

ESSAYS ON MIGRATION

BY

XIAOYING XIE

DISSERTATION

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 Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Associate Professor Darren Lubotsky, Chair  
Professor Daniel McMillen  
Associate Professor David Albouy  
Assistant Professor Tatyana Deryugina

## **Abstract**

Migration is one of the main forces shaping our society as we know it. Focusing on the determinants of migration and its influence on local communities, this dissertation consists of three chapters. Chapter 1 provides a brief introduction to the thesis, covering the motivation for the research, the methodologies used, and policy implications.

Chapter 2 estimates the impact of education on two key outcomes: migration probability and distance. Migration greatly affects the regional economy, and hence, the out-migration of highly educated workers has raised serious concerns for regional development. The OLS estimator indicates a small but positive effect of education on both outcomes, which is similar to other studies. However, using compulsory schooling law changes as an instrumental variable, the 2SLS estimator suggests that education increases migration distance but decreases the probability to migrate. To guide the analysis, this paper expands the basic migration model to include distance as another element in people's decisions. The intuition is that by searching broader distances, people could obtain higher expected incomes, but must also pay higher costs. The overall effect of education on migration is determined by the trade-off between the cost and benefits of migrating longer distances.

Chapter 3 estimates the influence of immigration on local housing prices. Housing price is crucial to people's well-being, as it not only affects their living conditions, but also affects homeowners' investment values. Both the OLS and 2SLS results suggest that on average immigration has a slight positive effect on housing prices. However, if we use quantile regression, we observe quite significant but heterogeneous effects of different neighborhoods. For census tracts with expensive housing, immigrants increase housing prices. For census tracts with cheap housing,

immigrants reduce housing prices. Lower housing prices make housing more affordable for tenants, but reduces homeowners' total wealth. We also look at possible sources of heterogeneity from both the supply side and demand side. In poor neighborhoods, for example, immigrants might drive natives to neighborhoods with better amenities while increasing housing supply in the area, hence reducing housing prices.

## Acknowledgements

There are many people deserved to be thanked.

For my Ph.D. (and general life and career) education, I'd like to thank my advisor, Darren Lubotsky. He is definitely an economist that I try to model myself after. Thanks also go to Dan McMillen, David Albouy and Tatyana Deryugina. They are great mentors and have given me much helpful advice. I am so appreciative to all of the professors in our department, who give me support and advice throughout the years. Thanks to Daniel Dias for his help and guidance. Finally, there are my coauthors, Cong and Dongwoo. It has been fun and inspiring to work with them.

There were a lot of people who helped me live a rich and full filling life in UI. The most important is my family, especially my mom. Being the most strong, optimistic and diligent woman I know of, she has always been my role model. There are the people who have been there for me for fun and everything else, like Vera, Yingyi and Lisa. Then there's Catrin, who supported me and believed in me unconditionally.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>How is Immigration Affected by Education</b>	<b>6</b>
2.1	Introduction . . . . .	6
2.2	Adding Distance to the Conceptual Framework . . . . .	10
2.3	Data and Variables . . . . .	14
2.4	Identification Strategy and Specification . . . . .	16
2.5	OLS results . . . . .	21
2.6	The Effect of Compulsory Schooling Laws on Education . . . . .	22
2.6.1	Visual Evidence: The Survival Function of Schooling . . . . .	22
2.6.2	The Impact of Compulsory Schooling Laws on State Education . . . . .	24
2.6.3	First Stage: The Effect of Compulsory Schooling Laws on Education . . . . .	25
2.7	The Impact of Education on Migration . . . . .	26
2.7.1	The Impact of Education on Migration Distance . . . . .	26
2.7.2	The Impact of Education on Probability of Migration . . . . .	29
2.7.3	Alternative Specification . . . . .	33

2.8	Using Quarter of Birth as Instrument Variables . . . . .	35
2.9	Discussion . . . . .	37
<b>3</b>	<b>The Heterogeneous Effect of Immigration on Housing Price</b>	<b>44</b>
3.1	Introduction . . . . .	44
3.2	Data . . . . .	46
3.3	Quantile Regression and Heterogeneity . . . . .	47
3.3.1	OLS and IV . . . . .	47
3.3.2	Quantile and Two Stage Quantile Estimation . . . . .	50
3.4	Estimation Details and Empirical Results . . . . .	52
3.4.1	OLS and IV . . . . .	52
3.4.2	Quantile and Instrumental Quantile Estimation . . . . .	54
3.5	Further Results: The Change Of Housing Supply . . . . .	56
3.6	Conclusion . . . . .	58
	<b>Bibliography</b>	<b>65</b>

# Chapter 1

## Introduction

Migration is one of the main forces shaping today's society. History has shown that migration is not only important for the transfer of manpower, but also important for the subsequent transfer of culture and skill, hence providing the knowledge and innovation needed for global growth. In the past decade, for example, China experienced what was probably the most extensive internal migration in human history. By 2011, more than 158 million rural workers left their hometown to work in factories in east coast cities. They provide the essential workforce for the development of the Chinese manufacturing industry and have changed the modern world economy as a whole. Migration has also provoked intense debates about social justice, equality, and political reform in China. Since migration has a great impact on both the regions of origin and destination, this dissertation focuses particularly on the determinants of migration and its influences on the destination society.

The second chapter uses the U.S. Compulsory Schooling Law changes as an instrumental variable to study the impact of education on two migration outcomes: the likelihood to migrate and how far to migrate within the U.S. in the early 20th century. My analysis indicates that

education reduces people’s likelihood to migrate out of their birth states, but leads them to reside in states that are farther away if they decide to move.

The reason why I am interested in this topic is that the out-migration of highly-skilled workers has long been considered detrimental to regional development, as suggested in the “Brain Drain” literature. If more schooling makes people more likely to move out of the region, then the local education policy might have a limited effect on regional development and equality. To correctly evaluate the return to local educational investment and to determine its optimal level, we need to understand how it would affect individuals’ migration behavior. Thus, in my job market paper, I focus on estimating the impact of education on both migration likelihood and distance.

Since the goal of the paper is to estimate the causal impact of education, a credible identification strategy is key. It is well known that education is correlated with many personal characteristics, which might also affect migration. Such factors include people’s motivation, health, social networks, and abilities among others. In this paper, I use Compulsory Schooling Law changes in the early 20th century as an instrument to isolate the effect of education from those factors. In the past century, the U.S. experienced a series of changes in compulsory schooling law legislation. These changes are exogenous to individual characteristics and had a significant effect on people’s education outcomes. Hence, they are widely used as instruments when studying the impact of education.

While OLS estimators suggest a positive correlation between education and migration, 2SLS estimators indicate a different story. First, contrary to OLS, 2SLS estimator suggests that one year of schooling decreases the probability of migrating out of one’s birth state by 6.6%. This results challenge the conventional view that education would necessarily increase one’s



likelihood to migrate, as suggested in the “Brain Drain” literature. Second, similar to OLS, 2SLS estimator implies that an additional year of schooling influences people to migrate to a state that is 23.3% farther. To the best of my knowledge, this paper is the first to estimate this causal relationship.

To correctly identify the coefficients, I include a comprehensive set of controls in the econometric model. Except personal characteristics such as one’s age, square of age and gender, I also included two sets of fixed effects. First, I include cohort fixed effect to control for nation-wide shocks that affect all states similarly. Second, I include birth state fixed effects to control for state specific amenities such as the weather, the size and the location of the states. Adding state-specific time variant variables are important to separate the effect of improved education from other state trends. Hence I also include linear birth state trends and state-level characteristics, such as state average annual wage in the manufacturing industry and the percentage of urban population.

To guide the empirical analysis, I include distance as another element in people’s migration decisions. Distance is an important aspect of migration. It provides a useful proxy for migration cost, and it determines the range of the spillover effect (that is, the range within which local education policies could affect neighboring communities). My assumption is that before migration, people have to choose a distance within which they will search for jobs. This searching distance determines their expected migration distance as well as their expected income. This is because if people search in a wider radius, they will find more job opportunities lead to a higher expected income. However, they must also pay higher costs. If schooling changes people’s searching distance, people with different education levels would be facing different migration costs. Hence, they might have different likelihoods to migrate. By adding a simple component,

this model is able to generate both positive and negative selection of migrants with respect to education in a more generalized way.

The main conclusion of this paper is that education does have a significant impact on migration outcomes. The conceptual framework suggests that the negative impact of education on migration probability indicates a high marginal cost of migration distance. This paper contributes to the literature in three ways. First, to design sustainable policies and reduce regional inequality, we need to understand what factors cause people to move. If increased education causes more talented people to stay, policies investing in education might have a larger effect on local development due to this overlooked channel. Second, understanding the determinants of migration is important if we are to treat the selection bias it causes. Third, when evaluating the overall effect of education at the individual level, we should also consider its influence on geographic mobility, which offers people wider range of location and job choices, allowing them to reside in suitable neighborhoods and to move to better opportunities after a local economic downturn.

In the third chapter, I use instrumental quantile estimation to study how immigration affects housing prices. The housing market is crucial to people's well-being. Expenses on housing not only affect living conditions, but also affect homeowners' investment values. Only recently, did researchers start to analyze how immigration affects the price of this particular commodity (Saiz (2003); Saiz & Wachter (2011)). However, most of the studies focus on the average effect, overlooking the fact that different neighborhoods could react to immigration differently. Immigrants may affect neighborhoods in multiple ways. For example, they can increase cultural diversity, which could drive up local housing demand or change local crime rates and local average education levels. The local amenity externality can either bring housing price up or

down depending on both the characteristics of the immigrants and local communities. Since different neighborhoods attract different immigrants, the induced housing price changes should not be the same. We thus analyze the neighborhood heterogeneous impacts of immigration on housing price. Specifically, in this paper, we use quantile regression, introduced in Koenker and Bassett (1978), to analyze how immigration affects different types of neighborhoods differently when it comes to the housing market. In other words, we not only ask the question, "Does immigration matter?", but also consider the question, "For whom does immigration matter?"

Though the average effect of immigration is to increase housing prices, quantile regression suggests quite heterogeneous effects of different areas. The OLS estimation suggests that the share of immigrants in the neighborhood has little effect on the housing price, while the 2SLS estimation suggests that a 1% increase in the share of immigrants increases housing prices by 0.2%. However, the quantile regression suggests that the impacts of immigrants are quite different for different neighborhoods, hence looking at the average effect alone will lead to a vastly underestimated effect. For example, at the 20th percentile of the housing price distribution, a 1% increase in the share of immigrants will lead to a 1.3% decrease in housing prices. Whereas at the 80th percentile of the housing price distribution, a 1% increase in the share of immigrants will increase the housing prices by 1%. This result suggests that when we discuss immigration policy, we have to keep in mind that it might affect different sub-populations quite differently. We need to first understand who benefits from policy changes and who pays the cost, then decide the best policy to improve the overall welfare. The results of immigrants' heterogeneous effects make homogeneous immigration policy inappropriate. This paper can then provide new guidelines for policy makers to better regulate immigrants.

## Chapter 2

# How is Immigration Affected by Education

### 2.1 Introduction

Education is a surprisingly strong predictor of migration. People with more education migrate more frequently and move longer distances (Todaro (1980); Lucas (1997); Greenwood (1997)). This relationship is of particular interest to policy makers, since the out-migration of highly educated workers has long been considered detrimental for regional development. To correctly evaluate the return to local educational policy, it is necessary to determine how it would affect different individual outcomes, including their likelihood to migrate and how far they move. However, we know relatively little about the causal effects of education on migration, even less about the mechanisms behind it.

To identify the impact of education, this paper utilizes Compulsory Schooling Law (CSL) changes in the early 20th century to disentangle the effect of education from other confound-

ing factors. The U.S. experienced a series of changes in compulsory schooling law legislation in the past century. These changes are widely used as instruments when studying the impact of education, due to their exogeneity to individual characteristics and their significant effects on personal education outcomes (Lleras-Muney (2005); Oreopoulos (2006); Pischke and Von Wachter (2008); Mazumder (2011)). The first-stage F-statistics suggests that the CSL changes have sufficient power to consistently estimate the effect of education. The second-stage results indicates that more education increases people’s migration distance if they decide to move, but reduces their mobility.

The empirical evidence suggests that education does have a positive impact on migration distance. To the best of my knowledge, this paper is the first to estimate this causal relationship. Both OLS and 2SLS suggest a positive impact of education, though 2SLS increases the estimated effect of education by a factor of ten (2.1% v.s. 23.3%). The 2SLS estimator implies that an additional year of schooling raises migration distance by 23.3% for individuals who migrate. This effect is relatively large, indicating that one year of education causes people to move an extra 162 miles, which is roughly the distance between New York City and Boston.

The 2SLS estimator also suggests that education decreases people’s probability of migration, which is the opposite of what OLS suggests. Estimators in the literature vary (Malamud and Wozniak (2012); Machin et al. (2012); McHenry (2013)). The negative effect suggested by the 2SLS estimator in this paper is consistent with the results in McHenry (2013)<sup>1</sup>. However, when using people’s draft avoidance behavior during the Vietnam War as an instrument, Malamud and Wozniak (2012) find that education increases people’s migration probability. The difference

---

<sup>1</sup>McHenry (2013) focuses on migration probability only. It uses CSLs as IV but focuses on different cohorts. Its main results are not robust after including state trends.

might indicate different local average treatment effects at different education levels, since the Vietnam War draft mainly affected college education, while the compulsory schooling laws mainly affected primary and high school education.

To guide the empirical analysis, this paper augments the conceptual framework to include migration distance. Distance is an important part of migration decisions (Schwartz (1973); Magrini and Lemistre (2013)), but is rarely incorporated into the migration model. To demonstrate how education might affect migration, this paper assumes people simultaneously choose whether to migrate and how far to search for jobs. Before migration, I assume people have to choose a distance within which they will search for jobs and migrate. If they search in a wider range, they have a higher likelihood of finding better jobs, but they must pay higher costs. People with different education levels search in different distances, and hence face different relocation costs. The costs they face will in return affect their likelihood to migrate. The model shows that the trade-off between the gain of migrating longer distances and the cost of it are important in determining the overall effect of education on migration. By adding a simple component, this model is able to generate both positive and negative selection of migrants with respect to education without assuming different wage distributions in different locations as in Borjas (1999).

This paper contributes to the literature studying the determinants of migration. Migration usually poses a strong influence on the local development at both the destination and source areas. To design sustainable policies and reduce regional inequality, we need to understand what factors cause people to move. If increased education causes more talented people to stay, policies investing in education might have a larger effect on local development due to this overlooked channel. Also, understanding the determinants of migration is important if we are to treat the

selection bias it causes. For example, Dahl (2002) suggests that to correctly estimate the return to education, one needs to compensate individual's likelihood to migrate. This essentially means estimating migration probability based on people's observable characteristics.

This study also contributes to the literature examining the impact of education. My results suggest that education does have a significant impact on migration. Therefore, when evaluating the overall effect of education at the individual level, we should also consider its influence on geographic mobility, which offers people wider range of location and job choices, allows them to reside in suitable neighborhoods, and allows them to move to better opportunities after a local economic downturn (Bound and Holzer (2000)).

Developing world could benefit from analyzing historical data in the U.S.. In the early 20th century, the U.S. shared a lot of similarities with modern developing countries. For example, in 1940, the urbanization rate in the U.S. was 56.5%<sup>2</sup>, which is very similar to the rate of 52.6% in China in 2012<sup>3</sup>. Also, in 1940, more than 30.3% of American worked in the service industry and 23.4% worked in the manufacturing industry. Those numbers are 36.1% and 30.3% in China in 2012<sup>4</sup>. The out-migration of highly educated workers, which could loosely referred to as "brain drain," have raised serious concern for developing regions. The negative causal impact observed in this study suggests that under certain conditions, local educational programs could encourage more highly educated workers to stay and may be more beneficial to local development than people originally believed.

---

<sup>2</sup>Data from china.org.cn

<sup>3</sup>Data from US Census, 1940

<sup>4</sup>Data from Ministry of human resources and social security of the People's Republic of China

## 2.2 Adding Distance to the Conceptual Framework

To guide the empirical analysis, I augment a simple conceptual framework to include distance as another choice variable. When considering migration, people are usually faced with two decisions: whether to migrate and where to migrate. I simplify the second decision as the choice of how far to migrate. The model shows that the trade-off between the gain of migrating a long distance and the cost of it is important in determining the overall effect of education on migration.

In most basic migration models, the only decision people make is whether to migrate after calculating the utility gain from migration (Borjas (1987), (1999); Chiquiar and Hanson (2005)). In Borjas (1987), the decision to migrate is further simplified by comparing the expected income at the current state and the destination state. Assume at current state 0, an individual faces the following income:

$$\omega_0 = \mu_0 + \beta_0 E, \tag{2.1}$$

where  $\omega_0$  is total income,  $\mu_0$  is base income,  $\beta_0$  is return to education, and  $E$  is the education level. If he migrates to destination state 1, he faces a new income  $\omega_1$ , and must pay the cost of migration  $\mu_c$ . The individual will decide to move if the net gain from migration ( $f$ ) is positive, defined as following:

$$f = \omega_1 - \omega_0 - \mu_c. \tag{2.2}$$

In reality, people not only decide whether to migrate or not, but also decide where to mi-



grate. Migration and destination decisions are usually made simultaneously. It is unlikely that someone would pack up the whole house without knowing where he is heading. To incorporate the location choice into the model, I simplified the geography of destination states using its distance from the current state ( $D$ ). This is essentially assuming that states with different distances are different destination choices, since they may represent totally different cultures, job opportunities and accessibility to social networks and information.

In the augmented model, I assume migration cost is an increasing function of distance, since it has long been considered a serious deterrence to migration in economics literature. Greenwood (1975) summarized three types of cost that could be proxied by distance. First of all, distance directly determines the transportation cost of migration. Secondly, distance directly links to the psychic cost of migration, such as being far away from family members or losing one's social network. Lastly, when distance increases, the availability of information decreases and uncertainty increases. People face higher costs if they want to obtain job information from states far away. Based on these reasons, I assume the migration cost is an increasing function of distance with a constant marginal cost,

$$\mu_c = \delta_c D.$$

In the model, I also assume future income is positively correlated with distance. One apparent advantage of migrating a longer distance is that people can reach out to outside labor markets and hence have more location choices and job opportunities. More job opportunities entail a better matching quality of final employment and higher future wages. Consider a simple scenario in which forward looking workers are searching for new jobs and their expected

migration distance is determined by their searching radius. If they look for jobs in nearby regions only, their expected income increase might be small since there might be limited jobs available. However, if they look for jobs in a broader region, their expected income increase should be higher because now they are exposed to more opportunities. Therefore, their expected incomes should be positively correlated with their expected migration distance.

At the same time, people with different education levels might be facing different job distributions, such as different average incomes and different income variations. For example, less-educated workers might find that jobs are quite similar regardless of where they locate. Hence their marginal return to researching distance might be lower than the highly-educated workers. In this paper, I use  $\tau(E)$  to capture the effect of education on future income. Based on the above rationale, I assume the expected wage after migration is a function of distance and their own education; specifically,

$$\omega_1 = D^\alpha \tau(E),$$

where  $\alpha > 0$ . In other words, I assume that the more people venture out of their current locations and look for jobs, the higher the potential future wages are. The marginal return to searching distance  $MRD = \frac{\partial \omega_1}{\partial D} = \alpha D^{\alpha-1} \tau(E)$  is different for people with different education levels. When  $\frac{\partial \tau}{\partial E} > 0$ ,  $MRD$  increases as education level increases. When  $\frac{\partial \tau}{\partial E} < 0$ ,  $MRD$  decreases as education increases.

In incorporating the cost and wage structure, the gain from migration could be rewritten as the following:

$$f = D^\alpha(\tau(E)) - (\mu_0 + \beta_0 E) - \delta_c D. \quad (2.3)$$

The optimal distance is chosen by analyzing the cost benefit trade-off of distance. The first order condition of Equation (2.3) with respect to  $D$  requires that the marginal cost ( $\delta_c$ ) is equal to the marginal benefit ( $\alpha D^{\alpha-1} \tau$ ) at the optimal distance. Solving for the optimal distance gives

$$D^* = [\alpha / \delta_c * \tau(E)]^{1/(1-\alpha)}.$$

When  $0 < \alpha < 1$ , the second order condition is satisfied and the optimal distance exists. Taking the derivative of  $D^*$  w.r.t.  $E$ , we can see that when  $\frac{\partial \tau}{\partial E} > 0$ , education increases optimal distance; when  $\frac{\partial \tau}{\partial E} < 0$ , education decreases optimal distance. Recall that  $\frac{\partial \tau}{\partial E}$  determines how  $MRD$  changes when education level changes. The model implies that if less-educated workers face a lower  $MRD$  due to reasons such as a more homogeneous labor market, they will search and migrate a shorter distance. However, if they face a higher  $MRD$ , they will migrate a longer distance.

Knowing their optimal distance  $D^*$  if they search and move, potential migrants will chose whether to migrate or not based on their net gain of migration,  $f(D^*)$ . The effect of their own education on migration benefit  $\frac{\partial f(D^*)}{\partial E}$  is determined by the structure of migration cost and return to distance. Recall that the optimal distance is  $D^* = [\alpha / \delta_c * \tau(E)]^{1/(1-\alpha)}$ . After plugging  $D^*$  into Equation (2.3), we obtain the function of net gain of migration  $f(D^*)$ . The first derivative of  $f(D^*)$  with respect to education determines how education affects people's decisions to migrate. Taking the derivative of the benefit function with respect to education

level could yield the marginal effect of education, which is

$$\frac{\partial f(D^*)}{\partial E} = \left(\frac{\alpha\tau}{\delta}\right)^{\frac{\alpha}{1-\alpha}} \frac{\partial \tau}{\partial E} - \beta_0.$$

In most scenarios, the education groups that have higher return to searching distance will migrate further and also more frequently. However, in some special cases, such as when  $\frac{\partial \tau}{\partial E} > 0$  and when the marginal cost is sufficiently large,  $\frac{\partial f(D^*)}{\partial E}$  could be negative while  $\frac{\partial D^*}{\partial E}$  is positive, indicating that highly-educated workers migrate longer distances but are less likely to move.

An example might help elucidate the mechanisms for the special case. Consider a computer developer and a construction worker. Assume that the offered incomes vary more for the developer than for the construction worker in different locations. Therefore, the developer has a higher incentive to conduct a nation-wide, sometimes even global, search before she switches job and migrates. Now knowing the long-distance searching cost and migration cost, the developer might not want to migrate frequently. On the other hand, the construction worker is more likely to look for jobs in nearby regions. Since he does not need to pay the long-distance searching and migration cost, he might migrate more frequently.

## 2.3 Data and Variables

The analysis in this section mainly relies on the Integrated Public Use Microdata Survey (IPUMS) by the U.S. Census from 1940-1960. Data with missing values are omitted. Cohorts were chosen such that the CSL reform has a great influence on their education outcome. In the sample, the highest age is set to be 60 to reduce the bias caused by mortality. The samples were also restricted to people aged 25 or above to reduce the bias due to people re-locating

to finish their education in different schools.

In the baseline study, this paper only includes the white sample for two reasons. First, according to Lleras-Muney (2002), CSLs have a similar effect on white males and females, but have no effect on blacks. Secondly, blacks showed a drastically different migration pattern in the early 20th century, which is commonly referred to as the Great Migration. Driven by harsh segregationist laws, many blacks migrated from the rural south to the urban north during that period, which is different than the whites. However, when all races are included, the results are robust, only change slightly.

The present study focuses on migration between states, since the between-state movements is more likely to be driven by employment opportunities or wage differences (Niedomysl (2011)). Also, only lifetime migration is observable in the census by comparing people's birth states and resident states during the survey. The variable of migration status is defined as a binary variable, equal to 1 if state of birth is different from the state of residence at the census year, and equal to 0 otherwise. (For simplicity, migration only refers to residing out of birth states hereinafter.) The individual characteristic variables used in this analysis, such as age, sex, education, state of birth and state of residence are directly extracted from the IPUMS.

In this paper, migration distance is proxied by the distance between birth state population center and resident state population center. The distance is calculated using an accurate ellipsoidal model of the Earth. The longitude of state population centers are obtained from the Census Bureau. Since migrating to Hawaii and Alaska might show different patterns than other U.S. states, this study excludes those individuals for simplicity. However, including the two states will not change the results since only small amount of the population live there.

Table (2.1) and Table (2.2) list the summary statistics of the data. The sample average

years of schooling is 10.2 years. Around 31% of the total population migrated at least once. For those who migrated, the average migration distance is 694 miles, proximately the distance between Chicago and New York City.

Table 2.1: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Education (year)	634,656	10.2	3.3	0	18
Migration Distance (km)	196,795	694.4	642	29	3031.7
Migration Status	634,656	0.31	0.46	0	1
Female	634,656	0.5	0.5	0	1
Age	634,656	39.4	8.8	25	59

Note: Included cohort born between 1901 and 1925, white population.

Data Source: Integrated Public Use Microdata Survey 1940-1960, U.S. Census Bureau

Since samples exposed to different CSL levels are very unequal (Table (2.2)), I divide CSLs into 4 broad categories to be more balanced. The 4 categories are: 1) Low CSL (with 0, 4 and 5 years of CSL, constituting 5.7% of total sample); 2) CSL=6 (with 6 years of CSL, 31.13%); 3) CSL=7 (with 7 years of CSL, 46.96%); and 4) High CSL (with 8, 9 and 10 years of CSL, 16.20%). As shown in Table (2.2), a majority of the states have compulsory schooling requirement at either 6 or 7 years of education. The categorical CSL defined here will be used as an alternative IV in the estimation, to supplement the results using linear CSL.

## 2.4 Identification Strategy and Specification

The main goal of the present study is to estimate the effect of education on a person's migration distance and probability. There is a long-standing belief that education is correlated with many characteristics which also affect migration, such as motivation, health, social networks, or abilities. If a highly motivated individual is more likely to finish school, and to accept jobs

Table 2.2: Summary Statistics of Compulsory Schooling Law

Compulsory Schooling Law (yr)	Frequency	Percentage (%)
CSL=0	7,929	1.28
CSL=4	17,927	2.88
CSL=5	7,405	1.19
CSL=6	190,282	30.62
CSL=7	295,241	47.51
CSL=8	54,597	8.79
CSL=9	19,541	3.14
CSL=10	28,541	4.59

Note: Included cohort born between 1901 and 1925, white population.

Data Source: Integrated Public Use Microdata Survey 1940-1960, U.S. Census Bureau

that require relocation, one might suspect that the OLS is biased and will overestimate the effect of education. Reverse causality is another issue that can not be addressed by OLS. If families move to seek better education for their children, OLS will overestimate the impact of education. Lastly, measurement error in schooling might also introduce bias when using OLS. But according to Angrist and Krueger (1991), the bias is relatively small. The present study proposes to use exogenous education variation that is caused by Compulsory Schooling Law (CSL) reform in the U.S. from 1915 to 1939 to control for these issues.

In this paper, I use state-level CSL changes from 1915 to 1939 as an instrument to study how education affects migration behavior. The first CSL in the U.S. was passed in 1852 by the Massachusetts General court in an attempt to transform education from a moral obligation into a legal requirement. At the time, CSL requirements were mostly symbolic, lacking enforcement mechanisms. Between 1900 and 1930, many states started enforcing the law more rigorously, i.e. establishing attendance offices and institutionalizing school census to record students' education progress. At the same time, CSLs themselves were also changed from simple schooling requirements into complex legal systems which also include child labor regulations. As a result,

CSLs were more strictly followed during this period (Katz (1976)). After 1940, the importance of CSLs were reduced since average education level increased steadily, hence the number of students who would drop out school if without CSLs reduced (Edwards (1978)). In addition, the reform of CSL was mainly driven by social forces at the time of passage, and hence were mostly independent of state-wide economic conditions. These factors make CSL reforms a valid IV for this time period due to its impact on child schooling and its independence from individual characteristics.

The CSL variables are constructed following Lleras-Muney (2005). In the early 20th century, child labor laws and compulsory schooling laws often were not coordinated (Lleras-Muney (2002)). Lleras-Muney (2002) suggest that CSLs affected schooling through three major aspects of these laws: the age at which a child had to enter school (*enterage*), the age at which a child could get a work permit to leave school and work (*workage*) and whether a state required a child with a work permit to finish the schooling requirement by studying part-time. When the school entrance age and the work permit age were enforced following CSL, the difference between these two ages represents the years that a child had to attend school. Following these studies, the CSL variable is calculated by combining the age at which a child had to enter school and the age required for a work permit, defined as  $CSL = workage - enterage$ . Currently, most states require all students to start school at age 7 and permit them to work when older than 16, so the effective compulsory schooling would be  $16 - 7 = 9$  years.

Based on the variables defined above, this paper estimates the following 2SLS model. The econometric specification is as following:



$$\log(D)_i = \alpha + E_i\beta + X_i\zeta + H_s\mu + \gamma_t + \lambda_c + b_s + trend_s + u_i, \quad (2.4)$$

$$m_i = \alpha + E_i\beta + X_i\zeta + H_s\mu + \gamma_t + \lambda_c + b_s + trend_s + u_i, \quad (2.5)$$

with first stage specified as below:

$$E_i = \alpha + CSL_{cs} \cdot \theta + X_i\lambda + H_s\mu + \gamma_t + \lambda_c + b_s + trend_s + v_i, \quad (2.6)$$

In the equations above,  $\log(D)_i$  is the log migration distance measured in kilometers;  $m_i$  is the migration status dummy of an individual  $i$ : equal to 1 if  $i$ 's birth state  $s$  is different from his resident state in census year  $t$ , and equal to 0 otherwise. The completed schooling years are represented by  $E_i$ . Its coefficient  $\beta$  is the center of estimation. Variable  $CSL_{cs}$  is the compulsory schooling requirement when a given individual is 14 years old for cohort  $c$  born in state  $s$ , which serves as the excluded instrument. Vector  $X_i$  are individual characteristics which are exogenous to unobservables at the individual level. Here I include gender, age and square of age. Since the Census data from various waves are being pooled together, it is necessary to include census year dummies  $\gamma_t$ . Cohort fixed effect  $\lambda_c$  is also included to control for nationwide shocks that affect all states similarly. I also include birth state fixed effects  $b_s$  to control for state specific amenities. A state with pleasant weather or beautiful landscapes might attract more highly educated people. Failing to control for state amenities might cause a spurious correlation between education and migration.

Adding state-specific time variant variables are important to separate the effect of improved

education from other state trends. Many state-wide changes were occurring at the same time with CSL increase and schooling improvement. One may argue that the effects of increased schooling on migration could simply reflect the impact of those omitted state-level shocks. For example, a prospering economy could both make education more affordable to local governments and make the labor market more attractive and decrease out-migration simultaneously. Omitting this confounding factor would underestimate the effect of education on migration. This model includes two sets of variables  $trend_s$  and  $H_s$  for this purposes. Variables  $trend_s$  are linear birth state trends, constructed by using birth state dummies multiplied by birth year.  $H_s$  are state-level characteristics at individual  $i$ 's birth state at age 14, which are non-linear and time variant.

State-wide characteristic  $H_s$  are chosen to control state-wide economic and labor market conditions that might affect education and migration together. The variables include average annual wage in the manufacturing industry, percentage of manufacturing employment of the total population, average value of a farm per acre, number of doctors per capita, state expenditures on education, the number of school buildings per acre, percentage of foreign born white population, percentage of black population, and percentage of urban population. Each cohort is matched with their birth-state characteristics including CSLs at the year when the cohort reached age 14, as it is the lowest common drop out age across states. However, the results show that excluding those variables does not affect the main conclusion.

## 2.5 OLS results

This section reports the OLS estimation results, which serve as the benchmark for comparison with the IV results. The OLS results are consistent with previous literature, suggesting that one additional year of education is associated with a 2.1% increase of migration distance and a 1.8% increase of the probability to migrate. These results are very robust when using different subsamples.

Part 1 of Table (2.4) reports the results using log migration distance as a dependent variable. The econometric model used for OLS estimation is summarized by Equation (2.4). Using the entire white population, the results suggest one additional year of schooling will increase travelling distance by 2.1%. The mean migration distance in this period is 694 miles. A 2.1% increase indicates an extra 14 miles of moving distance, which is a quite small effect.

Part 2 of Table (2.4) analyzes how education affects the probability to migrate following Equation (2.5). The estimated coefficient of education is positive, statistically significant and robust to different sample choices and specifications. When use the entire white sample, the marginal effect of education on migration is 1.8%, implying that one extra year of education will increase the likelihood of migration by 1.8%. Since the mean migration rate for the entire sample is only 32.5%, an 1.8% increase is not trivial. This effect is consistent with many other studies on both size and direction (Todaro (1980); Greenwood (1997)).

The OLS specification is robust to different specifications and subsamples. Table (2.4) includes estimation results using three main subsamples. First of all, the Dust Bowl was a major weather disaster that drove a large amount of the farm population to move in the 1930s. Column (2) reports the results excluding the main Dust Bowl-affected states (Oklahoma, Kansas,

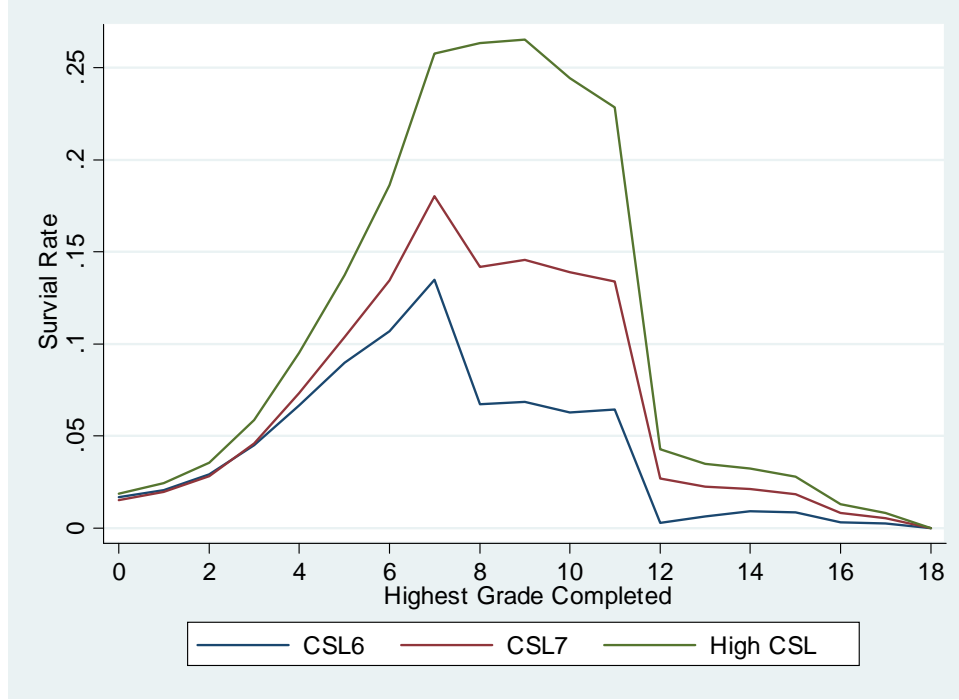
Colorado and New Mexico). Secondly, since CSLs mainly affect people with less education, column (3) reports the results using only people with less than 12 years of schooling. Thirdly, even though the First and Second World War did not happen on the American homeland, they may have influenced internal migration through the draft. Column (4) shows the results with only male non-veterans. All above subsamples return very similar results. Table (2.5) reports estimation using different specifications. The estimator barely changed at all when using different trends or state characteristics.

## 2.6 The Effect of Compulsory Schooling Laws on Education

### 2.6.1 Visual Evidence: The Survival Function of Schooling

Figure (2-1) not only provides visual evidence of the first stage regression results, but also provides some insight into how CSLs affect educational distribution. Following Acemoglu and Angrist (2001), each line displays the survival function of education under a particular CSL minus the baseline distribution. The survival function of education is defined as  $1 - \text{Cumulative Distribution Function}$  ( $SF = 1 - CDF$ ). The baseline distribution is the survival function when there was Low CSL, denoted as  $SF_{CSL=0,4,5}$ . Intuitively, each line represents the difference between the rate of surviving certain grades when subject to a particular CSL and Low CSL. For example,  $CSL=6$  is defined as  $SF_{CSL=6} - SF_{CSL=0,4,5}$ . On line  $CSL=6$ , point (7, 13.2%) means compared with people subject to Low CSL, people subject to 6 years of CSL are 13.2% more likely to finish at least 7 years of education. Intuitively, the lines display the difference between groups affected by different CSLs. Each point on the line represents the probability of completing grades higher than grades on the x-axis compared with baseline distribution.

Figure 2-1: Survival Function of Education by Severity of Compulsory Schooling Law



Note: Each line represents the survival function of education distribution under different Compulsory schooling law (CSL) compared with baseline distribution. The points on each line represent the probability of complete grades higher than grades on x-axis compared with when expose to low CSL requirement. CSLs are calculated at birth state-level when individual is 14. Survival function of education is equal to 1-Cumulative Distribution Function ( $SF = 1 - CDF$ ). Baseline distribution is survival function when there was low CSL, defined as  $SF_{CSL=0,4,5}$ . The line CSL6 represents  $SF_{CSL=6} - SF_{CSL=0,4,5}$ . The line High CSL is  $SF_{CSL>7} - SF_{CSL=0,4,5}$ .

Data Source: Integrated Public Use Microdata Survey 1940-1960, U.S. Census Bureau

There are two features of the graph that support the choice of using CSL change as IV.

First of all, the survival rate of any grade increases with increased CSL. All the lines are above 0, meaning those who are exposed to CSL are more likely to survive any level of schooling than those who are exposed to less than 6 years of CSL. Also, the lines are monotonically shifting up as CSLs increase. For example, CSL=7 is significantly above the CSL=6 line, especially between 4-12 grades, even though they have similar shapes. This implies that people living

in areas with 7 years of CSL requirement were much more likely to complete grade 4-12 than those who live in areas with 6 years of CSL. Overall, figure (2-1) demonstrates that the higher the CSL, the higher the probability of completing certain grades. This observation is consistent with the first stage result in table (2.3) when using broad CSL categories.

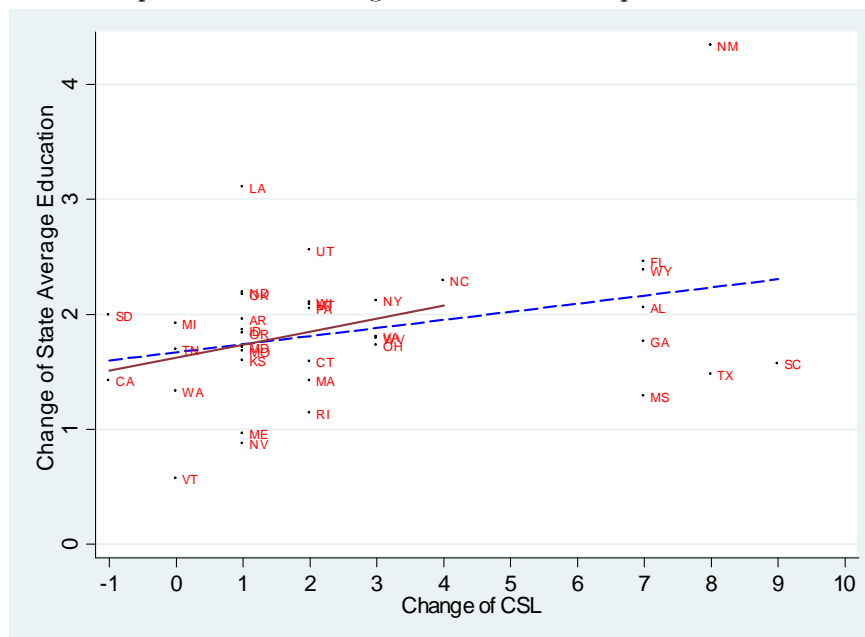
The second feature is that the survival rate difference increases as schooling increases, then drops drastically after grade 12. The trend implies that the main impact of CSL was on people's middle- and high-school education. Also, the figure suggests the problem of omitted variables such as macroeconomics conditions or preferences for schooling across states might not be correlated with CSL. The reason for this is that if they are highly correlated, since those omitted variables are also correlated with the likelihood of college education, then a stricter CSL should be associated with a higher chance of finishing collage. That is not the case here since the impact of CSLs on schooling dropped sharply after grade 12.

### **2.6.2 The Impact of Compulsory Schooling Laws on State Education**

To show the association between state level education and the change of CSLs, Figure (2-2) plots changes in state average education between cohorts born in 1901 and 1925 against the changes in CSLs between 1915 and 1939. The dash line is the fitted line of the entire sample, with a slope of 0.07. The positive slope implies a positive effect of CSLs on education. The solid line is the fitted line that excludes states with CSLs that increased more than 7 years, with a slope of 0.113. Those states are mainly southern states with less educated people. The solid line also slopes upward but is steeper than the dash line. The slope difference indicates that the excluded southern states flatten the slope, with a large increase of CSLs but only a moderate increase of education. If we treat states with more than 7 years of CSL increment

as outliers and drop them, we will observe an even larger effect of CSLs on education, which indicates that the positive effect of CSLs is not driven by some particular states.

Figure 2-2: The Impact of CSL Change on Education Improvement: Cohort 1901-1925



Notes: The change of state average education is calculated using cohorts born in 1901 and 1925. The changes of CSLs are between 1915 and 1939. Sample contain only white population and states that experienced CSL changes. The dash line is the fitted line of the entire sample. The solid line is the fitted line of the subsample that excludes states with CSLs that increased more than 7 years.

Data Source: Integrated Public Use Microdata Survey 1940-1960, U.S. Census Bureau

### 2.6.3 First Stage: The Effect of Compulsory Schooling Laws on Education

Consistent with visual evidence in Figure (2-1) and Figure (2-2), the OLS estimator suggests that CSLs do have a statistically significant influence on educational attainment. Table (2.3) summarizes the first stage results following Equation (2.6). The coefficient of interest is the effect of CSLs, which is positive, statistically significant, robust to different sample choices and relatively small. Using the entire white sample, the results indicate that one extra year

of compulsory schooling requirement when a person was 14 would increase that person's total schooling by 0.027 years. Lleras-Muney (2005) estimated the impact of CSLs on education to be 0.046, which is also positive but with twice the magnitude. The results differ because the present study includes more control variables, i.e. age and race dummies, and uses birth state trends instead of regional trends. Since each region contains several states, using regional trends might not be able to separate state-specific developments from the CSL increments, and hence would overestimate the effect. Using state trends and adding the additional controls significantly reduces the magnitude but reflects a cleaner effect of CSLs. In addition to OLS results discussed above, the first stage F-statistics are all statistically significant, rejecting the weak IV assumption. This result could alleviate concerns that states passed CSLs when they were no longer binding, or states did not enforce them strictly. If that were the case, the weak correlation between CSLs and completed schooling years would lead to an inconsistent estimation in the second stage.

## **2.7 The Impact of Education on Migration**

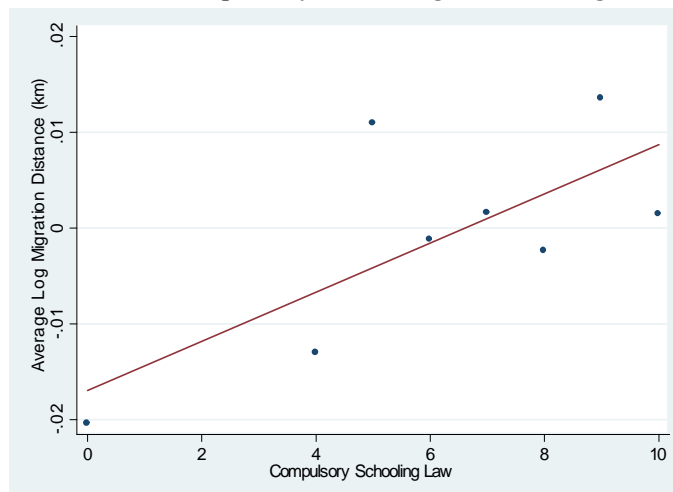
### **2.7.1 The Impact of Education on Migration Distance**

This section presents the empirical results, suggesting that education increases people's migration distance. The reduced form results indicate a direct positive correlation between CSLs and distance. The 2SLS estimator suggests the marginal effect of education is to increase distance by 23%. The positive impact of education could be consistently observed using various specifications and subsamples.



## Reduced Form Estimation

Figure 2-3: Effect of Compulsory Schooling Law on Migration Distance



Notes: Figure is drawn using the white sample born between 1901 and 1925 in the U.S., who lived outside of their birth state. Dots are average log migration distance for each CSL with the effect of cohort, birth state and birth state trend taken out. The solid line is the fitted line of the entire sample.

Data Source: Integrated Public Use Microdata Survey 1940-1960, U.S. Census Bureau

The reduced form estimators in Table (2.3) suggest that one extra year of compulsory schooling requirement is associated with 3% greater migration distance when use the white migrants born between 1901 and 1925 in the U.S. as the sample. The fitted line in Figure (2-3) is the visual display of the association, which slopes upward after controlling for cohort, birth state and birth state trend effects. When using categorical CSL as IV, the omitted category is CSL that require less than 6 years of education. Compared with this, CSL that require 7 years of education or above increase people's migration distance. The only exception are CSLs that require 6 years of education, which reduce migration distance by 9.5%. The negative effect is mostly driven by the migrants subject to 5 years of CSL. Since those migrants only constitute 1.19% of the total sample, their greater average distance might be an outlier (Table (2.2)).

## 2SLS Results

Both OLS and 2SLS results in Table (2.4) show that education increases migration distance. However, 2SLS results indicate an impact that is 10 times larger (0.233 v.s. 0.021), suggesting that an additional year of schooling increases migration distance by 23.3% for individuals who migrated. This effect is quite large. Since the average migration distance is 694 miles, a 23.3% increment means moving an additional 162 miles, which is roughly the distance between New York City and Boston.

This paper also uses a different specification of CSLs to test the robustness of the results. As discussed before, CSLs are divided into four broad categories: Low CSL (CSL=0, 4, 5), CSL=6, CSL=7, and High CSL (CSL=8, 9, 10). When using these categorical dummy variables as IV, the results are very similar to the baseline results (column (6)). The magnitude of the effect decreases slightly, from 23.3% to 21.8%, and the standard deviation also decreases which leads to a tighter estimation.

Additionally, Table (2.4) reports three robustness tests results using different subsamples. Excluding the main Dust Bowl-affected states decreases the estimated effect slightly, from 23.3% to 17%, indicating that education has a larger impact on those states. Other robustness tests, such as excluding the sample with more than 12 years of education or excluding females and veterans, consistently suggest a positive effect of education on migration distance. Though not reported in the main table, excluding California residents also suggests a positive impact. Based on total inflow of migrants, California is the most popular destination state in the U.S. before 1940. It attracted more than twice the number of migrants as the second popular state New York. Approximately 3.6 million migrants reside in California in 1940, which constitutes 15.4%

of total migrants in the U.S.. Due to its specialty, a robustness test is done by excluding all the California residents, and the estimated impact is positive but not statistically significant.

Based on the conceptual framework, the positive impact of education on distance indicates an increasing return to distance when people become more educated. One possible reason is that jobs for highly-educated workers are more disperse than jobs for less-educated workers, hence searching a longer distance is more beneficial for the highly-educated workers. If the marginal cost of distance is similar for both types of workers, then the highly-educated workers would migrate longer distances on average.

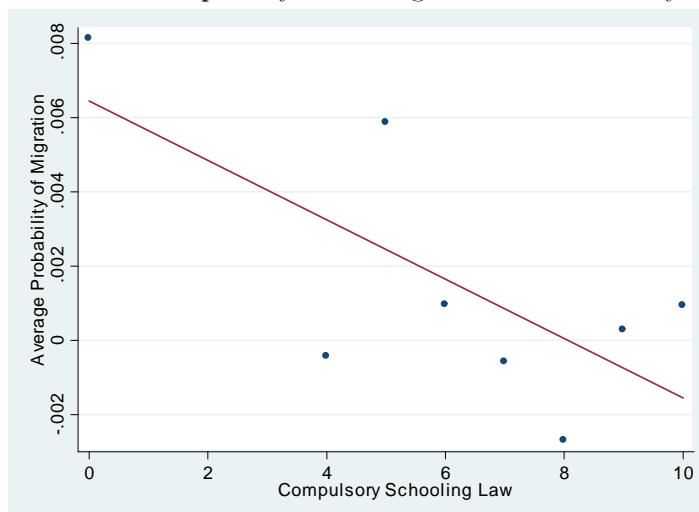
### **2.7.2 The Impact of Education on Probability of Migration**

This section shows that education decreases people's probability of migration and makes them more stable when using CSL as IV. 2SLS results suggest that one additional year of schooling decreases the migration probability by 6.6%. The result is robust under different specifications and using different subsamples.

#### **Reduced Form Estimation**

Table (2.3) displays the reduced form estimation and suggests that one extra year of CSL requirement decreases the probability of migration by 0.2% (column (5)). The fitted line in Figure (2-4) visually displays the negative correlation, which slopes downward after controlling for cohort, birth states, and birth state trends. The results using CSL categories as explanatory variables suggest that CSLs are negatively correlated with migration probability only when states require 7 years of schooling or above. Figure (2-4) suggests that this is mostly because individuals subject to 4 years of CSL have low probability of migration.

Figure 2-4: Effect of Compulsory Schooling Law on Probability of Migration



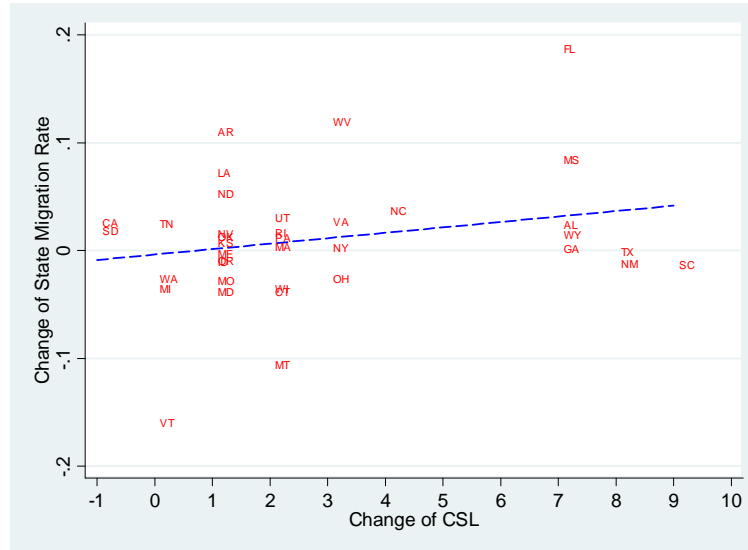
Notes: Figure is drawn using the entire whites sample born between 1901 and 1925 in the U.S.. Dots are average probability of migration for each CSL with cohort, birth state and birth state trend effects taken out. Solid line is the fitted line.

Data Source: Integrated Public Use Microdata Survey 1940-1960, U.S. Census Bureau

## 2SLS Results

When comparing OLS and 2SLS results in Table (2.4), the coefficients not only have different magnitudes but also have opposite signs. The baseline includes the entire white sample. The OLS estimator shows 1 more year of education is associated with a 1.8% higher probability of migration (column (5)). In contrast, the 2SLS estimator indicates that one additional year of education decreases the probability of migration by 6.6%. When using categorical CSL as IV, results are very similar, showing that the marginal effect of education is -7.4% (column (6)). If we focus on the population with less than 12 years of education, the difference between OLS and 2SLS is slightly larger. The OLS estimator suggests the correlation is 1.4%, while the 2SLS estimator suggests the effect is -7.2%. According to the conceptual framework, the estimated negative impact suggests there might be a substantial marginal cost of migration distance,

Figure 2-5: Effect of Compulsory Schooling Law on Probability of Migration:By States



Notes: Figure is drawn using the entire whites sample born between 1901 and 1925 in the U.S.. Dots are average probability of migration against the change of compulsory schooling laws for each states. Solid line is the fitted line.

Data Source: Integrated Public Use Microdata Survey 1940-1960, U.S. Census Bureau

especially considering the poor transportation and communication system in the early 20th century.

To see whether the results are driven by particular events during 1915-1960, two robustness tests are done. First, weather disasters like Dust Bowls have large impacts on poor and low educated people, hence they might cause a negative correlation between education and migration. After excluding main Dust Bowl affected states, the estimator is still negative, only changed slightly, from -6.6% to -6.3% (column (7)). This indicates that the disaster induced migration could not explain why lower educated people are more likely to move. Secondly, veterans are usually more mobile than non-veterans. In the data, they are also more educated on average. Restricting the sample to only men who are non-veterans, the OLS result is almost identical. However, since the sample size is greatly reduced by more than 60%, the 2SLS result is no

longer significant (column (9)). However, the point estimator is still negative and very close to the baseline results.

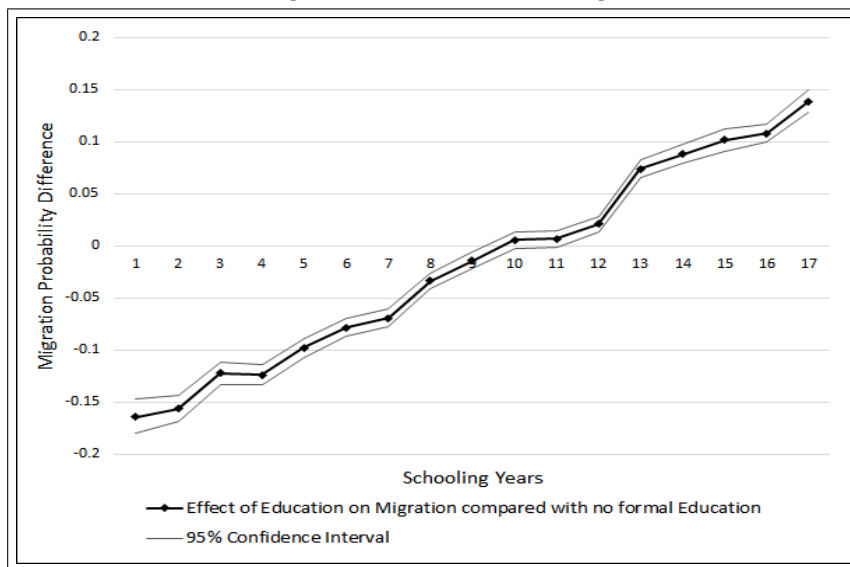
### **Is the Impact of Education U-shaped?**

Since Figure (2-1) implies that CSLs mainly affect people with 12 years of education or under, it is natural to ask whether the difference between OLS and IV estimation simply reflect the difference between average treatment effect and local treatment effect. Imagine when the impact of education is "U-shaped," decreasing at first then increasing when education increases. If this is the case, it is possible that OLS estimation with a large portion of a highly educated sample shows a positive effect of education on migration, since the average treatment effect could be positive due to the upward sloping area. And when using CSL as IV, 2SLS results, which estimate the local treatment effect, are negative due to the downward sloping area at the low end of education distribution.

Figure (2-6) suggests that the difference between average treatment effect and local treatment effect could not explain the observed difference between OLS and IV estimation. Figure (2-6) is drawn with the coefficients obtained from OLS estimation of Equation (2.5) using categorical years of education as explanatory variables instead of total years of education. The omitted variable is the dummy variable representing no formal schooling, which serve as baseline effect. The coefficients of other categorical schooling years imply how achieving certain years of education changes the probability of migration compared with no education at all. The estimated coefficients are plotted in Figure (2-6), which shows clearly that people with higher education level are always more likely to migrate when using OLS. The only exception is when people increase their schooling from 3 to 4 years, their migration likelihood decreases by 0.002,

however, this difference is statistically insignificant. In conclusion, the difference between OLS and 2SLS is consistent using different subsamples, and is not caused by local treatment effect.

Figure 2-6: Effect of Categorical Education on Migration: OLS Estimation



Note: This figure displays the coefficients of categorical education attainment from OLS estimation of Equation (2.5) using categorical schooling instead of total years of education as  $E_i$ .

Data Source: IPUMS 1940-1960, U.S. Census Bureau

### 2.7.3 Alternative Specification

Recently, several papers have pointed out that including a location specific time trend will drastically alter the results when using state schooling requirement change as IV. Because when using CSL as an IV, the identification is from the variation of policy changes within the state which affect different cohorts. As state policy, compulsory schooling reforms might coincide with other locational trends such as the change of economic or labor market conditions, and other social reforms. Hence, it is important to include location specific time trends to separate the effect of increased schooling requirements from birth place trends. Mazumder (2008) suggests

that ignoring state trends tends to overestimate the impact of education on health outcomes, and hence, causal interpretation of the results is unwarranted. Stephens and Yang (2014) also suggest that including regional time trend will eliminate some observed effects or alter the sign of the effect of education on personal outcomes. The reason why adding time trends will change the results drastically remains unclear. Their suggestion is to split the sample and test the robustness of the results. Although not reported in the table, the main results hold when all the southern states are excluded as a robustness test.

The results in this paper suggest that using detailed time trends is important when using CSL as an IV, but has limited influence on the OLS results. Table (2.5) reports the estimation results when different birth place trends specifications are used. The estimated effects of education on migration outcomes are almost identical under different specifications when using OLS estimation. On the other hand, the point estimators using 2SLS suggest that location trends significantly affect the estimated effect of education. The different effects of locational trends when using OLS and 2SLS might indicate that the locational trend is particularly important to identify the effect of CSLs in the first stage. Regional trends show no particular power in improving the identification of the impact of education, suggesting that it is not able to fully separate the influence of state development.

While it is not reported in the main table, analysis has also been done using samples that are younger than 18. The OLS estimators are statistically significant at 1% level, suggesting that 1 year of education is associated with a 0.9% higher probability to migrate and a 4.2% longer migration distance if they decided to move. The 2SLS estimators are much less precise due to the restricted sample size. None of the estimator is statistically significant at 10% level, but the sign of the point estimator is consistent as before, suggesting that education causes



people to migrate further away from their birth place, but reduces their likelihood to move<sup>5</sup>. The results suggest that we could not rule out the possibility that families migrate for better education since a higher CSL reduces the migration likelihood for those who were under 18.

## 2.8 Using Quarter of Birth as Instrument Variables

This section presents results when using quarter of birth as an alternative IV to estimate the effect of education on migration. This identification strategy is well-founded and easy to understand. Angrist and Krueger (1991) were the first to point out that seasons of birth are related to education attainment due to the school starting age policy and compulsory schooling laws. Most school districts require a child to have turned 6 years of age by January 1st of the year in which she/he enters school. Hence, individuals who are born at the beginning of the year start school slightly older than those born at the end of the year. Most states also require students to remain in school until they turn 16 or 17. Those who are born in the early months of the year could legally leave school with less education since they are older compared with other students in the same grade. After establishing the impact of birth date on schooling, Angrist and Krueger (1991) then utilized the education variation caused by birth date as an IV to identify the effect of schooling on future earnings.

Since then, many studies have been done to discuss the validity of the method. One main critique of the original method is that one's quarter of birth is only weakly correlated with their education outcomes (Bound et al. (1995)). It is well known that when the instruments are weakly correlated with the endogenous variable, 2SLS can be severely biased (Rothen-

---

<sup>5</sup>The point estimator and variance of the effect of education is 77.3% and 0.883 when using migration distance as dependent variable; and -4.8% and 0.135 when using migration probability as dependent variable.

berg (1984)). Staiger and Stock (1997) suggests that limited information maximum likelihood (LIML) estimation is less biased than 2SLS in the existence of weak instrument, particularly when there is only one endogenous variable. Anderson and Rubin (1949) proposed a test of structural parameters (the AR test) that has the correct size under a wide variety of violations of the standard assumptions of IV regression, including under weak instrument conditions. In this section, I present estimation results from OLS, 2SLS and LIML results, using AR test as a robustness test when using LIML estimation.

The main specification, which directly follows Angrist and Krueger (1991), is presented in the following equations:

$$E_i = X_i\pi + \sum_c Y_{ic}\delta_c + \sum_c \sum_j Y_{ic}Q_{ij}\theta_{jc} + u_i$$

$$mig\_outcomes_i = X_i\zeta + \sum_c Y_{ic}\xi_c + E_i\beta + \mu_i$$

In the above equations,  $E_i$  is education of individual  $i$  measured in years.  $Q_{ij}$  are dummy variables indicating whether the individual was born in quarter  $j$  ( $j = 1, 2, 3$ ). Fixed effect for the first quarter of birth  $Q_{i0}$  is used as the baseline and is thus omitted in the regression.  $Y_{ic}$  are dummy variables indicating whether the individual was born in year  $c$  ( $c = 1, \dots, 10$ ).  $X_i$  are vector of covariates including age (detailed to quarters), square of age and gender, cohort fixed effect, state fixed effect and state trends.  $\theta_{jc}$  is the fixed effect of being born in year  $c$  and

quarter  $j$  compared with being born in year  $c$  but in the first quarter. Instead of assuming a similar effect of quarter of birth for all cohorts, this specification allows the effect to differ

among the cohorts, and hence more flexible.

Using quarter of birth as an IV, the results in Table (2.6) are different compared with using CSL to estimate the effect of education on people’s probability of migration. There could be several reasons behind this. First of all, weak IV might bias the results when using quarter of birth even though LIML is used and Anderson Rubin test is done. Secondly, as a state policy, CSL not only affects education at the individual level, but also affect education at the state level. Increased average education at the state level might also increase state amenities, hence encouraging people to stay. On the contrary, quarter of birth will not have that effect.

## 2.9 Discussion

The main conclusion of this paper is that education does have a significant impact on migration outcomes. Compared with other studies, the OLS results are quite similar, which indicate that one additional year of education is associated with a 1.8% higher probability to migrate and 2.1% greater distance of that migration (Greenwood (1997)). Using state CSL reforms as an IV, this paper suggests the causal impact of education on migration distance is 23.3%, which is in the same direction but much larger than the OLS results. However, using the same IV, this paper also reports a negative point estimator of the impact of education on migration probability, which suggests that the marginal effect of schooling is to reduce people’s probability to migrate by 6.6%. These results are similar to McHenry’s (2013), in which negative impact of education on both lifetime and 5 year migration probability is reported using a similar identification. However, other studies suggest different effects using various identifications. For example, Malamud and Wozniak (2012) reports that an additional year of postsecondary

schooling increases probability of migration by around 9%, using variation in college attainment due to draft-avoidance behavior during the Vietnam War. One main difference between this study and theirs is that the reforms used here affect primarily the lower part of the educational distribution, while the draft-avoidance in their paper mainly affects college education. This could be one of the reasons why our results differ.

The drastic difference between OLS and 2SLS result shows that it is important to treat endogenous unobservables when study the effect of education, since omitted variables such as ability could lead to largely biased results. For example, assume that compared with low educated worker, the social network of high educated worker makes it easier for them to find new jobs, but also result in a higher loss if they move away from their network. If this is the case, then the unobserved social network will increase the chance of migration but reduce their incentive to move far away. Without controlling it, OLS will overestimate the effect of education on migration but underestimate the effect of education on distance.

Several policy implications may be derived from the results. First of all, the effect of a local educational program on migration is theoretically indeterminate, and the results question the conventional idea that more education would necessarily lead to more out-migration of highly-educated workers, which is loosely referred to as "brain drain" (Beine et al. (2001)). This paper suggests that under certain conditions, education could make people more stable. At the state level, to determine the optimal investment on education, the government should consider how it would affect local economy. Knowing the migration effect of education would allow us to accurately estimate the return of education program. Secondly, the ability to migrate to a suitable location with better job offers and neighborhood amenities is essential for people's well-being. When evaluating the overall effect of education at the individual level, we should

also consider its influence on geographic mobility. The results suggest that more education leads to a broader migration distance hence more location choices.

The policy implications should be viewed with caution. First of all, as pointed out by Imbens and Angrist (1994), the effect estimated using IV are a local average treatment effect, which represents the effect on the binding population only. Those who are pushed into more education by compulsory schooling laws might be different than the whole population. Second, the cohorts studied in this paper are historical cohorts in the early 20th century, which resemble the developing world more than the modern developed world. To draw inference for the developed world, more recent instrument variables are needed.

To improve our understanding of the decision making process of migration, more research need to be done. This paper mainly focuses at the individual level. The model suggests that the negative impact of education on migration probability indicates a high marginal cost of migration distance. However, this paper could not rule out other potential mechanisms. For example, state CSL reforms not only affect individual schooling, but also affect the average education level in a state. An improved average education level might make a state more attractive to highly educated workers, hence making them more stable. This group effect could be regarded as one of the external effects of education, which are essentially inseparable when using state policy as IV. Since CSL reform binds a relatively small portion of the total population, this paper chose to focus on individual effect and overlook the general equilibrium effect.

Table 2.3: The Effect of CSLs:First Stage and Reduced Form Results

Dependent Variable:	First Stage		Reduced Form			
	Schooling Years		Log Migration Distance		Migration Probability	
	(1)	(2)	(3)	(4)	(5)	(6)
	Linear	Categ.	Linear	Categ.	Linear	Categ.
CSL requirement	0.027*** (0.006)		0.030*** (0.002)		-0.002** (0.001)	
CSL=6		0.088*** (0.030)		-0.095*** (0.010)		0.001 (0.004)
CSL=7		0.128*** (0.032)		0.055*** (0.010)		-0.008* (0.005)
High CSL		0.187*** (0.037)		0.099*** (0.011)		-0.010* (0.005)
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Individual Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
State-level Controls	Yes	Yes	No	No	No	No
Birthyear Dummy	Yes	Yes	No	No	No	No
Birth State Dummy	Yes	Yes	No	No	No	No
Census Year Dummy	Yes	Yes	No	No	No	No
State Dummy*Cohort	Yes	Yes	No	No	No	No
F Statistics for Weak IV	34.976	90.388				
N	634656	634656	196795	196795	634656	634656
Standard errors in parentheses						
* p<0.10; ** p<0.05; *** p<0.01						

Note: The first and second columns present the first stage estimation following Equation (2.6) using CSL and CSL categories. The remaining columns present the reduced form estimation with migration distance and probability as dependent variables. In those columns, the control variables are gender, age and age squared. All estimations use the white sample born between 1901 and 1925.

Data Source: Integrated Public Use Microdata Survey, 1940-1960, U.S. Census Bureau

Table 2.4: The effect of Education on Migration: OLS and 2SLS estimation results

Migration Distance									
OLS					2SLS				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
White		No DB	Low Edu	Male non Vet	White	CSL Categ.	No DB	Low Edu	Male non Vet
Education (yr)	0.021*** (0.001)	0.023*** (0.001)	0.029*** (0.001)	0.023*** (0.001)	0.233* (0.127)	0.218** (0.101)	0.170* (0.101)	0.167 (0.114)	0.223 (0.190)
F Statistics					34.976	12.689	37.203	27.049	7.02
N	196795	186615	150714	60493	196795	196795	186615	150714	60493

Migration Probability									
OLS					2SLS				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
White		No DB	Low Edu	Male non Vet	White	CSL Categ.	No DB	Low Edu	Male non Vet
Education (yr)	0.018*** (0.000)	0.018*** (0.000)	0.014*** (0.000)	0.018*** (0.000)	-0.066* (0.035)	-0.074** (0.032)	-0.063* (0.033)	-0.072* (0.038)	-0.089 (0.057)
F Statistics					90.388	36.177	95.766	82.777	24.791
N	634656	616880	522796	205540	634656	634656	616880	522796	205540

Included Control Variables:

Constant, Individual Characteristics, State-level Controls, Birth State Fixed Effect, Birth Year Fixed Effect, Birth State Dummy\* Cohort

Standard errors in parentheses  
\* p<0.10; \*\* p<0.05; \*\*\* p<0.01

Note: This table present the OLS and 2SLS estimation results following Equation (2.4) and (2.5), with first stage estimation following Equation (2.6). All specifications include birth year dummy, birth state dummy and census year dummy. First part of the table use log migration distance as dependent variable. The second part of the table use migration dummy as dependent variable. Though not reported in detail, all of the estimations include personal characteristics, state-level control variables listed in previous sections, birth year dummies, birth state dummies and birth state trend. The White column report the results using the entire white population within Census 1940-1960 born between 1901 to 1925. Different subsamples and different IV specification have been used in following columns to test the robustness: the NoDB column exclude individuals born in Oklahoma; the Low Edu column exclude individuals with more than 12 years education; the Male non Vet column exclude female and veterans; the CSL Categ. column use broad CSL categories defined in section 7 as IV instead of linear CSL variable.

Data Source: Integrated Public Use Microdata Survey 1940-1960, U.S. Census Bureau.

Table 2.5: Results using Alternative Specification of State Trend

Specification:	State Trend	Regional Trend	No Trend	No State Characteristic	Cohort*Region Fixed effect
Migration Distance					
OLS					
Education (yr)	0.021*** (0.001)	0.021*** (0.001)	0.021*** (0.001)	0.022*** (0.001)	0.021*** (0.001)
2SLS					
Education (yr)	0.233* (0.127)	0.093 (0.060)	0.092 (0.060)	0.038 (0.053)	-0.024 (0.720)
N	196795	196795	196795	196795	196795
Migration Probability					
OLS					
Education (yr)	0.018*** (0.0002)	0.018*** (0.0002)	0.018*** (0.0002)	0.018*** (0.0002)	0.018*** (0.0002)
2SLS					
Education (yr)	-0.066* (0.035)	-0.028 (0.020)	-0.045** (0.022)	-0.163*** (0.050)	-0.003 (0.018)
N	634656	634656	634656	634656	634656
Included Control Variables:					
Cohort					
Fixed Effect	Yes	Yes	Yes	Yes	No
State					
Characteristics	Yes	Yes	Yes	No	Yes

Note: This table presents the OLS and 2SLS estimation results following Equation (2.4) and Equation (2.5), with first stage estimation following Equation (2.6). Though not reported in detail, all of the estimations include constant, age, age square, gender, birth state dummies and census wave dummies. This table use sample contains the white population born between 1901 and 1925 in the U.S.. When using log migration distance as dependent variable, only individuals who have migrated are included. Column (StateTrend) includes linear time trend of each state, which is used in previous sections. Column (Regional Trend) uses time trends of 4 census region. Column (No Trend) doesn't contain any location variable time trend. Column (No State Characteristics) uses state linear trend but doesn't contain state-wide variables such as average annual wage in the manufacturing industry, percent of manufacture employment of total population, etc. Column (Cohort\*Region Fixed Effect) drops cohort fixed effect, but add region specific cohort fixed effect, allows cohort fixed effect to differ at different regions.

Data Source: Integrated Public Use Microdata Survey, 1940-1960, U.S. Census Bureau



Table 2.6: Impact of Education on Migration: Using Quarter of Birth as IV

	OLS			Migration Distance			2SLS		
	Age 50-60	Age 40-50	Age 30-40	Age 50-60	Age 40-50	Age 30-40	Age 50-60	Age 40-50	Age 30-40
	0.018*** (0.000)	0.013*** (0.001)	0.012*** (0.000)	0.346 (0.280)	0.026 (0.055)	0.076* (0.044)	0.070 (0.057)	0.025 (0.046)	0.049 (0.033)
Years of Education									
F stats for Weak IV	364247	342228	458485	364247	342228	458485	364247	342228	458485
N									
	OLS			Probability of Migration			2SLS		
	Age 50-60	Age 40-50	Age 30-40	Age 50-60	Age 40-50	Age 30-40	Age 50-60	Age 40-50	Age 30-40
	0.023*** (0.000)	0.026*** (0.000)	0.029*** (0.000)	-0.022 (0.023)	0.053** (0.026)	0.045*** (0.017)	-0.004 (0.018)	0.023 (0.021)	0.040*** (0.014)
Years of Education									
F stats for Weak IV	953853	899088	1207427	953853	899088	1207427	953853	899088	1207427
N									
Included Control Variables:									
Indiv. Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
*Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Sample include white male who are born in the U.S.. In LIML, the Data Source: Integrated Public Use Microdata Survey 1980, 5% sample, U.S. Census Bureau

## Chapter 3

# The Heterogeneous Effect of Immigration on Housing Price

### 3.1 Introduction

Housing price is crucial to people's well-being. It not only affects their living conditions, but also affects the homeowners' wealth. Only recently, did researchers begin to analyze how immigration affects the price of this particular commodity (Saiz (2003); Saiz and Wachter (2011)). However, those studies are mostly focused on the average effect, ignoring existing neighborhood heterogeneity. In this paper, we propose to use quantile regression to analyze how immigration affects different types of neighborhoods in the housing market. In other words, we not only ask the question, "Does immigration matter?", but we also ask the question, "For whom does immigration matter?" (Eide and Showalter (1998)).

Previous literature analyzing immigration impacts has focused mainly on the displacement effects in the labor market. Kerr and Kerr (2011) provide a comprehensive survey of recent

empirical studies on the economic impact of immigration. Using U.S. data, many studies (Card (1990); Altonji and Card (1991); Longhi et al (2006); Peri (2007)) find that both the wage and employment displacement effects are very small when using various identification strategies. More recent studies also attempt to estimate how immigrants affect the prices of various commodities. Saiz (2003, 2007) are the first to show that US housing prices rise with immigration at the city level. On the other hand, Saiz and Wachter (2011) pointed out that U.S. housing prices decrease with immigration at the census tract level. These different results show that there is large spatial heterogeneities within a city, and simply estimating the average effect will overlook the most interesting within-city dynamics. To analyze the neighborhood heterogeneous effect of immigration directly, we use quantile regression.

The obvious advantages of using quantile regression is that it can estimate the effect of immigration on the whole conditional distribution of housing price, and it is less affected by outliers. Using ordinary least square (OLS) and instrumental variable (IV) estimation, the results suggest that immigrants have a very limited effect on the local housing prices on average. However, both quantile regression (QR) and the two stage quantile regression (2SQR) suggest that at both higher and lower tails of the housing price distribution, immigrants have significant and heterogeneous effects on the housing price. For example, when using the pull effect of existing immigrants as an IV, the 2SQR suggests that a 1% increase of foreign-born in the neighborhood will increase the housing price by 1% at the 80th percentile while decreasing the housing price by 1.3% at the 20th percentile. This paper also looks at potential channels through which immigrants affect housing prices in different neighborhoods. The results suggest that neighborhoods react differently when immigrants move in at both the demand and supply side of the housing market.

The welfare implications of our results are complicated. A decrease in housing price implies that the housing is more affordable for renters, but it also means homeowners are losing their wealth. As pointed out in the previous literature, people who are more likely to be credit-constrained and older homeowners, who are likely to be “trading down” on their housing stock, are the ones whose daily consumption is mostly affected by housing price changes (Calomiris et al (2012)). Hence, the adverse effect that immigrants have on poor neighborhoods could potentially bring about large negative welfare implications.

## **3.2 Data**

The empirical analysis in this paper is based on the decennial Neighborhood Change Database from 1970 to 2000, which contains housing price and demographic information at the census tract level. Census tract is a small geographic unit, with a population of less than 4000 on average. In comparison, the more commonly used Metropolitan Statistical Area (MSA) has more than a 0.9 million population on average. Using MSA as the unit of analysis could uncover the average effect of immigration at the city level, but would overlook interesting changes happening within a city. Using census tract instead will allow us to look inside each MSA and understand the effects of immigration on different neighborhoods within a city.

The Neighborhood Change Database allows us to access the decennial U.S. Census data with the geographic boundaries normalized to the 2010 boundaries. With the census tract boundaries normalized, we are able to study how neighborhoods evolve over time within the same boundary definition, and hence makes the historic comparison more accurate. Note that some areas, especially some rural areas, were not tracted in 1970 and 1980. Hence no data is

available for those years. Specifically, around 40% of the 2010 census tracts are missing for 1970 and 24% are missing in 1980. Those census tracts, along with the ones with zero population or zero housing prices are dropped from the analysis. We did not include data from year 2010 census to avoid the subprime mortgage crisis which started in 2007; even though it is available.

All MSAs are included in this analysis, unlike Saiz and Wachter (2011), who only included the MSAs with substantial immigration population. The MSAs that have low immigrant population could still have neighborhoods that are heavily affected by immigrants. Hence, they should be included in our study. Summary statistics in Table (3.1) show that the average inflation adjusted housing price has increased almost seven times in three decades, from 22.2 thousands in 1970 to 155 thousands in 2000. At the same time, the average foreign-born population more than tripled, indicating a positive time trend for both housing price and immigration. Among all the foreign born populations, the proportion with Mexican origin increased more than three times.

### **3.3 Quantile Regression and Heterogeneity**

#### **3.3.1 OLS and IV**

Studies analyzing the immigration effects have primarily relied on estimation approaches such as Ordinary Least Squares (OLS) or Instrumental Variables (IV), which estimate the average effects of immigration on neighborhood housing prices. While less robust, understanding how immigration affects neighborhoods on average could still provide a useful insights. Hence, we start our analysis with conventional OLS and IV estimation as a benchmark. The OLS estimation is specified as below:

$$\Delta \log(price)_{it} = \alpha + \beta_1 \Delta imm\_share_{it-1} + Tract'_{it} \beta_2 + Year_t + MSA_i + \epsilon_{it}. \quad (3.1)$$

where  $\Delta \log(price)_{it}$  is the change of log housing price for census tract  $i$  from period  $t - 1$  to  $t$ . Taking the first difference of the housing price eliminates census tract fixed effects, controlling for time invariant factors such as the climate, history or location of the census track. To control for reverse causality issue, lagged change of immigration population is used, which is defined as  $\Delta imm\_share_{it-1}$ , the change of immigrant population between  $t - 2$  and  $t - 1$  divided by the total population in that census tract at  $t - 1$ . The coefficient  $\beta_1$  is the parameter of interest. A positive value would imply that neighborhoods that are becoming more immigrant dense are the same ones that experience faster housing value appreciation. An alternative way would be to use the total immigrant population as an independent variable, which could reveal the overall effect of the inflow of immigrants, including both the effect from a changing population and changing neighborhood demographics. By using the share of immigrants instead of the total immigrant population, we will be able to focus on the effect of the change of racial composition, which is more important to understanding the unique influence of immigrants.

Year fixed effect  $Year_t$  is included to control for national events that affect all regions simultaneously. Since different regions might experience drastically different housing market conditions, Metropolitan Statistical Area fixed effect  $MSA_i$  is included. Hence, we can focus on the effect of immigration within each MSA. We also include a complex set of variables  $Tract_{it}$  to control for housing and demographic characteristics of a census tract, such as average rooms, kitchen or plumb facilities, average housing tenure, average education level and average

household income.

Reverse causality and omitted variables are two main concerns when using OLS to estimate the causal effect of immigration on housing prices. First of all, housing prices might be one of the factors that affect immigrants' location choice. A neighborhood with cheaper housing could potentially attract more immigrants, who, on average, have a lower skill level. Although using lagged immigration variable could partially control this issue, immigrants could still anticipate future housing prices while choosing where they want to reside. Secondly, omitted variables such as the change of neighborhood amenities could affect both the immigrants' location choice and the local housing price; and hence bias the relevant coefficients. While taking first difference in housing price should eliminate the effect of time invariant census tract characteristics, it may not control for different trends. This means that it may still be an issue to identify the effect of immigration.

$$\Delta \log(price)_{it} = \alpha + \beta_1 \Delta immi\_share_{it-1} + Tract'_{it} \beta_2 + Year_t + \epsilon_{it} \quad (3.2)$$

$$\Delta immi\_share_{it-1} = \alpha + \theta_1 IV + Tract'_{it} \beta_2 + Year_t + \epsilon_{it}$$

A common way to deal with the reverse causality and the omitted variable problem is by using instrumental variables. An ideal instrument will have a significant impact on housing prices through its effect on immigration population only. We can then implement two stage least square estimation specified in Equation (3.2) to identify the coefficient correctly. Previous literature suggests that immigrants tend to cluster in proximity of early immigrants enclaves from the same source country (Bartel (1989); Card (2001)). This clustering tendency has more

to do with their culture or language preferences than the neighborhood economic trends that might cause endogenous problems. Hence, it could be used as an instrument to study the impact of immigration. Since the Geolytics data does not contain detailed information about the source country of the foreign born population, we use the pull effect of the overall foreign born as our instrument. Census tract is a very small geographic unit that contains less than 4000 residents on average, new immigrants might be attracted to the existing immigrants not only in the specific census tract, but in the nearby regions as well. To capture the overall geographic attraction of a census tract  $i$ , we use a simplified version of the gravity equation derived in Saiz and Wachter (2011):

$$Pull_{it} = \frac{county\_imm\_pop_{ct} - tract\_imm\_pop_{it}}{county\_pop_{ct} - tract\_pop_{it}}. \quad (3.3)$$

$county\_imm\_pop_{ct}$  is the total foreign born population and  $county\_pop_{ct}$  is the total population in county  $c$  at time  $t$ . Similarly,  $tract\_imm\_pop_{it}$  is the total foreign born population and  $tract\_pop_{it}$  is the total population in census tract  $i$  at time  $t$ . Hence  $Pull_{it}$  is the share of historical immigrants in the nearby areas, which summarizes the geographic attraction to new immigrants of census tract  $i$  located in county  $c$  at time  $t$ . We use  $Pull_{it}$  as the instrument variable in Equation (3.2) to recover the causal impact of immigration.

### 3.3.2 Quantile and Two Stage Quantile Estimation

Neighborhoods at the census tract level show great heterogeneity in terms of racial composition, economic development, and political atmosphere. Some of the heterogeneities are observable, while others are not. Due to their differences, neighborhoods might respond to immigration



differently. For example, the main residents of neighborhood with the cheapest housing might be lower class workers, whose political view towards immigrants might be different from other areas in general. If that is the case, then the mean effect estimated using ordinary least squares provides a rather poor estimate of the conditional mean for the poorest neighborhood in the sample. The obvious advantages of using quantile regression is that it can estimate the effect of immigration on the whole conditional distribution of housing price, and it is less affected by outliers.

As described by Koenker and Bassett (1978), we specify the  $\tau$ th conditional quantile function as  $Q_y(\tau|x) = x'\beta(\tau)$ , in which  $\beta(\tau)$  is estimated by solving:

$$\underset{\beta \in \mathbb{R}^p}{Min} \sum_{i=1}^n \rho_{\tau}(\Delta \log(price)_{it} - \alpha - \beta_1 \Delta immi\_share_{it-1} - Tract'_{it} \beta_2 - Year_t). \quad (3.4)$$

In the equation above,  $\rho_{\tau}$  is the piecewise linear loss function:

$$\rho_{\tau}(u) = u(\tau - I(u < 0)). \quad (3.5)$$

The focus of this study is to see whether the effect of immigration varies across different quantiles, which require us to test whether  $\beta_1(\tau) = \tilde{\beta}_1$  for any  $\tau$ . To achieve this end, we need to estimate the standard error of the coefficients to test and construct confidence intervals comparing coefficients describing different quantiles. It is well documented in literature that using empirical quantile function to construct standard errors, as introduced in Koenker and Bassett (1978), usually under-estimate the variance (Bucinsky (1995)). To construct robust standard error, we use bootstrap methods in which both dependent and independent variables

are re-sampled simultaneously.

To estimate the impact of immigration on the entire distribution of housing price while treating the reverse causality and omitted variable issues, we will use two stage quantile regression methods following Portney and Chen (1996). Arias et al (2002) applied similar method to estimate the return to education at different quantiles while addressing simultaneity and measurement error biases. This estimator is essentially a quantile analog of two stage least squares (2SLS) whose large sample properties were established in Powell (1983). The estimation is done by first projecting the endogenous variable  $\Delta immi\_share_{it}$  on the matrix of exogenous variables including the instrument  $Pull_{it}$ , just like the first stage in 2SLS. The first stage OLS projection will then replace  $\Delta immi\_share_{it}$  when solving the model specified in Equation (3-1). We also report the results using instrumental quantile regression introduced by Hansen and Chernozhukov (2005, 2006).

## 3.4 Estimation Details and Empirical Results

### 3.4.1 OLS and IV

The first stage results in table (3.2) show that the existing immigrant population have a significant pull effect on the new immigrants. Specifically, a census tract in a county that had a 1% higher share of foreign born population 10 years ago will have a 0.8% higher share of foreign born population at the present. The F-statistics in the first stage is 239.41, much higher than the critical value of weak IV, suggesting that the instrument has enough power. The Durbin-Hausman test rejected the exogenous consumption at 0.01 significance level, confirming that endogeneity is an issue and a correct identification strategy is necessary.

Both OLS and 2SLS results in table (3.2) suggest that immigrants have a positive effect on housing prices at the census tract level. The results are obtained following Equation (3.1) and Equation (3.2) respectively. The OLS results suggest that a 1 unit increase of the share of the foreign born in the last period is associated with a 0.002 greater log value of the housing price, which is an extremely small effect. In percentage term, it suggests that a 1% increase of the share of foreign born is associated with a 0.002 percent increase in the housing value. Since the median home value in 2000 is only 155,000 dollars, the increase in the housing value is only 3 dollars.

On the other hand, the 2SLS estimate the coefficient to be 0.096, which is a much more substantial impact. It indicates that a 1% increase of the share of foreign born will increase housing prices by 0.096 log value, which is equal to a 0.2% increase of the housing value. The mean value of a single family house in 2000 is 155,000 dollars and the mean change of the share of foreign born is around 0.08. Hence the average housing value appreciation due to immigrants is around 2480 dollars. It is still a small effect, but much more substantial compared with the OLS results.

One possible reason for the difference between OLS and 2SLS could be reverse causality. If immigrants move into a poor neighborhood, which has a slower increase in housing price, then OLS will underestimate the positive effect of immigration. By disentangling the reverse causal effect, 2SLS would return a higher positive effect.

Previous literature, such as Saiz (2003, 2007) and Ottaviano and Peri (2006), have suggested that immigrants bring up housing value at the metropolitan level. However, within a metropolitan area, Saiz (2011) suggests that a 1% increase of immigrant share decreases housing value by 1%. In that paper, Saiz assumes that neighborhoods that are located near the

existing immigrant enclaves are more attractive to new immigrants. He then constructs the instrument variable for each census tract, which is essentially a weighted neighborhood immigrant population using distance to that particular census tract as weights. Compared with our approach, except in regard to different IV, Saiz (2011) also selects a different sample. His paper only focuses on metropolitan areas and years for which the decennial change in immigrant population is substantial (at least 5% of the MSA population). In this paper, since our goal is to test whether different neighborhoods react to immigrants differently, we included all census tracts in all the available years to cover the whole spectrum. These could be the main reasons why the results from our analysis are different from the ones by Saiz (2011).

### **3.4.2 Quantile and Instrumental Quantile Estimation**

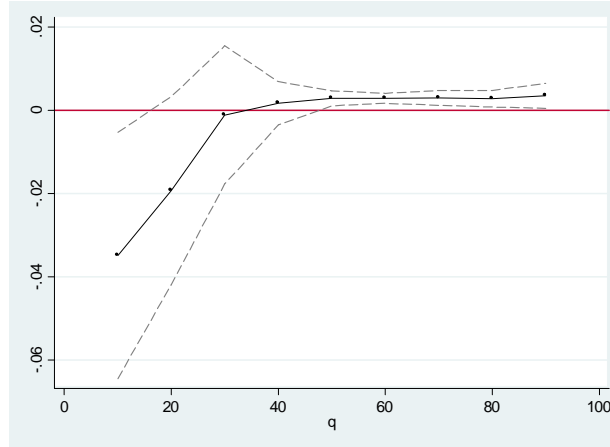
Using quantile regression, we observe significant heterogeneity of immigrant's effect on housing price. In the first part of Table (3.3), we estimate the quantile regression that specified in Equation (3-1) at the deciles from 0.1 to 0.9. Figure (3-1) presents the same results in a more intuitive way. While the OLS suggests that on average immigrant have a positive effect on housing price, quantile regression suggests the effect is quite different for different neighborhoods. For census tracts with median to expensive housing, the quantile regression suggests a similar effect of immigrants both in its direction and size. For example, for census tracts with the housing price at the median, a 1 unit increase in the share of foreign born in the last period is associated with a 0.003 higher log value of the housing price, which is almost negligible. The 95% confidence band is quite narrow around this area, suggesting that we have very precise estimation at the upper end of the distribution.

On the other hand, quantile regression also suggests that for census tracts with very cheap

housing, the share of foreign born population is negatively correlated with housing price and the effects are much more substantial. For example, for census tracts with the housing price at the 10th percentile of the price distribution, a 1% increase of the share of foreign born in last period is associated with a 0.3% lower housing price. Though the coefficients are still significantly different from 0, the 95% confidence interval is relatively wide, suggesting that there is more noise at the lower end of the distribution.

Although it returns significantly larger coefficients, the two stage quantile regression shows a similar pattern as the quantile regression. In the second column of Table (3.3), we report the coefficients of the two stage quantile regression specified in Equation (3-1), replacing endogenous variable  $immi\_Inflow_{it}$  with its projection on the matrix of exogenous variables including the instrument  $Pull_{it}$ , at the deciles from 0.1 to 0.9. Figure (3-2) graphs the coefficients against their quantiles, which shows a clearly rising trend across different points in the conditional distribution of housing price. The figure suggests that at the lower end of the distribution, the effect of increased share of foreign born in the neighborhood decrease housing price. However, the effect is positive for the 70th percentile and higher. For example, if the share of foreign born increases by 1% in the census tract with housing price at the 80th percentile, the value of log housing price will increase by 0.98, which is equal to an increase of 8% in the housing value. The 95% confidence interval bands are very narrow in Figure (3-2), suggesting a very precise estimation.

Figure 3-1: Immigrant Inflow and Housing Price Change: Quantile Regression



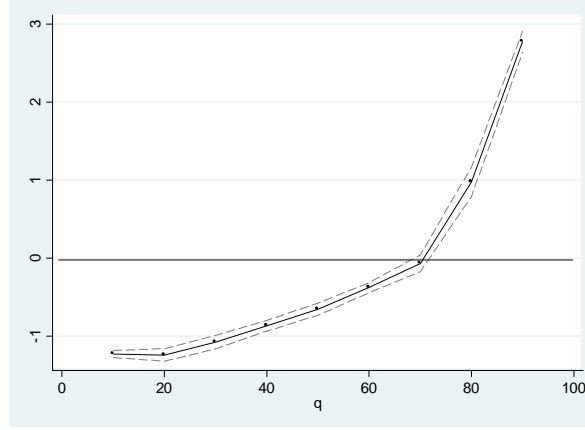
Note: This figure displays the coefficients of immigrant inflow on housing price change at different quantile levels. Coefficient estimates are on the vertical axis, while the quantile index is on the horizontal axis. Coefficients are obtained by solving Equation (3-1). The dash lines are the 95% confident interval constructed using bootstrap method. Data Source: Neighborhood Change Database (NCDB) Tract Data from 1970-2000, GeoLytics.

### 3.5 Further Results: The Change Of Housing Supply

The obvious increasing pattern of the coefficients of immigration along the quantiles suggests that there are neighborhood heterogeneities that are not captured by census tract fixed effect, housing characteristics, or MSA trend. In a housing market, the equilibrium housing price is determined by both the demand and supply. In this section, we look at the supply side stories in search for reasons why neighborhoods react differently to immigrants, using the Neighborhood Change Database.

After immigrants move in, real estate developers might respond differently in different neighborhoods. A rising new housing supply might slow down housing prices and prevent it from growing fast. Since the Neighborhood Change Database does not record housing supply changes directly, we use the change of total housing unit between surveys to infer the new constructions.

Figure 3-2: The Effect of Immigrant Inflow on Housing Price Change: Instrumental Variable Quantile Regression



Note: This figure displays the effect of immigrant inflow on housing price change at different quantile levels, using historical immigrant share in the county as an instrument. Specifically, coefficients are from two stage quantile regression that specified in Equation (3-1), replacing endogenous variable  $immi\_Inflow_{it}$  with its projection on the matrix of exogenous variables including the instrument  $Pull_{it}$ , at the deciles from 0.1 to 0.9. Coefficient estimates are on the vertical axis, while the quantile index is on the horizontal axis. The dash lines are the 95% confident interval calculated using bootstrap.

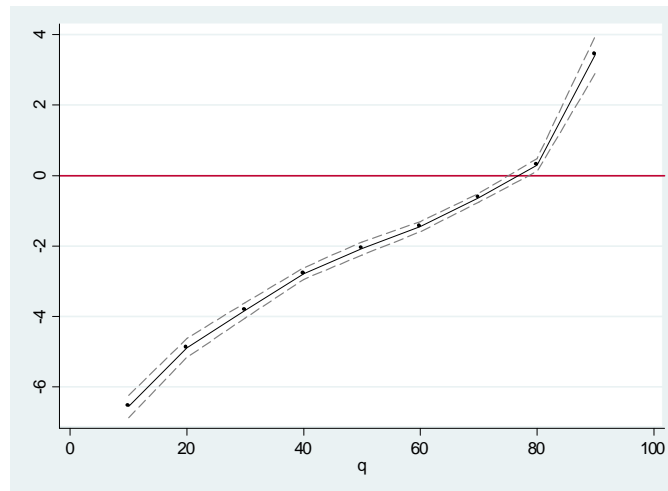
Data Source: Neighborhood Change Database (NCDB) Tract Data from 1970-2000, GeoLytics.

Since developers will try to build houses in neighborhoods with a prospering housing market, and that those areas might be too expensive for new immigrants, reverse causality might be an issue to identify the effect of immigration on housing supply. We also use the pull effect of the existing immigrants in the county 20 years ago as an IV to control this issue.

We test the supply side neighborhood heterogeneity using the two stage quantile regression and report the results in Table (3.4) and Figure (3-3). The dependent variable here is  $D \log(NewHousing_{it})$ , which is the change of the log of the new housing supply, calculated by taking difference of the total housing unit between period  $t$  and  $t - 1$ . Hence the results in Table (3.4) and Figure (3-3) are similar to taking a difference in difference approach. The Figure

(3-3) shows that the more active a neighborhood is in developing new houses, the less negative impact immigrants have on new housing supplies. At the highest quantiles of the distribution of new housing development, immigrants actually increase the new housing supply. Hence they should slow down the rate of housing price increase. If poor neighborhoods happen to be those that attract a lot of new development, then the supply side story could partially explain why the impact of immigration on housing prices is increasing as neighborhoods get wealthier.

Figure 3-3: Immigration and Housing Supply Change: Instrumental Variable Quantile Regression



Note: This figure displays the effect of immigrant inflow on housing supply change at different quantile levels, using historical immigrant share in the county as an instrument. Coefficient estimates are on the vertical axis, while the quantile index is on the horizontal axis. The dash lines are the 95% confident interval.

Data Source: Neighborhood Change Database (NCDB) Tract Data from 1970-2000, GeoLytics.

### 3.6 Conclusion

The findings in this article point to two substantive conclusions. First of all, on average, the inflows of new immigrants have a very limited influence on housing prices at census tract



level. Using OLS, the effect of a 1% increase of the share of immigrants has little effect on housing prices. Using 2SLS, the effect is much more substantial but still small. On average, an increase of 1% in the share of immigrants increases the housing prices by 0.2%. However, if we look at the effect more closely, we will notice that the impacts of immigrants are quite different for different neighborhoods, and looking at the average effect alone will lead to a vastly underestimated effect. Hence, our second conclusion is that there are significant neighborhood heterogeneous effects at different points in the housing price conditional distribution when we use quantile regression. For example, at the 20th percentile of the housing price distribution, a 1% increase in the share of immigrants will lead to a 1.3% decrease in housing price. At the 80th percentile of the housing price distribution, a 1% increase of the share of immigrants will increase housing prices by 1%. Both marginal effects are much larger, but since the effects in some neighborhoods is negative and others are positive, the effect is almost cancelled out when taking the average.

In this paper, we also study the potential reasons behind the heterogeneity, from both the demand side and supply side. Using two stage quantile regression, the results suggest that neighborhoods could react quite differently to immigrants moving in. Neighborhoods with a less active real estate development market tend to build fewer houses after immigrants move in. The slowed supply might slow down the housing price growth for those neighborhoods. These differences between neighborhoods could potentially explain why the effects of immigrants are heterogeneous amongst neighborhoods.

To design an immigration policy that improves the overall welfare of the country, we need to understand who benefits from it and who pays the initial cost. We can then use other social welfare policies to compensate those who are most negatively affected by immigrants.

The results in this paper suggest that the housing price in poor neighborhoods are negatively affected by immigrants, while housing prices in rich neighborhoods are positively affected. In other words, homeowners in poor neighborhoods are hurt by immigrants while homeowners in rich neighborhoods benefit from them. This difference could have a large impact on households' wealth and should be taken into consideration when debating immigration-related policies.

Table 3.1: Summary Statistics by Year

Year = 1970					
Variable	Obs	Mean	Std. Dev.	Min	Max
Average Housing Price	44145	22227.99	68138.86	0	13300000
Total Population	73057	2026.15	2230.55	0	82584
Foreign Born Population	73057	119.60	252.99	0	8427
Population of Mexican Origin	73057	27.65	150.48	0	5932
White Population	73057	1759.68	2017.53	0	81163
Year = 1980					
Variable	Obs	Mean	Std. Dev.	Min	Max
Average Housing Price	59535	152012	2228354	0	20900000
Total Population	73057	2478.58	2109.07	0	89545
Foreign Born Population	73057	182.20	344.88	0	11087
Population of Mexican Origin	73057	105.61	388.50	0	8797
White Population	73057	2036.68	1893.14	0	76199
Year = 1990					
Variable	Obs	Mean	Std. Dev.	Min	Max
Average Housing Price	71798	109042.40	83560.32	14	1178614
Total Population	73057	3404.33	1835.78	0	98443
Foreign Born Population	73057	270.57	487.92	0	11927
Population of Mexican Origin	73057	183.33	544.13	0	13602
White Population	73057	2735.22	1724.42	0	73216
Year = 2000					
Variable	Obs	Mean	Std. Dev.	Min	Max
Average Housing Price	72114	155000.90	118078.90	216	1798581
Total Population	73057	3852.09	1932.13	0	101300
Foreign Born Population	73057	425.80	655.74	0	14636
Population of Mexican Origin	73057	282.05	680.09	0	15492
White Population	73057	2941.20	1805.31	0	70879

Note: This table shows the summary statistics by census year. The unit of analysis is census tract. The housing price is inflation adjusted average housing price, normalized to 2000 dollars.

Data Source: Neighborhood Change Database [NCDB] Tract Data from 1970 to 2000, GeoLytics.

Table 3.2: Immigrant Inflow and Housing Price Change: OLS and 2SLS

	OLS	2SLS	
		Second Stage	First Stage
Lagged Change of Share of Foreign Born Population	0.002*** (0.0007)	0.096*** (0.030)	
Neighborhood Pull Effect			0.797*** (0.310)
P-Value of Durbin-Wu-Hausman test of Endogeneity			0.001
F-Statistics for excluded instrument			239.41
N	106304	106304	106304
Standard errors in parentheses			
* p<0.10; ** p<0.05; *** p<0.01			

Note: This table shows the coefficients of immigrant inflow on housing price following Equation (3.1) and (3.2). For column (1) and (2), the dependent variable is the change of log housing price in the census tract. For column (3), the dependent variable is the change of share of foreign born population. Though not reported in the table, all regression included time fixed effect, MSA fixed effect and census tract characteristics. Data Source: Neighborhood Change Database [NCDB] Tract Data from 1970 to 2000, GeoLytics.

Table 3.3: Quantile Regression

Dependent Variable: First Difference of Log Housing Price			
Quantile	Lagged Change of Share of Foreign Born		
	Quantile Regression	2 Stage Quantile	IV Quantile
0.1	-0.035** (0.015)	-1.228*** (0.023)	-0.720 (0.676)
0.2	-0.019* (0.011)	-1.24*** (0.040)	-0.057 (0.070)
0.3	-0.001 (0.008)	-1.077*** (0.043)	0.644 (0.473)
0.4	0.002 (0.003)	-0.866*** (0.034)	0.832 (0.586)
0.5	0.003 (0.001)	-0.655*** (0.039)	0.990 (0.688)
0.6	0.003*** (0.001)	-0.377*** (0.032)	1.132 (0.795)
0.7	0.003*** (0.001)	0.065 (0.054)	1.280 (0.933)
0.8	0.003*** (0.001)	0.980*** (0.094)	1.452 (1.145)
0.9	0.004** (0.002)	2.775*** (0.068)	-0.058 (0.106)
N=106304			
Standard errors in parentheses			
* p<0.10; ** p<0.05; *** p<0.01			

Note: This table shows the coefficients of immigrant inflow on housing price changes at different quantile levels. In the first column, coefficients are from quantile regression obtained by solving Equation (3-1). In the second and third column, coefficients are from two stage quantile regression and instrumental quantile regression that specified in Equation (3-1), replacing endogenous variable  $immi\_Inflow_{it}$  with its projection on the matrix of exogenous variables including the instrument  $Pull_{it}$ , at the deciles from 0.1 to 0.9. Though not reported in the table, all regression included time fixed effect, MSA fixed effect and census tract characteristics. Robust standard errors are calculated using bootstrap.

Data Source: Neighborhood Change Database [NCDB] Tract Data from 1970 to 2000, GeoLytics.

Table 3.4: Immigration and Housing Supply Change: Instrumental Variable Quantile Regression

Dependent Variable: The Change of Log Housing Supply Increament					
	q10	q20	q30	q40	q50
Immigrant Inflow	-6.559***	-4.893***	-3.824***	-2.784***	-2.069***
	(0.159)	(0.133)	(0.110)	(0.082)	(0.093)
	q60	q70	q80	q95	
	-1.451***	-0.630***	0.301***	3.442***	
	(0.073)	(0.061)	(0.092)	(0.263)	

Note: This table displays the effect of immigrant inflow on housing supply change at different quantile levels. Though not reported in the table, all regression included time fixed effect, MSA fixed effect and census tract housing characteristics.

Data Source: Neighborhood Change Database (NCDB) Tract Data from 1970-2000, GeoLytics.

# Bibliography

- Acemoglu, D. and Angrist, J. 2001. How large are human-capital externalities? evidence from compulsory-schooling laws. *NBER Macroeconomics Annual 2000, Volume 15*, pages 9–74.
- Altonji, J. G. and Card, D. 2007. The effects of immigration on the labor market outcomes of less-skilled natives. *Immigration, Trade, and the Labor Market*, page 201.
- Anderson, T. and Rubin, H. 1949. Estimators of the parameters of a single equation in a complete set of stochastic equations. *The Annals of Mathematical Statistics*, 21.
- Angrist, J. D. and Krueger, A. B. 1991. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014.
- Bartel, A. P. 1989. Where do the new us immigrants live? *Journal of Labor Economics*, pages 371–391.
- Beine, M., Docquier, F., and Rapoport, H. 2001. Brain drain and economic growth: Theory and evidence. *Journal of Development Economics*, 64(1):275 – 289.
- Borjas, G. 1999. The economic analysis of immigration. *Handbook of labor economics*, 3:1697–1760.

- Borjas, G. J. 1987. Self-selection and the earnings of immigrants. *The American Economic Review*, pages 531–553.
- Bound, J. and Holzer, H. J. 2000. Demand shifts, population adjustments, and labor market outcomes during the 1980s. *Journal of labor Economics*, 18(1):20–54.
- Bound, J., Jaeger, D. A., and Baker, R. M. 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association*, 90(430):443–450.
- Buchinsky, M. 1995. Estimating the asymptotic covariance matrix for quantile regression models a monte carlo study. *Journal of Econometrics*, 68(2):303–338.
- Calomiris, C. W., Longhofer, S. D., and Miles, W. 2012. The housing wealth effect: The crucial roles of demographics, wealth distribution and wealth shares. *NBER Working Paper*, (w17740).
- Card, D. 1990. The impact of the mariel boatlift on the miami labor market. *Industrial and Labor Relations Review*, 43(2):245–257.
- Card, D. 2001. Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64.
- Chen, L.-A. and Portnoy, S. 1996. Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equation models. *Communications in Statistics-theory and Methods*, 25(5):1005–1032.
- Chernozhukov, V. and Hansen, C. 2005. An iv model of quantile treatment effect. *Econometrica*, 73:379–398.



- Chernozhukov, V. and Hansen, C. 2006. Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, 132(2):491–525.
- Chiquiar, D. and Hanson, G. H. 2005. International migration, self-selection, and the distribution of wages: Evidence from mexico and the united states. *Journal of Political Economy*, 113(2).
- Dahl, G. B. 2002. Mobility and the return to education: Testing a roy model with multiple markets. *Econometrica*, 70(6):2367–2420.
- Edwards, L. N. 1978. An empirical analysis of compulsory schooling legislation, 1940-1960. *Journal of Law and Economics*, 21(1):203–222.
- Eide, E. and Showalter, M. H. 1998. The effect of school quality on student performance: A quantile regression approach. *Economics letters*, 58(3):345–350.
- Greenwood, M. J. 1975. Research on internal migration in the united states: a survey. *Journal of Economic Literature*, pages 397–433.
- Greenwood, M. J. 1997. Internal migration in developed countries. *Handbook of population and family economics*, 1:647–720.
- Imbens, G. W. and Angrist, J. D. 1994. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):pp. 467–475.
- Katz, M. 1976. *A history of compulsory education laws*. Phi Delta Kappa Educational Foundation.
- Kerr, S. P. and Kerr, W. R. 2011. Economic impacts of immigration: A survey.

- Koenker, R. and Bassett Jr, G. 1978. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.
- Lleras-Muney, A. 2002. Were compulsory attendance and child labor laws effective? an analysis from 1915 to 1939. *Journal of law and economics*, 45(2):401–437.
- Lleras-Muney, A. 2005. The relationship between education and adult mortality in the united states. *The Review of Economic Studies*, 72(1):189–221.
- Longhi, S., Nijkamp, P., and Poot, J. 2006. The impact of immigration on the employment of natives in regional labour markets: A meta-analysis. *IZA Working Paper*.
- Lucas, R. 1997. Internal migration in developing countries. *Handbook of population and family economics*, 1:721–798.
- Machin, S., Salvanes, K. G., and Pelkonen, P. 2012. Education and mobility. *Journal of the European Economic Association*, 10(2):417–450.
- Magrini, M.-B. and Lemistre, P. 2013. Distance–income migration trade-off of young french workers: An analysis per education level. *Regional Studies*, 47(2):282–295.
- Malamud, O. and Wozniak, A. 2012. The impact of college education on geographic mobility: evidence from the vietnam generation. *Journal of Human Resource*.
- Mazumder, B. 2008. Does education improve health? a reexamination of the evidence from compulsory schooling laws. *Economic Perspectives*, 32(2).
- Mazumder, B. 2011. The effects of education on health and mortality. *Nordic Economic Policy Review*.

- McHenry, P. 2013. The relationship between schooling and migration: Evidence from compulsory schooling laws. *Economics of Education Review*, 35:24–40.
- Niedomysl, T. 2011. How migration motives change over migration distance: evidence on variation across socio-economic and demographic groups. *Regional Studies*, 45(6):843–855.
- Oreopoulos, P. 2006. Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *The American Economic Review*, pages 152–175.
- Peri, G. 2007. Immigrants’ complementarities and native wages: Evidence from california. Technical report, National Bureau of Economic Research.
- Pischke, J.-S. and Von Wachter, T. 2008. Zero returns to compulsory schooling in germany: evidence and interpretation. *The Review of Economics and Statistics*, 90(3):592–598.
- Powell, J. L. 1983. The asymptotic normality of two-stage least absolute deviations estimators. *Econometrica: Journal of the Econometric Society*, pages 1569–1575.
- Rothenberg, T. J. 1984. Approximating the distributions of econometric estimators and test statistics. *Handbook of Econometrics*, 2:881–935.
- Saiz, A. 2003. Room in the kitchen for the melting pot: Immigration and rental prices. *Review of Economics and Statistics*, 85(3):502–521.
- Saiz, A. 2007. Immigration and housing rents in american cities. *Journal of Urban Economics*, 61(2):345–371.
- Saiz, A. and Wachter, S. 2011. Immigration and the neighborhood. *American Economic Journal: Economic Policy*, pages 169–188.

- Schwartz, A. 1973. Interpreting the effect of distance on migration. *The Journal of Political Economy*, pages 1153–1169.
- Staiger, D. and Stock, J. H. 1997. Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586.
- Stephens Jr, M. and Yang, D.-Y. 2014. Compulsory education and the benefits of schooling. *American Economic Review*, 104(6):1777–92.
- Steven Ruggles, J. Trent Alexander, K. G. R. G. M. B. S. and Sobek, M. 2010. Integrated public use microdata series: Version 5.0 [machine-readable database]. *Minneapolis: University of Minnesota*.
- Todaro, M. 1980. *Internal migration in developing countries: a survey*. University of Chicago Press.