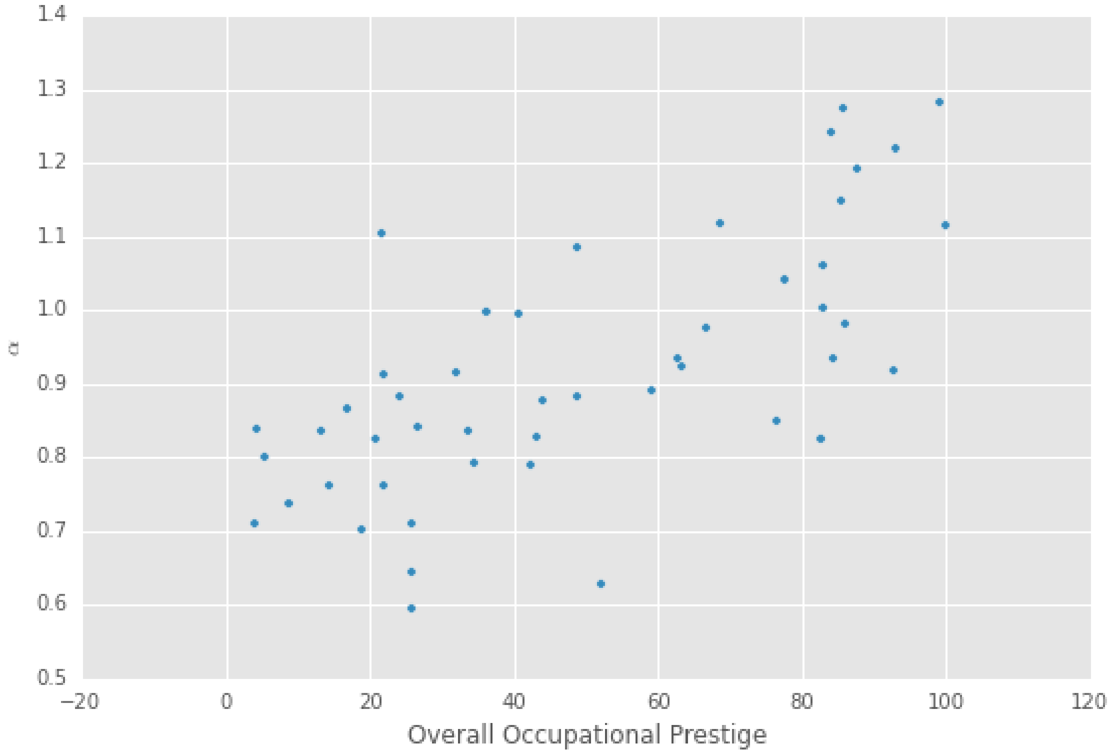

Figure 34. Relationship Between Overall Occupational Prestige and Regression Coefficient α

occupational status, where a higher score represents a higher socioeconomic status of the occupation and score is between 0 and 100.

The distribution of the overall occupational prestige store is shown in Figure 33. Clearly the top 50 occupations fall into two main clusters with one cluster around an occupational prestige score at 20 and the other cluster at 90. The occupation with the lowest prestige score is *Waiter's assistant* (a score of 4.1), whereas the highest prestige score belongs to occupation *Physicians* with a score of 99.9. I further show the relationship between overall occupational prestige score and the regression coefficient

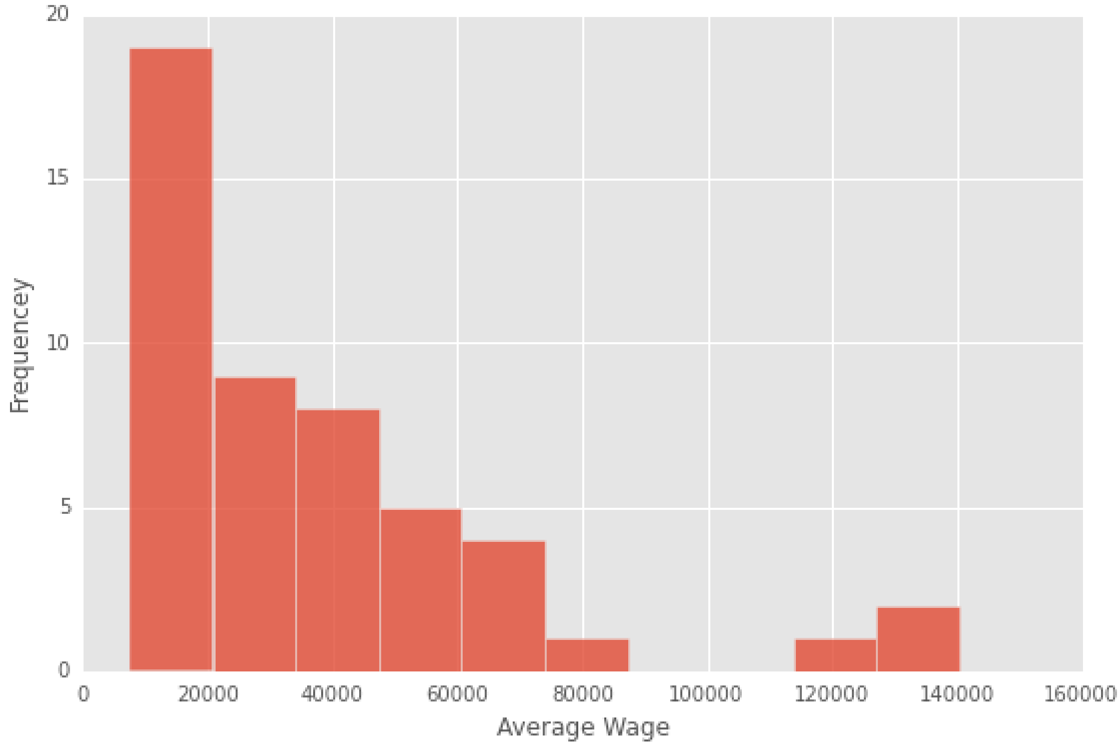

Figure 35. Average Wage Distribution Across Top 50 Occupations

α in Figure 34. It is clear that an occupation with a higher prestige score is more likely to have a higher coefficient α value. Notice that *child care workers* is one exception shown in the figure. Its occupational prestige score is as low as 20, but the α value is higher than 1.1. This positive correlation demonstrates that persons from a more prestigious occupation tend to migrate to bigger cities, probably due to the abundant resources, strong demand, sound public facilities and other benefits associated with big cities.

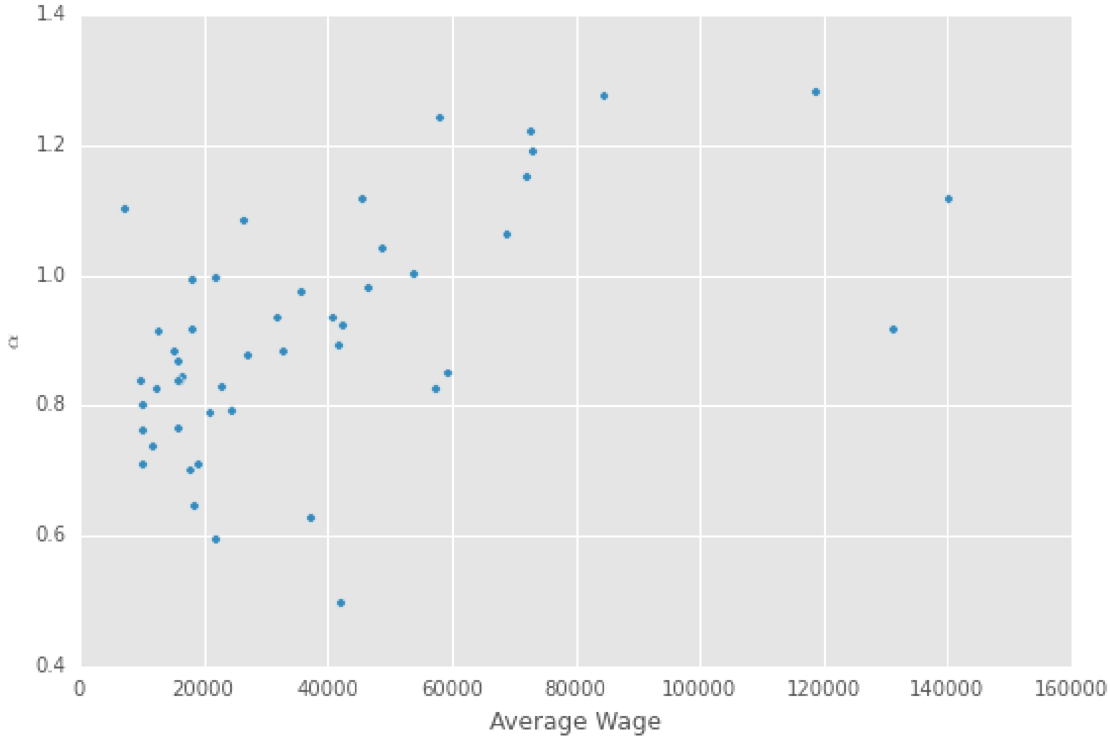

Figure 36. Relationship Between Average Wage and Regression Coefficient α

4.5.5 Average Wage

Another important attribute associated with the occupation is wage. The average wage of an occupation naturally indicates the general financial outcome for the given occupation. The distribution of average wage for the top 50 occupations is illustrated in Figure 35. For those top 50 popular occupations in the United States, the average wages follow a long tail distribution rather than a normal distribution. Specifically, 40% of the occupations have an average wage below $\$20k$. The occupation with the lowest wage is *child care workers* with an average wage at \$7,508. There are three occupations with average wage above $\$100k$, and they are *chief executives and public administrators*, *lawyers* and *physicians*.

To discover if wage is correlated with the patterns of migration, I show the relationship between average wage and regression coefficient α in Figure 36. One can observe that workers in occupations with lower average wage are more likely to migrate to smaller cities. In the meantime, occupations with average wage higher than $\$60k$ generally have α values greater than 1, indicating their tendency to migrate to big cities. One can see that there are three outliers in the top right part of Figure 36. Compared with other occupations, these three have much higher average wages. They are respectively *lawyers*, *physicians* and *chief executives and public administrators*. It is worth noticing that the average wage of *chief executives and public administrators* is almost $\$130k$, but its α value is below 1.

4.5.6 Average Residential Population

As different occupations cluster in different areas across the country and this may also affect the decisions on migration, it is valuable to measure the characteristic of occupations in the perspective of population. Here I calculate the average residential population for each occupation and use it as an attribute that represents the geographic nature of occupations. Similar to the approaches before, the distribution of average residential population is shown in Figure 37. The population distribution is bell shaped and *military* as an occupation has the lowest average population at 6727 hundreds. The occupation with the largest average residential population is *designers* at 23988 hundreds. The relationship between average residential population and regression coefficient is shown in Figure 38. It is clear that these two variables have a strong positive relationship and α increases as average residential population goes up. This demonstrates that for an occupation that is popular in large cities, the workers for that

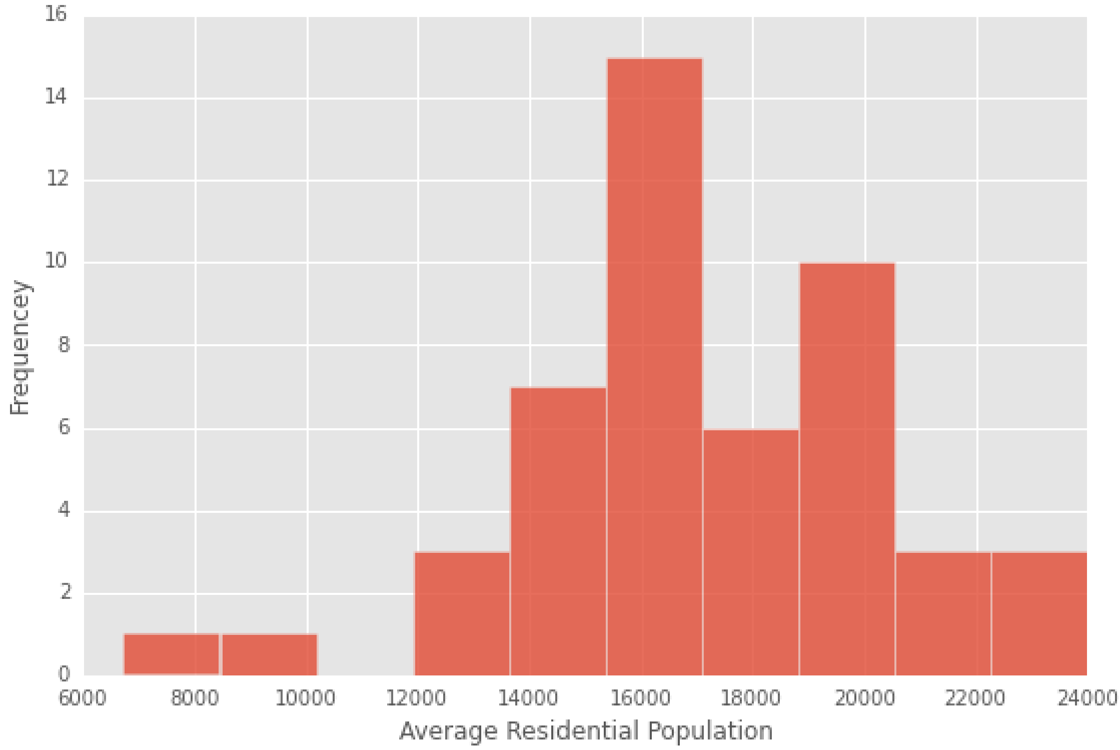

Figure 37. Average Residential Population Across Top 50 Occupations

occupation are more likely to be attracted to large cities as well since the employment opportunities are abundant.

4.5.7 Unemployment Rate

The last attribute that I study is the unemployment rate. Does the unemployment rate associated with occupations affect the workers' decisions on migration? Are workers in occupations with low unemployment rate more likely to migrate? These questions are critical to understand the migration patterns from the economic view for each occupation.

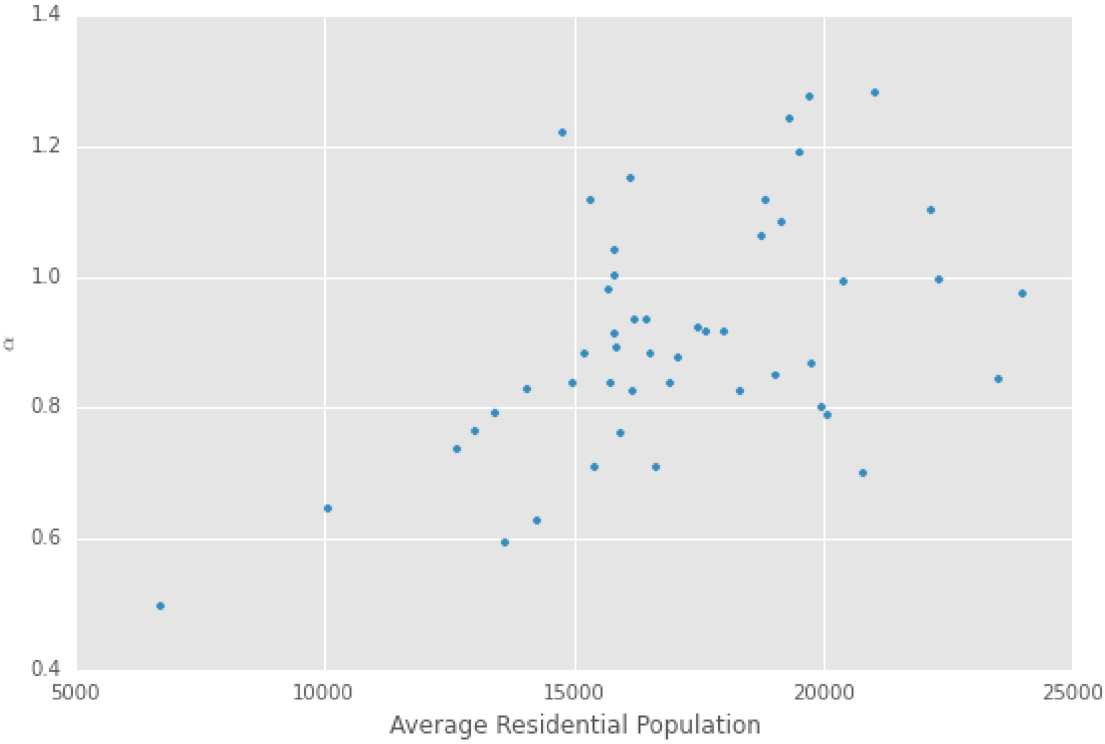

Figure 38. Relationship Between Average Residential Population and Regression Coefficient α

I calculate the unemployment rate for each occupation in the sampled year and show the histogram in Figure 39. The data roughly follows a bell shaped distribution with the majority occupations having unemployment rates less than 0.15. *Physicians* is the occupation with the lowest unemployment rate at 0.69%, and *construction laborers* has the highest unemployment rate at 18.1%. To uncover the correlation between unemployment rate and the tendency of migrating to big cities, their relationship is demonstrated in Figure 40. It is obvious that they have a negative correlation as the value of α decreases when an occupation has a higher unemployment rate. This is because an occupation with a higher

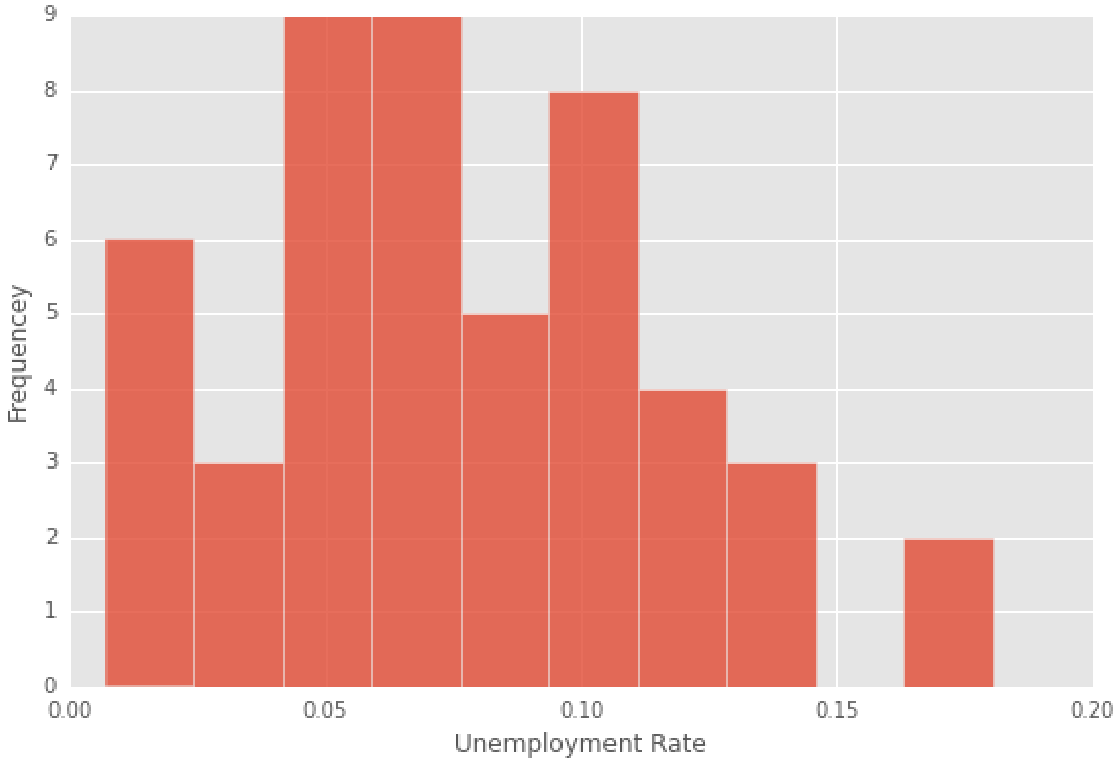

Figure 39. Unemployment Rate Across Top 50 Occupations

unemployment rate has less power to attract migrants, people have less drive and incentive to migrate to large metropolitan areas where are potentially more costly and competitive to live.

4.5.8 What factors affect migration patterns?

As a number of attributes on occupation have been extracted and they measure different aspects and characteristics of occupation, it is important to discover and identify their relationships with the disparity of migration phenomenon. I have shown that some measures are highly correlated with migration patterns on moving to large cities in either positive or negative way, and some other attributes are seem-

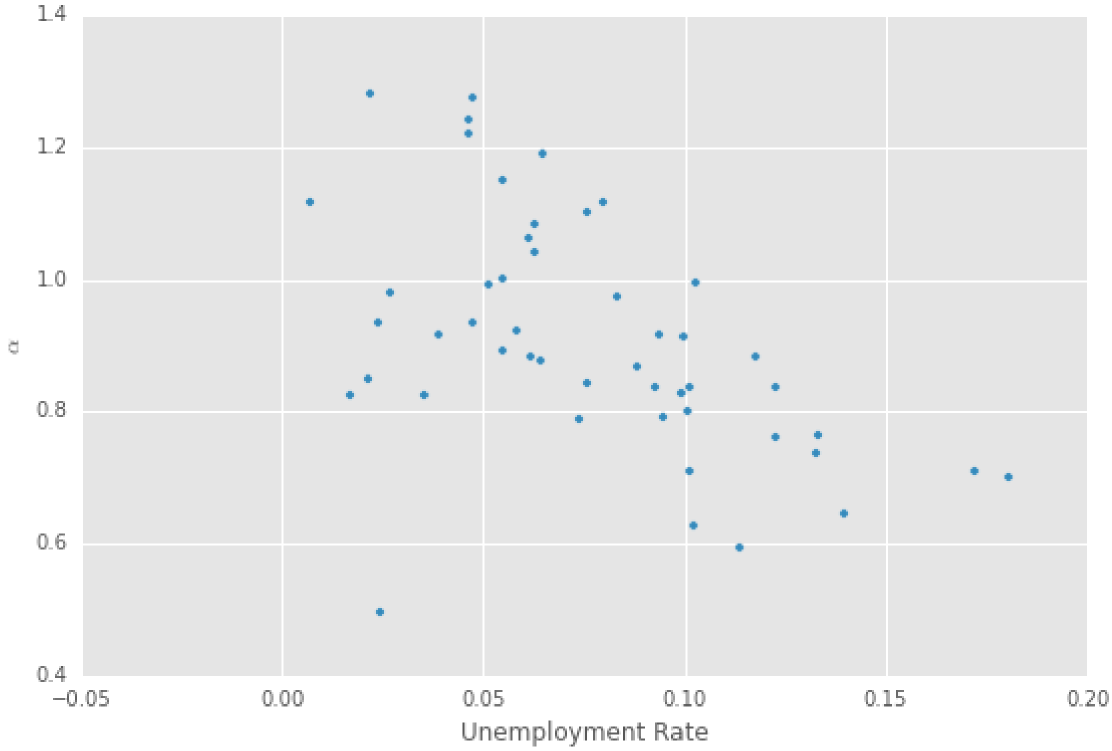

Figure 40. Relationship Between Unemployment Rate and Regression Coefficient α

ingly not related to migration. It is important to have a quantitative understanding on what factors are statistically significant in terms of migration, especially on the tendency of moving to large metropolitan areas.

To study the effects of the attributes listed in this chapter, I establish a regression analysis by treating the coefficient α in Equation 4.1 as the dependent variable. Since the value of α indicates the tendency of migrating to big cities, the regression analysis will establish a relationship between the seven extracted critical attributes and the tendency of moving to large metropolitan areas. The regression model will

Dep. Variable:	Y	R-squared:	0.718
Model:	OLS	Adj. R-squared:	0.670
Method:	Least Squares	F-statistic:	14.95
Prob (F-statistic):	1.67e-09	Log-Likelihood:	44.577
AIC:	-73.15	BIC:	-58.02
Df Model:	7		

	coef	std err	t	P>|t|	[95.0% Conf.	Int.]
Intercept	0.7102	0.293	2.426	0.020	0.119	1.301
age	-0.0112	0.004	-2.527	0.028	-0.020	-0.002
edu	0.0514	0.020	2.520	0.006	0.010	0.093
industry	0.0009	0.000	3.673	0.002	0.000	0.001
prestige	-0.0004	0.000	-2.796	0.026	-0.001	-0.000
wage	1.021e-06	9.02e-07	1.132	0.264	-8e-07	2.84e-06
population	1.381e-05	5.63e-06	2.455	0.008	2.45e-06	2.52e-05
unemploy	-0.9999	0.748	-1.338	0.188	-2.510	0.510

Figure 41. Regression Results on Occupational Factors

also confirm if the proposed attributes can fit the output well and test the statistical significance of each individual factor. Through the regression analysis, it will generalize what factors fundamentally lead to various behaviors on migration for different occupations.

$$\begin{aligned}\alpha = \beta_0 + \beta_1 \cdot age + \beta_2 \cdot edu + \beta_3 \cdot industry + \beta_4 \cdot prestige \\ + \beta_5 \cdot wage + \beta_6 \cdot population + \beta_7 \cdot unemploy\end{aligned} \tag{4.2}$$

The regression model is formulated as shown in Equation 4.2 and its results are illustrated in Figure 41. One should note that the independent variables *age*, *edu*, *industry*, *prestige*, *wage*, *population* and *unemploy* represent the attributes discussed earlier, namely average age, average education level, industry coverage, overall prestige, wage, average residential population and unemployment rate, respectively.

The regression model gives a R^2 value at 0.718, which is a major improvement on using the education level as the independent variable alone as shown in Figure 30. The proposed regression model using seven occupational factors accounts for 71.8% of the variance of α. From the regression results, five attributes are significant, namely age, average education level, industry coverage, occupational prestige and average residential population. They all have p-values less than 0.05. Specifically, education, industry coverage and average residential population are the most significant attributes with regard to α since their p-values are very close to zero. The coefficients for these three attributes being positive further demonstrate that all these three attributes have positive relationships with the tendency of migrating to big cities. One should also note that average age and occupational prestige have p-values at 0.028 and 0.026, respectively. If 97% confidence interval is applied, these two attributes are considered marginal and do not provide strong evidence against the hypothesis, although they both show some degrees of correlation with the dependent variable.

Education level is not surprisingly statistically significant, since workers for an occupation with higher education levels indicate that the specific occupation is more likely to be skill intensive and the corresponding workers tend to migrate to larger cities for more job opportunities, higher compensations and better environments on career development. Despite the common claims from previous research that education is the only main factor that affects migration decisions, the regression results further disclose that two occupation related attributes, industry coverage and average residential population, are also significant. On the one hand, if an occupation is required by a large number of industries, this shows that the workers could find job opportunities in large cities more easily since there are more diversified industries and employers in large cities. On the other hand, if only a handful industries need workers from a specific occupation, the workers have less flexibility on migrating and they do not have strong motivations on moving to larger metropolitan areas. Usually these occupations are more specific and they tend to cluster in smaller areas.

The significance of average residential population indicates that the higher average residential population of an occupation, its workers are more likely to migrate to large cities. This is reasonable since for occupations condensed in cities with large populations, their workers are more likely to migrate to these cities for opportunities. In the meantime, workers tend to flow to smaller cities if their occupations are more popularly located in smaller areas. These clearly demonstrate the consequences of agglomeration economies.

4.6 Conclusion

In this chapter, I study the relationship between city size and migration. Based on the regression analysis, I find that migration is positively related to the city population with coefficient approximately equal to 1, and this relationship has been held across 7 years. This result further indicates that larger cities do not necessarily have higher ability of attracting in-migrants compared with smaller cities, although the number of in-migrants is bigger in larger areas.

In order to more deeply understand the effect of city size on migration and specify the category and type of migration, I conduct the more detailed analysis on city size and migration by occupation. According to the theory of agglomeration economy, the nature of occupation determines whether it needs to be spatially clustered together in big areas to take advantage of benefits from agglomeration economy.

Hence, individual's occupation largely determines his migration choice - whether moving to larger cities or not. In the regression model, the value of coefficient α suggests the preference and sensitivity of an occupation to the large cities. When α is greater than 1, larger cities are more attractive for individuals with this occupation to migrate in compared with smaller cities. The representative occupations include financial managers, accountants, computer software developers, lawyers, physicians, etc.

Based on above findings, I further analyze what factors on characterizing occupation-migration patterns. I test 7 related factors including age, education level, industry coverage, occupational prestige, wage, average residential population and unemployment rate, and find that education level, industry coverage and average residential population significantly affect the migration patterns.

APPENDICES

Year	1980					1990				
City Size Bin	1	2	3	4	5	1	2	3	4	5
Managers & Administrators	45067.07	43303.92	43114.87	49875.69	47361.69	56875.01	51152.06	54121.01	53632.32	62683.41
Secretaries	19180.07	19138.56	20436.35	22001.12	23630.76	21726.58	20862.88	22320.28	23531.44	27103.46
Secretaries	19180.07	19138.56	20436.35	22001.12	23630.76	21726.58	20862.88	22320.28	23531.44	27103.46
Salespersons, n.e.c.	22968.15	22631.23	21379.62	24924.01	25957.55	25745.20	23412.56	25637.22	25742.41	28277.50
Cashiers	11113.95	11275.06	11348.69	12325.90	12688.87	10008.59	9679.12	9886.62	9597.76	12627.48
Truck, delivery, and tractor drivers	29357.74	28562.33	28949.87	31252.02	28260.25	26905.62	24425.35	24564.32	26138.26	26218.50
Primary school teachers	26885.98	26464.65	28662.31	30540.80	32605.65	31713.67	29615.35	31266.74	31615.20	35986.27
Janitors	16312.37	15682.73	16300.11	18637.40	21498.91	15767.98	14642.18	15208.67	16101.12	20007.48
Cooks, variously defined	12118.11	12533.59	13143.01	14519.53	17790.65	12406.28	12732.36	12979.81	12498.84	17110.64
Supervisors and proprietors of sales jobs	35377.23	34053.68	33370.52	36548.08	36537.63	40920.19	36941.69	38849.11	38717.92	46331.20

Year	1980					1990				
City Size Bin	1	2	3	4	5	1	2	3	4	5
Nursing aides, orderlies, and attendants	14191.63	13422.47	15248.89	16344.25	20598.31	16724.62	15557.53	16330.94	16034.15	20177.28
Retail sales clerks	15271.62	14461.70	11892.20	13739.67	15720.22	13169.10	12358.79	13518.56	13872.06	18894.36
Waiter/waitress	8512.94	9314.16	9423.36	11129.98	13789.07	10514.25	11058.53	12012.30	11669.58	15358.72
General office clerks	16701.96	16961.66	17890.93	18448.20	19368.11	17916.92	17248.89	18911.75	18312.15	21250.86
Registered nurses	26778.72	26384.20	28571.84	31366.96	32195.81	35399.53	35280.39	35426.86	36756.18	42897.81
Accountants and auditors	34061.26	33468.91	33499.58	36607.33	38080.76	38479.48	34816.04	38306.34	38562.82	45465.28
Laborers outside construction	19593.53	19566.14	20729.64	21319.46	20724.57	18950.73	17368.79	18152.09	18142.77	20147.44
Bookkeepers and accounting and auditing clerks	19846.33	19917.54	20501.74	21440.35	23188.40	21715.56	20787.88	22247.41	22154.70	25351.64
Customer service reps, investigators and adjusters, except insurance	26578.69	26670.92	27298.57	27874.00	30050.64	25727.28	24041.02	24847.10	24972.46	27241.64

Year	1980					1990				
City Size Bin	1	2	3	4	5	1	2	3	4	5
Assemblers of electrical equipment	22163.31	22008.45	23263.11	20229.15	18102.13	21351.15	19614.92	21995.12	18628.59	18364.50
Machine operators, n.e.c.	24323.32	23706.52	26483.70	24138.10	19301.46	24507.28	21192.20	24187.61	22989.48	19900.54
Office supervisors	34113.76	33609.33	34415.93	33884.23	35270.04	35336.66	32594.12	33631.75	34329.97	38994.56
Subject instructors (HS/college)	34123.96	32871.72	31774.92	35863.00	37769.06	34745.15	33421.22	35933.21	33880.49	42408.63
Housekeepers, maids, butlers, stewards, and lodging quarters cleaners	12205.53	11912.69	13046.54	13389.45	17702.96	12745.19	12409.40	12644.56	12957.82	16853.22
Construction laborers	20066.10	19278.50	20325.58	22361.27	24787.97	21197.38	19286.76	18123.04	20010.68	22388.69
Military	23107.70	22194.45	21401.21	22322.10	20419.39	24695.58	24282.95	27024.96	23794.60	26229.41
Teachers , n.e.c.	18750.55	18917.24	20036.24	22156.59	20537.87	23509.74	21946.20	21967.45	22434.83	23432.67
Computer systems analysts and computer scientists	44871.38	43878.46	47574.05	48568.59	48088.83	50576.65	44954.55	48617.91	48747.90	54491.91

Year	1980					1990				
City Size Bin	1	2	3	4	5	1	2	3	4	5
Stock and inventory clerks	20583.57	20624.27	21564.22	20392.49	18729.74	18921.56	18778.55	17493.97	18617.26	18563.79
Guards, watchmen, doorkeepers	20321.08	18025.60	18956.04	19808.77	22239.76	20682.93	18482.70	18946.51	18576.11	22479.30
Receptionists	12467.07	12722.99	13182.60	15061.72	15410.57	13503.41	13576.91	13956.91	14976.95	16286.20
Managers and specialists in marketing, advertising, and public relations	48637.47	45850.15	44378.77	51421.41	48973.55	57176.92	47494.42	52426.91	52377.80	61918.85
Lawyers	50196.21	50133.87	48868.99	58008.03	60146.99	76324.37	75752.21	83951.03	79878.19	94507.30
Production supervisors or foremen	41753.61	39159.79	39599.42	39508.75	35629.99	42794.59	37661.73	37599.05	37480.19	38445.17
Computer software developers	33841.13	33086.25	36737.64	37348.63	36940.97	42594.06	39569.90	40129.40	39904.05	48135.95
Misc food prep workers	7869.25	8109.08	9120.24	9414.42	13217.88	8944.57	9957.28	10519.00	10153.30	13516.50

Year	1980					1990				
City Size Bin	1	2	3	4	5	1	2	3	4	5
Child care workers	7554.65	7900.32	7635.25	9124.18	10224.05	7769.61	8353.80	8196.71	9082.01	11199.56
Carpenters	24801.88	24476.63	25198.56	28332.63	26977.75	25891.57	23275.83	23436.73	25045.55	29048.70
Social workers	26107.63	25597.05	27276.10	28151.24	28531.95	28285.39	27478.17	29717.86	28486.24	33065.10
Other financial specialists	37113.93	35590.21	36727.06	41038.61	46162.99	42559.10	41426.24	46826.57	46045.85	56060.71
Financial managers	46576.10	46365.28	45781.47	48982.98	51310.19	53976.37	55044.15	56111.36	56492.00	68881.15
Shipping and receiving clerks	25894.17	25328.93	26830.33	24899.99	26964.62	25869.25	23951.63	25069.23	25041.42	26958.14
Personnel, HR, training, and labor relations specialists	35341.52	33837.29	34484.08	35666.74	36833.62	38073.75	32704.65	36065.76	37891.35	42490.87
Gardeners and groundskeepers	12898.39	13979.75	15099.71	16021.92	18978.50	14465.29	13949.47	12709.13	13487.34	16417.60
Physicians	80280.38	77743.95	70312.31	77759.20	66579.73	116196.72	100503.90	101971.51	84452.57	96630.40
Textile sewing machine operators	13744.42	13207.75	14408.09	13564.98	13113.84	14047.71	13316.06	14028.68	14313.29	11950.89

Year	1980					1990				
City Size Bin	1	2	3	4	5	1	2	3	4	5
Automobile mechanics	28321.77	26608.83	28115.18	28207.71	25852.58	30047.25	24373.83	26197.80	23956.21	27051.13
Police, detectives, and private investigators	41954.25	39432.94	43350.50	46574.43	47816.17	47785.79	42399.38	42753.57	45444.75	53877.52
Designers	26346.27	26438.56	27195.96	33779.78	38881.49	31787.43	28101.79	31958.09	32028.94	42861.23
Administrative support jobs, n.e.c.	24256.69	22595.11	22798.20	25754.13	26584.91	26081.05	23853.43	25263.89	24044.01	29786.76
Data entry keyers	19049.80	19700.81	19760.88	20508.46	21592.20	18798.95	17320.15	18023.15	19025.82	21599.30

TABLE XIV

CITY SIZE BIN AND AVERAGE WAGE FOR OTHER OCCUPATIONS (YEARS 1980 AND 1990)

Year	2000					2010				
City Size Bin	1	2	3	4	5	1	2	3	4	5
Managers and administrators, n.e.c.	60945.12	58116.49	60822.31	62347.32	66080.04	63587.52	60232.37	67516.37	66295.52	71492.91
Secretaries	23393.33	22835.26	24679.93	25824.84	28719.69	23394.10	22457.86	25099.92	24253.02	29059.65
Salespersons, n.e.c.	37605.80	36832.74	38821.72	38297.07	40504.01	37571.73	40712.74	48394.58	49324.19	44364.51
Cashiers	10639.12	10825.77	11709.84	11756.72	12585.21	9221.19	9704.61	10235.69	9505.28	12459.07
Truck, delivery, and tractor drivers	27347.19	26301.56	27026.33	26849.96	25853.79	23706.18	24635.71	23331.11	23716.08	23039.54
Primary school teachers	32576.37	31674.57	31680.77	32180.38	36707.31	35130.69	32279.67	33592.48	32745.09	41977.55
Janitors	17016.77	16684.40	17450.09	16551.18	20837.84	14560.06	14703.22	14843.77	14543.91	18593.27
Cooks, variously defined	15456.36	15290.39	16750.81	16400.46	19372.14	13567.90	14461.37	15127.98	15108.43	17899.17
Supervisors and proprietors of sales jobs	38687.45	37303.65	41338.52	37330.30	42238.81	36591.66	39321.64	41372.65	42797.95	43325.63

Year	2000					2010				
City Size Bin	1	2	3	4	5	1	2	3	4	5
Nursing aides, orderlies, and attendants	17232.12	16313.67	18574.42	18343.90	20742.09	15182.12	14800.16	15147.60	14745.63	17499.24
Retail sales clerks	19206.36	18563.16	20219.53	19914.73	20190.72	16634.42	17367.24	21208.77	16811.09	18448.54
Waiter/waitress	10994.12	11845.57	13102.87	13227.00	16171.47	9940.45	10865.30	11805.52	11583.31	15562.33
General office clerks	17816.12	17740.56	18650.00	19108.36	20205.07	19645.73	18455.74	19379.90	19484.73	21338.24
Registered nurses	40123.61	38980.33	40374.94	40808.36	51173.08	46717.37	45061.16	48705.01	47732.37	56136.46
Accountants and auditors	42758.28	39778.56	40698.67	44800.72	50637.39	47096.72	44222.59	50419.93	50822.84	65086.76
Laborers outside construction	18911.50	18016.81	18416.32	18039.19	20977.69	16261.25	14791.01	14019.19	15304.04	18725.64
Bookkeepers and accounting and auditing clerks	23254.83	22886.96	24658.74	25214.67	27501.58	25325.86	24447.16	24412.26	26418.98	29514.56
Customer service reps, investigators and adjusters, except insurance	22404.16	22288.37	24318.17	23565.63	25394.08	19734.83	20379.11	21503.73	21812.59	23931.88

Year	2000					2010				
City Size Bin	1	2	3	4	5	1	2	3	4	5
Assemblers of electrical equipment	21901.18	22079.41	25172.15	18309.13	18860.59	18637.25	20905.57	20187.76	17600.39	17829.44
Machine operators, n.e.c.	22765.50	22045.53	23350.86	23116.25	20055.94	22004.62	19772.89	20484.11	22571.98	21157.33
Office supervisors	34553.38	33385.15	35149.53	34749.27	37530.02	36865.40	36420.80	37457.71	35706.73	42848.19
Subject instructors (HS/college)	36061.71	36251.05	37156.73	40028.51	43645.90	37160.53	36761.23	37927.88	41661.86	46967.93
Housekeepers, maids, butlers, stewards, and lodging quarters cleaners	13305.16	13363.48	14678.82	13230.69	16009.59	11687.00	12348.94	12750.01	12128.43	15350.52
Construction laborers	22176.15	22194.25	20137.33	22128.88	22064.61	19966.75	18525.87	17697.43	19494.64	22303.64
Military	27598.30	26573.56	25578.48	32856.92	30889.75	33875.86	36125.54	30101.89	28988.08	36163.22
Teachers , n.e.c.	14267.65	14773.65	16497.39	16671.32	16154.97	15825.32	17516.61	19261.05	18554.06	19294.04
Computer systems analysts and computer scientists	43655.68	42925.52	45323.05	46167.40	50724.58	46862.87	43894.12	48224.18	51049.30	51867.70

Year	2000					2010				
City Size Bin	1	2	3	4	5	1	2	3	4	5
Stock and inventory clerks	16747.19	16973.46	16765.24	17275.49	16389.87	14603.96	14232.79	15454.06	14208.58	15088.58
Guards, watchmen, doorkeepers	20727.82	19930.22	20966.65	20851.50	22432.85	19398.92	19439.40	19696.20	20298.54	22378.06
Receptionists	14468.19	14658.66	15644.01	16859.29	17466.76	14621.20	14163.30	15305.00	15178.83	16184.02
Managers and specialists in marketing, advertising, and public relations	57141.49	54099.03	58409.82	61581.75	65419.06	54586.05	58812.58	67322.22	69459.35	72689.92
Lawyers	90133.33	89762.87	100443.30	95285.18	118760.36	105813.63	93143.08	106140.37	105892.56	144677.78
Production supervisors or foremen	39302.83	36923.34	36191.38	34878.81	36624.82	41528.55	36119.14	36774.03	39459.65	35129.51
Computer software developers	56252.55	51283.65	56730.81	52182.88	59389.61	62297.48	58450.26	62589.91	59359.27	70503.25
Misc food prep workers	10449.59	10365.48	10512.47	10904.06	13301.35	8787.02	8656.66	9761.32	9439.97	11549.91

Year	2000					2010				
City Size Bin	1	2	3	4	5	1	2	3	4	5
Child care workers	10359.89	11086.57	12543.00	13189.20	14284.93	8487.35	8720.33	10012.58	10590.79	12559.40
Carpenters	25844.59	24109.80	24311.09	25431.21	28476.60	23155.12	20919.56	21669.95	23226.00	27975.68
Social workers	29699.31	29679.86	30053.39	29783.85	33684.97	31939.35	30480.94	30922.27	31811.55	36568.26
Other financial specialists	49442.14	48094.04	58437.80	53607.54	79310.32	53389.99	50095.79	68426.60	61468.80	95552.45
Financial managers	59825.07	59136.48	63393.47	62822.17	87233.85	63928.89	59916.46	77976.24	69304.20	94964.34
Shipping and receiving clerks	21976.28	21039.93	22077.22	21482.55	20082.65	19716.84	19445.28	19644.19	19040.67	20576.08
Personnel, HR, training, and labor relations specialists	38500.37	36535.57	41704.72	43078.31	49241.91	42071.71	40101.20	43210.46	42389.49	47967.41
Gardeners and groundskeepers	16037.73	15484.65	15651.85	15519.84	17821.22	14375.48	13615.35	11472.39	12491.18	15122.41
Physicians	127744.9	128481.2	109410.7	101287.2	112335.8	127100.1	133535.6	113274.2	119263.8	109080.7
Textile sewing machine operators	15134.01	16089.29	22592.61	16337.59	12256.54	13582.54	13077.23	14322.55	13363.22	12666.84

Year	2000					2010				
City Size Bin	1	2	3	4	5	1	2	3	4	5
Automobile mechanics	28151.89	26186.39	26989.37	25438.81	24933.53	24884.27	26596.88	22882.37	22134.87	24229.18
Police, detectives, and private investigators	48847.50	44949.12	44822.64	48290.83	48244.41	52400.15	48181.49	44856.98	52399.71	53457.60
Designers	32820.65	31985.81	35825.99	34676.73	44830.60	31609.53	31698.92	39420.86	35644.54	45779.25
Administrative support jobs, n.e.c.	27265.67	24951.25	27255.39	26165.26	32664.70	25537.22	26442.27	24981.42	27832.38	29610.71
Data entry keyers	17333.84	18083.82	18189.27	18836.77	19562.35	20319.95	17591.09	18131.32	19268.54	24496.11

TABLE XV

CITY SIZE BIN AND AVERAGE WAGE FOR OTHER OCCUPATIONS (YEARS 2000 AND 2010)

CITED LITERATURE

Arnott, R.: Optimal city size in a spatial economy. Journal of Urban Economics, 6(1):65–89, 1979.

Athey, S. and Imbens, G.: Identification and inference in nonlinear difference-in-differences models, 2002.

Baum-Snow, N. and Pavan, R.: Understanding the city size wage gap. The Review of economic studies, 79(1):88–127, 2012.

Baum-Snow, N. and Pavan, R.: Inequality and city size. Review of Economics and Statistics, 95(5):1535–1548, 2013.

Borjas, G. J., Bronars, S. G., and Trejo, S. J.: Self-selection and internal migration in the united states. Journal of Urban Economics, 32(2):159–185, 1992.

Bowles, S.: Migration as investment: Empirical tests of the human investment approach to geographical mobility. The Review of Economics and Statistics, pages 356–362, 1970.

Chen, Y. and Rosenthal, S. S.: Local amenities and life-cycle migration: Do people move for jobs or fun? Journal of Urban Economics, 64(3):519–537, 2008.

Chiswick, B. R. and Miller, P. W.: The determinants of post-immigration investments in education. Economics of education review, 13(2):163–177, 1994.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S.: The productivity advantages of large cities: Distinguishing agglomeration from firm selection. Econometrica, 80(6):2543–2594, 2012.

Combes, P.-P. and Gobillon, L.: The empirics of agglomeration economies. 2014.

Davis, D. R. and Dingel, J. I.: The comparative advantage of cities. Technical report, National Bureau of Economic Research, 2014.

De la Roca, J. and Puga, D.: Learning by working in big cities. 2012.

Dustmann, C., Glitz, A., et al.: Migration and education. Handbook of the Economics of Education, 4:327–439, 2011.

Echeverri-Carroll, E. L. and Ayala, S. G.: Urban wages: Does city size matter? Urban Studies, 48(2):253–271, 2011.

Flood, S., King, M., Ruggles, S., and Warren, J.: Integrated public use microdata series, current population survey: Version 4.0 [machine-readable database]. Minneapolis: University of Minnesota, 2015.

Gabaix, X. and Ioannides, Y. M.: The evolution of city size distributions. Handbook of regional and urban economics, 4:2341–2378, 2004.

Glaeser, E.: Mare, d., 2001, cities and skills. J. LAB. ECON., 19:316, 2001.

Glaeser, E. L. and Resseger, M. G.: The complementarity between cities and skills*. Journal of Regional Science, 50(1):221–244, 2010.

Harvey, D.: Reshaping economic geography: the world development report 2009. Development and change, 40(6):1269–1277, 2009.

Hoover, E. M.: Location theory and the shoe leather industries, volume 55. Harvard University Press, 1937.

Karlsson, C.: Handbook of research on cluster theory, volume 1. Edward Elgar Publishing, 2010.

Khan, A. H.: Post-migration investment in education by immigrants in the united states. The Quarterly Review of Economics and Finance, 37:285–313, 1997.

Lemieux, T.: Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? The American Economic Review, pages 461–498, 2006.

Long, L. H.: Migration differentials by education and occupation: Trends and variations. Demography, 10(2):243–258, 1973.

Marshall, A.: Principles of economics, 8-(edition, 1920.

McCann, P.: Urban and regional economics. OUP Catalogue, 2001.

Melo, P. C., Graham, D. J., and Noland, R. B.: A meta-analysis of estimates of urban agglomeration economies. Regional science and urban Economics, 39(3):332–342, 2009.

Meyer, P. B. and Osborne, A. M.: Proposed category system for 1960-2000 census occupations. US Department of Labor, Bureau of Labor Statistics, Office of Productivity and Technology, 2005.

Mueser, P. R. and White, M. J.: Explaining the association between rates of in-migration and out-migration. In Papers of the Regional Science Association, volume 67, pages 121–134. Springer, 1989.

Nam, C. B. and Boyd, M.: Occupational status in 2000; over a century of census-based measurement. Population Research and Policy Review, 23(4):327–358, 2004.

O'Hagan, S. B. et al.: In-migration and return migration to cities in northern ontario, canada: Benefits by city size in the context of today?s knowledge economy. Journal of Population and Social Studies, 22(1):87–100, 2014.

Ohlin, B.: International and interregional trade. Harvard Economic Studies, Cambridge, MA, 1933.

Quinn, M. A. and Rubb, S.: The importance of education-occupation matching in migration decisions. Demography, 42(1):153–167, 2005.

Rao, N.: Migration, education and socio-economic mobility. 2010.

Rappaport, J.: Moving to nice weather. Regional Science and Urban Economics, 37(3):375–398, 2007.

Rosenbloom, J. L. and Sundstrom, W. A.: The decline and rise of interstate migration in the united states: Evidence from the ipums, 1850-1990. Research in economic history, 22:289–325, 2004.

Rosenthal, S. S. and Strange, W. C.: Evidence on the nature and sources of agglomeration economies. Handbook of regional and urban economics, 4:2119–2171, 2004.

Ruggles, S., Genadek, K., Goeken, R., Grover, J., and Sobek, M.: Integrated public use microdata series: Version 6.0. Technical report, 2015.

Walker, R., Ellis, M., and Barff, R.: Linked migration systems: Immigration and internal labor flows in the united states. Economic Geography, pages 234–248, 1992.

Weber, A.: Alfred weber's theory of the location of industries, 1909.

VITA

NAME: Zhao Gu

EDUCATION:

B.S., Economics, Beijing University of Technology, China, 2007.

Second Major, Modern Management and Public Relations, Beijing University of Technology, China, 2007.

M.S., Statistics, Acadia University, Canada, 2009.

Ph.D., Economics, University of Illinois at Chicago, Illinois, 2016.

TEACHING:

Department of Economics, University of Illinois at Chicago, various economics courses, 2009-2014

HONORS:

Teaching Assistantship, University of Illinois at Chicago, 2009-2014