Assimilation Estimation Is Less Biased Than Curved: Using Cross-Sectional Data in the U.S. from 1990 to 2010

David You
University of Missouri

Abstract

Assimilation is the process in which immigrants improve earnings as they become more adapted to the host country society. Cross-sectional studies show that immigrants have lower earnings upon arrival and faster earnings growth compared to natives. Longitudinal studies conclude that estimations with cross-sectional data are positively biased with decreasing cohort quality and negatively selected outmigration. I reproduce the estimation with the most recent data. The estimation shows "bias" with cohort fixed effect added, but there is no sign of decreasing cohort quality nor enough outmigration. Next, I propose a non-parametric method to make wage distribution visually comparable across cohorts and time. I find that the linear specification of assimilation is misleading. Finally, I revisit the classic model with a quadratic assimilation term and expand it to explore the assimilation process's heterogeneity. I find that the "bias" disappeared with a quadratic assimilation effect. The assimilation effect is sensitive to age at arrival and country of origin.

I. Introduction

Forty-seven million people living in the U.S. were born outside of the U.S. as of 2015. Throughout American history, immigrants played vital roles in the development of all aspects of this country. However, surviving in the U.S. is a challenge for new immigrants. Learning and adapting to a new environment is much needed to explore immigrants' full potential thoroughly. Assimilation is how immigrants learn skills that bridge the gap between them and the receiving society.

Assimilation can be measured in terms of socioeconomic status, spatial concentration, language assimilation, and intermarriage (Waters & Jimenez, 2005). This paper focuses on the economic assimilation of immigrants, which is measured in earnings growth. The idea of economic assimilation is that immigrants may not be fully adapted to the U.S. labor market upon arrival due to a lack of necessary knowledge or skills. Immigrants acquire skills to accommodate the U.S. labor market as they stay longer. As a result, their earnings growths are due to increases in their job experience and better adaptation to the job market. For example, the increase in earnings may be related to English proficiency improvement (Bleakley & Chin, 2010), intermarriage with natives (Qian & Lichter, 2007), or cultural factors that can make immigrants successful in the U.S. labor market.

Measuring accumulation in human capital directly is difficult, as it is often not observable. It is also difficult to pin down a set of skills necessary for assimilation because immigrants are heterogeneous. The skills needed for assimilation depend on the idiosyncratic situation. For example, a foreign professor needs to improve his English to express ideaa, while a foreign gardener may need to find a boss who can handle communications with clients. Both cases enhance the efficiency of working, which is essential to economic success. Fortunately, these skills share three features. First, most of them are communication- or cultural-related. Second, most natives have acquired these skills. Third, immigrants are likely to experience more significant growth in skills as they stay longer in the U.S. As a result, we can proxy the accumulation of human capital due to assimilation using earnings growth due to length of stay.

The U.S. Census and the American Community Survey (ACS) ask immigrants about their earnings and arrival times. The difference between the arrival year and the year of the survey tells the length of stay. The relationship between the length of stay and earnings level can be found after controlling for age and other earnings-determining factors.

Both the U.S. Census and the ACS provide cross-sectional data. Critics claim that the estimation of assimilation using cross-sectional data is problematic. The potential problem stems from the cohort bias caused by changing cohort quality and the outmigration bias caused by selected immigrants leaving the U.S.. In cross-sectional data of a given year, differences in length of time in the U.S. imply differences in date of arrival or "arrival cohort." Borjas (1985) points out that the more permissive immigration policy since 1970 has allowed more low-skill immigrants from Asia and Latin America to come to the U.S. As a result, older immigrant cohorts have higher earnings than newer cohorts because of higher productivity, not assimilation. The estimation is positively biased.

When repeated cross-sectional data from multiple years are available, cohort bias due to cohort quality changes can be eliminated by sampling people from the same cohort at different times. However, estimation of assimilation may still be biased insofar as immigrants leave the U.S., changing the cohort's composition. If we assume low-skilled workers have a more challenging time in the host country and are more likely to leave the U.S., the average earnings of the cohort increases as the fraction of high-skilled workers increases, which is not due to assimilation.

The conventional way of correcting both cohort and outmigration biases is to construct panel data (Borjas G. , 1985; Hu, 2000; Edin, Fredriksson, & Aslund, 2003; Duleep & Dowhan, 2002; Constant & Massey, 2003). Panel data include observations of the same sample over time. As a result, panel data can track earnings and exclude outmigration for a set of respondents within a period. Forming a panel dataset in the U.S. raises some technical challenges as there is no public panel dataset of immigrants in the U.S.. Analysts must match observations between cross-sectional data to construct panel data and merge the data with earnings data. Due to difficulties in matching observations, the method of construction may create selection bias.

In this paper, there is a similar "bias" in assimilation estimation when comparing estimations using cross-sectional and repeated cross-sectional analysis. However, partially control for differences between cohorts does not mitigate the bias. The interpretation of the bias is in question. I implement a non-parametric method to estimate assimilation with no constrain in the specification of assimilation. This method is inspired by Chiquiar and Hanson (2005), in which they adjust the earnings distribution of immigrant cohorts based on observed characteristics. The result confirms that the immigration assimilation process is not linear. The process is fast initially and slows down in later years, indicating the conventional model is not correctly specified. Using this result, I revisit the conventional model and find that allowing curvature in assimilation reduces the estimation bias significantly.

The remainder of this paper proceeds as follows. Section II provides the literature review and explains the biases in assimilation estimation in detail. Section III describes the dataset and some key variables. Section IV presents the structure of the conventional regression method and the method that adjusts migrant characteristics' distribution. Section V explains the results of both methods. Section VI concludes.

II. Literature Review

Whether new immigrants can succed in the U.S. is always raised since many immigrants are from developing countries. The two most recent immigrants' waves are the late 19th century to the early 20th century and the late 20th century to the early 21st century. There are some common features shared between these two waves. First, national survey data record the change in the immigrant population from both waves. Using the 1970 Census, an early article by Chiswick (1978) finds that immigrants in the country longer have higher earnings than new immigration cohorts. In response, Borjas (1985) argues that the improvement in earnings is subject to the cohort bias. The problem stems from the fact that years since arrival and arrival date are perfectly correlated in cross-sectional data. When the cohort's quality is declining over time, recent immigration cohorts are subject to lower earnings trajectories. Projecting their future earnings to achieve the earnings levels of older immigrants is too optimistic. Indeed, the current immigration wave contains more low-skilled workers from Asia and Latin America, whereas older migrants from Europe were generally high-skilled. The critical factor is skill difference, not the length of stay. In an article considering migrant assimilation during the age of mass migration, 1880-1920, Abramitzky, Boustan, and Eriksson (2014) show the same effect. Immigrants who arrived earlier in that period had higher earnings than immigrants who arrived later because the primary source countries of the majority of immigrants shifted from Britain and Germany to Russia and Italy[1].

The cohort bias can be addressed by merging cross-sectional data from different survey years to form repeated cross-sectional data. With repeated cross-sectional data, we can estimate the assimilation effect by controlling the cohort fixed effect. Ideally, we are sampling from the same group of people at different times to estimate their earnings growth. However, as noted above, the estimation is subject to the second source of bias – the outmigration bias. Abramitzk et al. (2014) show that negative selected outmigration biases the estimation of assimilation. Immigrants who are not successful in the U.S. job market are more likely to emigrate. When immigrants with lower earnings leave, the average earnings increase.

Since Borjas (1985), it has been recognized that cross-sectional studies have severe disadvantages in estimating the assimilation. As summarized by Abramitzky and Boustan (2017), "Much of the contemporary literature on immigrant assimilation has been focused on solving methodological issues in order to properly measure changes in immigrant earnings with time spent in the United States."

Panel data provide information about the same individuals over time. Also, it is easy to identify and exclude observations that leave the survey. As a result, the estimation of assimilation is not subject to cohort and outmigration biases. Many economists (Borjas G. , 1985; Duleep & Dowhan, 2002; Edin, Fredriksson, & Aslund, 2003; Lubotsky,

[1] Immigrants from West Europe were more skilled than immigrants from East and South Europe at that time.

2007; Kim, 2009; Abramitzky, Boustan, & Eriksson, 2014; Hu, 2000) believe regression results using longitudinal panel data are more compelling than using cross-sectional data. However, panel surveys covering immigrants are not available in the U.S.

Economists construct longitudinal panel data from multiple resources in the U.S. The critical step is to find matching observations from cross-sectional data for two periods or match observations between cross-sectional data and panel data. For example, Kim (2009) exploits the Current Population Survey (CPS) overlapping rotation feature, where individuals who are covered twice by the CPS in one year can be matched. Then, he forms a longitudinal panel dataset using income data from the Survey of Income and Program Participation (SIPP).

Articles using constructed longitudinal data generally show moderate or no assimilation in contrast to studies using cross-sectional data. The findings suggest that Chiswick's estimation may just be due to bias due to decreasing cohort quality and negatively selected outmigration in the 1970s.

Abramitzky, Boustan, and Eriksson (2014) match individuals between 1900, 1910, and 1920 Census to form a panel dataset. All matched individuals must have stayed in the U.S. for at least ten years. They impute immigrants' earnings based on the occupation reported. Using this panel data, they find that immigrants' initial penalty is small, and there is no sign of assimilation. Their methods have two flaws. First, if the assimilation process is significant in the first ten years and diminishes after that, dropping individuals arrived less than ten years earlier drops observations when assimilation is most significant. Second, although the example above assumes people with lower earnings will leave, there are many reasons that migrants may return home.

For example, Borjas and Bratsberg (1996) find that the outmigration population does not necessarily consist of immigrants with the lowest earnings. When immigrants from the sending country are negatively selected, people with lower earnings have greater incentives to migrate. After arrival to the host country, people with relatively higher earnings are more likely to return or remigrate. The opposite holds too. When immigrants are highly skilled in their home country, people making relatively lower earnings in the host country are more likely to return or remigrate.

Meanwhile, some studies using longitudinal data find that assimilation exists. Duleep and Dowhan (2002) use the Current Population Survey (CPS) in 1994 and obtain longitudinal data on the earnings of immigrants from the Social Security Administration (SSA). No data are dropped to meet a minimum year of stay. Estimates using longitudinal data formed from CPS and SSA from 1988 to 1993 imply that immigrants experience extra earnings growth in the first ten years of stay compared to native workers. Lubotsky (2011) finds that the difference in earnings growth rates between skilled and unskilled workers leads to an overly pessimistic picture of more recent immigrants' skills. He suggests that high-skilled workers' earnings grow much faster than the earnings of low skilled workers during this period. The widening earnings gap results from the higher return to skills and widening earnings inequality, not declining cohort quality. If the return to skill were stable, he argues that the assimilation effect would be more apparent, especially for low skilled immigrants.

Many European countries have longitudinal panel data of immigrants. Constant and Massey (2003) question if negatively selected outmigration distorts assimilation effect estimates. They use the first 14 waves of German Socioeconomic Panel data

(1984-1997). The regression is run separately for stayers and all immigrant workers. The estimated assimilation effects from both regressions are tested with an equality test. Their analysis finds that the estimated coefficients are not statistically different. Since immigrants make up about 20% of the population surveyed, it is unlikely that their results are due to the small sample size. They conclude that return migration by immigrants does little to change the assimilation estimation significantly in Germany.

This paper presents evidence that is inconsistent with the findings from previos longitudinal studies. I verify such findings using an alternative method to account for potential biases using cross-sectional and repeated cross-sectional data. I find the conventional model using repeated cross-sectional data can be modified to estimate assimilation consistently.

III. Data and Description

A. Census and ACS Data

The analyses use the 5 percent public use microdata sample (PUMS) of the 1990 and 2000 U.S. Census and the American Community Survey (ACS) of 2010. U.S. censuses are conducted on April 1st of each census year. The PUMS uses a 1-in-20 national random sample. ACS conducts annual surveys of 3.5 million households (about 2.7 percent of all households). The Census long-form was eliminated after 2000, but the ACS inherited many questions asked in the long form of the 2000 decennial census, making it the most consistent data for the subsequent decade. For analysis of the nation as a whole, Census and ACS are two of the best resources. I choose the data from the Integrated Public Use Microdata Series (IPUMS), a dataset from the University of Minnesota (Ruggles, et al., n.d.) that collects, preserves, and harmonizes

Census and ACS data from multiple years. The observations are representative of the entire U.S. population across time, and they have a comprehensive collection of socioeconomic variables. All observations are weighted to assure the sample is representative of the U.S. population.

In this paper, earnings used are the annual pre-tax earnings, that is, money received as an employee. The sources include wages, salaries, commissions, cash bonuses, tips, and other money income from an employer. Earnings in Census reflect the income during the previous calendar year, while earnings in ACS refer to income for the past 12 months. The earnings are adjusted for inflation with the Consumer Price Index[2]. Other socioeconomic variables include the highest educational attainment, marital status, race, gender, age, and usual residence census division.

Immigrants are defined as people born outside of the U.S. The length of stay is calculated as the difference between the year of observation and the year of arrival. Immigrants are grouped into 9 cohorts based on the arrival time: before 1970, 1970-74, 1975-79, 1980-84, 1985-1990, 1991-95, 1996-2000, 2001-05, and cohort 2006-10.

B. Demographic Description

Cross-sectional data show demographic information of immigrants in the U.S. from 1990 to 2010. Table 1 summarizes the population of natives and immigrant cohorts. The predicted numbers of deaths and outmigration in each decade are listed below the population. The death prediction is based on the age structure of each immigrant cohort. The number of deaths is calculated with the number of immigrants in each age

[2] The CPI data is available in the IPUMS dataset. Search "CPI99" on https://usa.ipums.org/usa-action/variables/live_search

cell and the fraction of deaths in the given age cell from the corresponding life table[3] of that year for ten years. Outmigration is estimated as the reduction in cohort size that is not due to death. Notice that the sizes of the cohort 1985 to 1990 and the cohort 1996 to 2000 increased between two census years. The increase is due to missed counts of immigrants who arrived in the U.S. in the census year after the census date. For example, an immigrant who arrived in October 2000 and missed the Census survey will tell Census 2010 that he arrived in 2000. These immigrants are captured in the next survey, causing an increase in the cohort population. Excluding these cohorts, the cohort population size shrinks at a rate of about one percent per year.

Prior studies[4] show 20 to 70 percent of migrants return or re-migrate within the first ten years after arrival. The outmigration estimates in Table 1 are much lower. However, there is no inconsistency between the estimates in Table 1 and the estimates from previous articles. Since many short-term immigrants would leave in less than ten years, most of them may not answer any decennial survey. The calculation relies on population counts both at the beginning and the end of a decade. As a result, the outmigration in Table 1 represents those who have stayed for at least ten years.

The native population has a net growth of 0.89 percent per year from 1990 to 2010. Immigration growth is much more substantial. The immigrant population grew at 4.6

---

[3] The life table is for all people in the U.S. in a given year.
1990 - https://www.cdc.gov/nchs/data/lifetables/life89_1_3.pdf
2000 - https://www.cdc.gov/nchs/data/nvsr/nvsr51/nvsr51_03.pdf

[4] See Ahmed and Robinson (1994), Beaujot and Rappak (1989), Borjas and Bratsberg (1994), Dustmann, Fadlon, and Weiss (2011), and Edin, LaLonde, and Aslund (2000)

percent per year. There are nine immigrant cohorts based on their years of arrival, as mentioned in the previous section. The size of immigration peaked in the 1996 to 2000 period, during which time over seven million immigrants arrived in the U.S. During the 2006 to 2010 period, around six million new immigrants arrived. The highest rate of outmigration estimated is 12 percent for the 1991 to 1995 cohort from 2000 to 2010. Notice that the estimated outmigration between 1990 to 2000 is extremely small. The reason is probably the economic boom at the end of the 1990s or some unknown changes between the censuses.

Table 1 Population Size of Each Cohort by Year (in thousands)

Population Size of Each Cohort by Year (in thousands)

			Immigrants Date of Arrival								
Year	**Natives**	**IMM Total**	**<1970**	**1970 -74**	**1975 -79**	**1980 -84**	**1985 -90**	**1991 -95**	**1996 -2000**	**2001 -05**	**2006 -10**
1990	225,201	22,906	7,600	2,454	3,114	4,211	5,527	-	-	-	-
Death		(1,752)	(1,255)	(111)	(114)	(129)	(143)	-	-	-	-
Emigrate		(463)	(395)	(58)	(7)	(3)	-	-	-	-	-
2000	246,766	34,656	5,950	2,285	2,993	4,079	6,520	6,049	6,780	-	-
Death		(2,120)	(1,084)	(158)	(161)	(179)	(214)	(175)	(149)	-	-
Emigrate		(1,661)	(96)	(8)	(178)	(205)	(444)	(730)		-	-
2010	265,215	44,134	4,770	2,119	2,654	3,695	5,862	5,144	7,151	6,807	5,932

Note: The sources of data are the 1990 U.S. Census, the 2000 U.S. Census, and the American Community Survey of 2010. Populations are in thousands. The cohort categorization rule before 1990 is based on the survey question from the census in 1990, in which respondents were asked "When did this person come to the United States to stay?" The answers include year ranges, 1987 to 1990, 1985 or 1986, 1982 to 1984, 1980 or 1981, and others. For immigrants with an arrival range, the last year in the range is used as the arrival year indicator in the data. In 2000 and 2010, respondents answer, "When did this person come to live in the United States?" with the answer recorded as the calendar year. Predicted death is calculated with the life table based on the age structure.

Table 2 presents the race composition of immigrants and natives based on our samples. Most natives are white. In contrast, less than 20 percent of the most recent wave of immigrants are white. The second-largest race in the U.S. is African American, representing approximately 11 to 12 percent of the native population. The proportion of black immigrants has incresed, but black accounts for less than 10 percent of the immigrant population. If immigration continues at its current rate, the black population will be surpassed by Asians and Latin Americans shortly.

The current immigration wave includes many immigrants from Asia and Latin America. Table 2 shows three race categories, Chinese, other Asian, and Hispanic. None of these groups account for more than 4.3 percent of the native population (less than 0.5 percent for Asian). However, the Asian and Hispanic groups make up more than 40 percent of immigrants arriving since 1970.

Table 2 Selected Race Composition of Each Cohort by Year

			Date of Arrival								
Year	**Native**	**IMM total**	**<1970**	**1970-74**	**1975-79**	**1980-84**	**1985-90**	**1991-95**	**1996-2000**	**2001-05**	**2006-10**
White											
1990	84.5%	32.3%	57.5%	21.2%	20.9%	15.7%	21.4%	-	-	-	-
2000	81.6%	23.6%	52.9%	22.3%	19.3%	13.1%	13.9%	19.3%	22.1%	-	-
2010	78.7%	20.2%	50.2%	23.0%	20.6%	14.5%	13.8%	16.6%	16.3%	15.5%	18.4%
Black											
1990	11.0%	7.4%	4.4%	9.7%	8.3%	10.3%	7.7%	-	-	-	-
2000	11.6%	6.9%	4.3%	8.4%	7.1%	8.8%	7.4%	6.8%	6.7%	-	-
2010	12.3%	8.2%	4.5%	8.6%	7.3%	9.6%	8.5%	7.6%	8.3%	9.1%	8.9%
Chinese											
1990	0.1%	5.1%	2.6%	5.2%	6.7%	7.9%	9.0%	-	-	-	-
2000	0.1%	5.1%	2.6%	4.5%	6.0%	6.9%	6.6%	6.6%	6.4%	-	-
2010	0.2%	5.5%	2.6%	4.4%	5.6%	6.5%	6.1%	6.2%	5.8%	5.3%	7.6%
Other Asian											
1990	0.2%	14.6%	4.2%	16.8%	22.1%	22.7%	20.8%	-	-	-	-
2000	0.3%	15.0%	4.6%	15.4%	19.0%	21.2%	17.0%	19.5%	17.0%	-	-
2010	0.5%	16.9%	5.0%	15.8%	19.9%	22.6%	17.9%	19.7%	15.9%	17.8%	23.6%
Hispanic											
1990	2.0%	20.0%	18.1%	23.4%	19.1%	21.1%	19.4%	-	-	-	-
2000	2.4%	21.2%	20.2%	21.6%	17.8%	20.9%	22.3%	20.6%	21.2%	-	-
2010	4.3%	29.3%	25.4%	30.1%	25.4%	28.5%	31.5%	29.2%	32.4%	30.8%	24.5%

Note: Census of 1990 and 2000, and the American Community Survey of 2010. The population includes people age 25 and above. Observations with missing education attainment are dropped. Unselected race categories include American Indian, Japanese, and other races.

The country of origin affects immigrants' performance in the receiving country (Borjas & Bratsberg, 1994; Abramitzky, Boustan, & Eriksson, 2014). Many factors, such as the development level, culture, and relationship with the U.S., determine how well an immigrant can fit into the U.S. There are two differences between the current wave and the mass migration wave between 1880 and 1920. First, immigrants during the earlier period were mostly from developed countries in Europe. Second, Asian and Hispanic immigrants and their descendants have unique physical peculiarities that distinguish them from the native-born white population. The shifting in race composition lays the foundation for changing cohort quality, which raises questions for the estimation of assimilation.

Table 3 shows the education composition of natives and immigrant cohorts. Education has three categories for simplicity, instead of five education attainment levels used in the analysis. Only people over 24 years old are included. Immigrants have heavy tails on both ends of the distribution. Natives are over-represented with some college or bachelor's degree.

No evidence shows decreasing cohort quality in education attainment. When comparing immigrants who stayed less than ten years in 1990, 2000, and 2010, later cohorts contain more people with bachelor degrees or advanced degrees. Since education attainment is an essential factor in skills, the cohort quality has been increasing since 1980. It is reasonable to have a pessimistic projection by looking at immigrants in the 1970s and 1980s since they have less education than previous immigrants. However, the trend turns to positive from 1980 to 2010. The assimilation estimates should be negatively biased if Borjas (1985) were right.

Table 3 Education Composition by Cohort and Year

Year	Native	Date of Arrival <1970	1970-74	1975-79	1980-84	1985-90	1991-95	1996-2000	2001-05	2006-10
High School or Less										
1990	54.1%	62.0%	61.2%	59.7%	60.9%	56.6%	-	-	-	-
2000	46.8%	55.6%	54.7%	54.4%	58.8%	61.6%	57.6%	52.1%	-	-
2010	40.6%	50.8%	51.8%	48.7%	56.1%	56.1%	55.5%	56.6%	55.2%	48.1%
Some College or College										
1990	38.9%	30.2%	30.3%	31.9%	31.0%	32.3%	-	-	-	-
2000	44.6%	34.4%	35.9%	36.4%	33.0%	29.8%	31.0%	33.5%	-	-
2010	49.1%	37.6%	37.9%	39.6%	35.1%	35.2%	34.0%	32.3%	32.9%	36.7%
Master's Degree and Higher										
1990	7.1%	7.7%	8.5%	8.4%	8.1%	11.1%	-	-	-	-
2000	8.7%	10.0%	9.4%	9.0%	8.3%	8.7%	11.4%	14.4%	-	-
2010	10.3%	11.7%	10.4%	11.7%	8.7%	8.7%	10.3%	11.0%	12.0%	15.2%

Note: Sources of data are the Censuses of 1990 and 2000, and the American Community Survey of 2010. The data include people with age strictly greater than 24.

Some studies claim that the shift in the country of origin changes the quality of education of immigrants. A Mexican immigrant with a college degree is probably less competitive than a college graduate from Germany because higher education is more advanced in Germany. This issue can be addressed by controlling for educational attainment and country of origin.

Table 3 also shows that the proportion of immigrants with the lowest educational attainment in a cohort drops from 1990 to 2010. This pattern holds for all cohorts, except the 1985 to 1990 cohort. The negatively selected outmigration of immigrants is not the only explanation. Other explanations include that most immigrants who die have lower levels of education; some immigrants finished higher education or both. These two reasons can be distinguished by comparing the older cohorts with newer cohorts. Since the decrease in the size of older cohort is dominated by death, the decrease in the proportion of less educated people in older cohorts is primarily due to death of less educated. On the other hand, the more recent cohorts are generally made up of young people who are less likely to die. The education distribution of cohort 85-90, 91-95, and 96-00 does not show substantial negatively selected outmigration in the early age of immigration. Negatively selected outmigration by education attainment is moderate, consistent with Constant and Massey's (2003) findings in Germany.

Age is essential in the assimilation study because it determines earnings both as a proxy for experience and the aging process. In general, older people are more experienced and productive with more human capital. Table 4 shows the age composition of natives and immigrants. Immigrants are overrepresented in the working-age range from 20 years to 60 years of age in all three cross-sectional datasets. Most immigrants, legal or illegal, asylee or economic immigrants, choose the U.S. partially for its developed economy and abundant job opportunities. The most recent immigrants are much younger than both the native population and previous immigrants.

Chueng et al. (2011) find that younger immigrants acculturate faster. As a result, the economic assimilation process should be faster when immigrants are young. In contrast, older immigrants are less able to adjust to the host culture. This finding raises the question, "Is the assimilation effect linear as newer cohorts are over-represented with young immigrants?" With a linear model, the assimilation effect is implicitly assumed to contribute to wage growth at a constant rate over time. Recently arrived cohorts may have faster assimilation rates because they have more young immigrants than older cohorts, but estimates may appear that the difference in wage growth is due to cohort differences.

Table 4 Age Composition of Natives and Immigrants by Cohort and Year

			Immigrant Arrival Year								
Age	Native	Imm	<1970	1970-74	1975-79	1980-84	1985-90	1991-95	1996-2000	2001-05	2006-10
1990											
0-9	15.6%	5.3%	-	-	-	5.8%	17.7%	-	-	-	-
10-19	14.3%	11.0%	-	9.7%	18.7%	17.5%	17.3%	-	-	-	-
20-29	15.6%	20.4%	8.3%	20.0%	20.6%	27.8%	31.3%	-	-	-	-
30-39	16.5%	21.2%	14.6%	26.3%	31.5%	26.9%	18.0%	-	-	-	-
40-49	12.5%	15.2%	19.0%	25.5%	16.2%	11.7%	7.7%	-	-	-	-
50-59	8.7%	10.2%	18.7%	10.5%	6.8%	5.2%	3.8%	-	-	-	-
60-69	8.4%	7.9%	17.2%	4.7%	3.6%	3.0%	2.7%	-	-	-	-
70+	8.5%	8.8%	22.2%	3.2%	2.8%	2.0%	1.4%	-	-	-	-
2000											
0-9	15.5%	4.4%	-	-	-	-	-	7.0%	16.1%	-	-
10-19	15.0%	9.9%	-	-	-	4.6%	15.5%	17.0%	17.7%	-	-
20-29	12.8%	19.4%	-	7.1%	16.7%	17.5%	21.3%	29.7%	31.8%	-	-
30-39	14.4%	22.5%	9.1%	21.9%	22.5%	31.1%	32.9%	23.6%	18.3%	-	-
40-49	14.9%	18.1%	17.4%	29.0%	34.1%	27.5%	17.4%	11.6%	8.6%	-	-
50-59	10.9%	11.3%	23.1%	25.9%	15.7%	10.8%	6.8%	5.5%	3.8%	-	-
60-69	7.2%	7.2%	22.3%	10.0%	6.4%	4.7%	3.3%	3.2%	2.3%	-	-
70+	15.2%	7.3%	28.1%	6.2%	4.6%	3.7%	2.7%	2.4%	1.3%	-	-
2010											
0-9	14.8%	2.7%	-	-	-	-	-	-	0.1%	6.3%	12.8%
10-19	14.9%	7.4%	-	-	-	-	0.1%	7.3%	14.3%	14.8%	14.1%
20-29	13.4%	15.9%	-	-	-	5.7%	15.9%	16.0%	19.3%	27.2%	31.0%
30-39	11.6%	21.3%	-	9.0%	19.0%	18.0%	20.4%	29.8%	32.5%	26.5%	20.1%
40-49	13.2%	20.0%	11.9%	21.9%	21.8%	30.0%	33.5%	24.9%	19.0%	13.4%	9.9%
50-59	13.3%	15.0%	21.4%	28.3%	34.1%	28.0%	18.0%	11.7%	8.7%	6.7%	5.4%
60-69	9.6%	9.4%	27.1%	27.7%	16.3%	11.6%	7.1%	5.9%	3.6%	3.0%	3.9%
70+	9.2%	8.3%	39.5%	13.1%	8.8%	6.7%	5.0%	4.5%	2.5%	2.2%	2.9%

Note: The source of data are the Census of 1990, the Census of 2000, and the American Community Survey of 2010. The fraction is over the cohort size.

C. Sampling Description

In order to make a meaningful inference, some tables in the previous section use different samples. This section briefly summarizes the sample selection criteria. The Appendix provides a detailed description of the sample.

The analysis excludes people who are less than 25 years of age for two reasons. First, many young people are graduates. Their wage increases are more related to graduation, less assimilation. The inclusion of young people would also create technical issues for implementing a multivariate logit model used in the later section since there are no people under 25 in the immigration cohort of people arriving before 1970 in the 2000 and 2010 surveys.

IV. Main Analysis

A. Conventional Regression Model Using Cross-Sectional Data

The conventional model used to estimate the immigrant assimilation is an extension of the classic wage function. The classic wage function, known as the Mincer (1958) earnings function, explains the wage as a function of experience and education. Mincer's model is prevalent in many labor economics studies with a broader set of covariates, such as gender, race, marital status, and region. This paper includes all the covariates above.

The base model using cross-sectional data can be written as

$$\ln(wage_{ij}) = \alpha + \beta \cdot ysm_{ij} + \gamma_1 \cdot age_{ij} + \gamma_2 \cdot age_{ij}^2 + \boldsymbol{\theta} \cdot \boldsymbol{X_{ij}} + \varepsilon_{ij} \quad (1)$$

where $wage_{ij}$ is the annual earnings of individual i in migration cohort j; ysm stands for years since migration; age_{ij} is age, a common proxy for experience when a direct measure is not available and education attainment is controlled. $\boldsymbol{X_{ij}}$ are control variables, like educational attainment, gender, and other factors; ε_{ij} is the error term with mean zero.

Coefficient β measures the average growth in earnings for staying one more year in the U.S. This model exploits the variance in the length of stay after the initial migration of different individuals while controlling for all earnings determinants. The idea is to compare immigrants' earnings residuals in addition to the prediction using the earnings function. The difference shows whether staying in the host country for more time contributes to earnings growth.

This model can be estimated with both cross-sectional data and panel data. Many articles using panel datasets, such as Kim (2006) and Abramitzky, Boustan, and Eriksson (2014), also include a regression model using cross-sectional data for comparison. This paper begins with an OLS regression using cross-sectional data to get the replicate result.

Critics argue that the OLS regression using cross-sectional data is subject to biases since the model compares immigrants in different cohorts. The sign of the bias is determined by whether the "cohort quality" increases or decreases over time. As per our previous discussion, recent immigration cohorts may have lower quality due to the inflow of a large share of low-skilled Hispanic immigrants. As a result, the average earnings of immigrants who arrived recently are expected to be lower than that of immigrants who arrived earlier. The cohort bias is positive.

With repeated cross-sectional data, the cohort bias can be corrected with the following model,

$$\ln\left(wage_{ijt}\right) = \alpha + \mu_j + \beta \cdot ysm_{ijt} + \gamma_1 \cdot age_{ijt} + \gamma_2 \cdot age_{ijt}^2 + year_t + \boldsymbol{\theta} \cdot \boldsymbol{X_{ijt}} + \varepsilon_{ijt} \quad (2)$$

The time subscript indicates the time when this data are collected. The year fixed effect, $year_t$, captures market conditions at time t. μ_j is a cohort indicator, which categorizes immigrants based on their arrival periods. In addition to Mincer's specification, I allow a quadratic age function. In practice, there is little difference except age can explain more earnings growth.

After adding the cohort fixed effect, μ_j, the estimated coefficient $\hat{\beta}$ uses the within-cohort variation to identify the assimilation rate. Earnings differences due to length of stay in the U.S. are compared between people in the same immigration cohort. As a result, the cohort bias is corrected, changing the assimilation estimate. Given our assumption that cohort quality is declining over time, the estimated $\hat{\beta}$ should be smaller.

We can run a simple test to verify if the estimate's decrease is due to bias correction. To determine if the bias is caused by decreasing cohort quality, we can run two regressions without other control variables and two regressions with control variables. Each pair includes one regression with the cohort fixed effect and one without. When there is no control variable, the assimilation rate estimates include changes in education, marital status, region, race, and other factors. After adding these control variables, the bias should be smaller as the comparison is limited in smaller cells.

The interpretation is that the cohort fixed effect controls for differences in cohort quality. The difference between two estimated $\hat{\beta}$ should decrease if more control variables are included, assuming the bias is due to measured cohort quality changes. For example, if the average educational attainment decreases from the older cohort to the newer cohort, what appears as assimilation in the cross-sectional data partially reflects the difference in educational attainment. When including education as an explanatory variable, the model identifies assimilation within the same education group. The bias should be smaller.

In actual data, cohort quality is probably a mix of both smooth trends and sudden shocks. Controlling cohort fixed effects may not be very useful since cohort groups are mostly arbitrarily chosen without identifying when a shock occurs in cohort quality.

Both estimates in equation (1) and (2) are subject to outmigration bias on unmeasured factors, which cannot be resolved with the conventional regression model. When leaving immigrants are predominately selected, for example, low skilled, the rest's average earnings become higher without any assimilation. Research using panel data that shows no assimilation effect concludes the estimates of assimilation using cross-sectional data are positive due to cohort bias and outmigration bias. The question regarding whether outmigration bias is large enough to explain assimilation estimates is tested in Section V.

B. Estimating Assimilation with Panel Data

A longitudinal panel dataset has repeated observations of the same person over time. This property is ideal for correcting the cohort bias since all the people are compared with themselves over time. The wage growth path is specific. Migrants who left are self-identified as missing in subsequent surveys. The room for outmigration bias is small. As a result, estimating assimilation with a longitudinal panel dataset is compelling in assimilation research.

The problem hides in the process of constructing the longitudinal panel data with cross-sectional data. To form longitudinal datasets as Abramitzky, Boustan, and

Erikson (2014) and others do, the most recent immigrants must be dropped from the data because they are not present in the prior survey. For example, Abramitzky, Boustan, and Erikson (2014) match observations across two censuses, which means all observations pertain to those who stayed in the U.S. for at least ten years. Only matched data merge with their income data by matching their Social Security numbers with a longitudinal income dataset, based on Social Security Administration Data and the SIPP. The final dataset includes the characteristics of all individuals in a particular year and their income history.

Although I focus on a different period than Abramitzky, Boustan, and Erikson (2014), their paper identifies relevant issues with the panel data construction process. A constructed panel dataset is generally small due to the low match rate and limitations on income history data. Whether the sample from the panel dataset represents all immigrants' assimilation process depends on the selection process. If the selection process is not random, the authors need to clarify who are excluded and how they are different. The general solution is to test if the matching process leads to a significant change in the distribution of selected characteristics. However, there may be unobservables that may lead to bias in estimation.

Even when return migration is not selective—i.e., when the characteristics of those immigrants who return are not different from those who stay—estimates of assimilation may be biased. The problem relates to the length of stay for those in the matched sample. Only individuals who stay in the U.S. for at least one complete survey period are valid in the matched sample. Suppose the two cross-sectional surveys used for matching separate by ten years, immigrants who were in the U.S. for less than ten years, are automatically excluded from the panel dataset[5]. This selection bias could be severe for two reasons. First, immigrants with abundant skills to survive the U.S. society upon arrival may be over-represented in the longitudinal dataset. These people may not be very different from native workers at the beginning. Second, short-term immigrants may be more sensitive to assimilation effects. Interpersonal and language skills may accumulate faster in the early years after arrival. Dropping short-term immigrants may reduce estimated assimilation.

Chiswick (1978) concludes that the assimilation process ends after 10 to 15 years. I use repeated cross-sectional data to replicate the conventional model's findings with data from 1990 to 2010. The findings call into question the size of the cohort and outmigration biases. Then, I use a method that adjusts for migrant characteristics to learn more about the assimilation processes and cohort bias. It turns out that assimilation is fast in the first ten years and slows down in later years. When we add a quadratic term of the years since migration variable to the existing models, the results are consistent with this explanation.

Lubotsky (2011) discusses both disadvantages mentioned above and three other challenges when matching cross-sectional data to construct a longitudinal dataset. These challenges include: some matched data cannot be matched with any earnings record; some observations are moving back and forth across borders; there may be measurement error in earnings data due to the top coding of data. As a result, the comparison between estimates using cross-sectional data and estimates using

[5] The most exreme case will be 19 years as one came right after a decennial survey.

longitudinal data is less informative. It is hard to conclude that cross-sectional estimates are biased when the validity of the longitudinal estimates is in question.

C. Controlling for Observed Characteristics

The non-parametric approach is inspired by Chiquiar and Hanson (2005). This approach reweights the observed distribution of immigrants' characteristics, such as gender, education, and age, to a native-born benchmark distribution and constructs a counterfactual earnings distribution based on this benchmark. The counterfactual earnings distribution is then compared with the native-born earnings distribution. Since differences in characteristics are accounted for, the comparison can be taken to indicate how employers treat immigrants, based on length of stay in the U.S.

The non-parametric approach relaxes the linearity specification of the assimilation. The estimation shows the shift of the earnings distribution, not the expectation. As a result, we can learn how does assimilation happen? What matters? And what adjustment is necessary to correctly specify the

The earnings distribution of immigrant cohort j is decomposed as,

$$g(w|C=j,D=1)=\int f^{j}(w|x)h(x|C=j,D=1)dx \quad j=0,1,\dots,J \text{(3)}$$

where w is the earnings. $f^{j}(w|x)$ is the earnings assignment function given characteristics x of cohort j; $h(\cdot)$ is the distribution of characteristics; and C stands for the cohort. There are $J+1$ cohorts, one native-born cohort, cohort 0, and J immigrant cohorts, based on when the immigrant arrived in the U.S. D is an indicator of being in the labor force.

The counterfactual earnings distribution of cohort j based on the native-born distribution of characteristics $h(x|C=0, D=1)$ from cohort 0 is written as,

$$g_0^j(w|D=1) = \int f^j(w|x)h(x|C=0, D=1)dx j = 1, \dots, J \quad (4)$$

This counterfactual earnings distribution can be tied to the observed distribution in (3) as,

$$g_0^j(w|D=1) = \int f^j(w|x)h(x|C=0, D=1) \times \frac{h(x|C=j, D=1)}{h(x|C=j, D=1)} dx$$

$$= \int \theta^j f^j(w|x)\, h(x|C_i=j, D_i=1)dx \quad (5)$$

$$\theta^j(x) = \frac{h(x|C_i=0, D_i=1)}{h(x|C_i=j, D_i=1)} \quad (6)$$

$\theta^j(x)$ is the weight adjustment ratio reflecting the difference between the distributions of the natives and an immigration cohort on observed characteristics. As shown in equation (5), the counterfactual earnings distribution can be derived from the earnings distribution of an immigrant cohort by adjusting the distribution to correspond to the distribution of characteristics for natives.

Using Bayes' theorem and the property of conditional probability,

$$h(x) = \frac{h(x|C_i=j, D_i=1) \cdot \Pr(C_i=j, D_i=1)}{\Pr(C_i=j, D_i=1|x)}$$

$$= \frac{h(x|C_i=j, D_i=1) \cdot \Pr(C_i=j, D_i=1)}{\Pr(D_i=1|C_i=j, x) \cdot \Pr(C_i=j|x)} \quad (7)$$

Using equation (7), $\theta^j(x)$ can be written as,

$$\theta^j(x) = \frac{\Pr(D_i=1|C_i=0, x)}{\Pr(D_i=1|C_i=j, x)} \cdot \frac{\Pr(C_i=0|x)}{\Pr(C_i=j|x)} \cdot \frac{\Pr(C_i=j, D_i=1)}{\Pr(C_i=0, D_i=1)} \quad (8)$$

The adjusting factor, $\theta^j(x)$, has three components. The last component is the joint probability that one is in cohort j and labor force regardless of skill. It is constant since it is irrelevant to x. We can ignore this component since the total weight after the reweighting process will be adjusted to ensure the total represented population does not change. The first term captures the labor force participation rate of cohort j given observed characteristics. For example, if native workers with higher skills tend to leave their highly educated partners at home to raise kids, and immigrant couples tend to maximize the total income regardless of skill level, high skilled immigrants would receive less weight using this ratio. However, such a shift in earnings distribution is not a function of the market. As a result, it is excluded so that the adjustment does not consider the difference in labor force participation across different cohort-skill cells.

The middle term is then used to adjust for the characteristic difference between cohort j and natives. As a result, the adjusting factor is,

$$\theta^j(x) = \frac{\Pr(C_i = 0|x)}{\Pr(C_i = j|x)} \tag{9}$$

For multiple-year data, the earnings distribution of immigrant cohorts in year t is reweighted with a factor for that year,

$$\theta_t^j = \frac{\Pr(C_i = 0|x, t)}{\Pr(C_i = j|x, t)} \quad (10)$$

where I suppress the dependence of θ_t^j on x. The earnings distribution of immigrant cohorts is made comparable to the earnings distribution of natives of that year, using the distribution of characteristics for natives in that year as the benchmark distribution. Then, earnings distributions are reweighted again by the following factor so that all earnings distributions of immigrant cohorts are comparable to the earnings distribution of natives in the base year,

$$\theta_t^T = \frac{\Pr(C_i = 0|x, t = base\ year)}{\Pr(C_i = 0|x, t)} \quad (11)$$

As a result, the weight factor for repeated cross-sectional data is,

$$\theta_t^{jT} \propto \theta_t^j \theta_t^T. \quad (12)$$

The advantage of this approach is that it adjusts different cohorts' characteristics at different times to the same benchmark. The earnings assignment to all immigrant cohorts is based on the same observed variables, so that neither cohort differences nor outmigration can affect the counterfactual earnings distribution[6] if all relevant variables are observed. For example, there is evidence of dropping cohort quality and negative selection in outmigration. As shown in Section III, the shift in race composition indicates that cohort quality may be dropping and that the average educational attainment of a cohort grows as immigrants die or leave the U.S. However, after adjusting the race and education distributions, the counterfactual earnings distributions of immigrant cohorts are not affected by these two variables' changes. In the real world, $\theta_t^{jT}(x)$ cannot be estimated in practice as unmeasured variables affect wages. We can only arbitrarily choose a set of variables to get $\widehat{\theta_t^{jT}}$. The earnings

[6] For example, Hu (2000) finds that the educational attainment improves in each immigrant cohort across 1970, 1980, and 1990 Censuses, indicating negatively self-selected outmigration on educational attainment.

differences between immigrant cohorts and natives are due to different immigration cohorts' different earnings assignment functions.

The constructed counterfactual earnings distribution,

$$g_{0t}^{j}(w|D_i = 1) = \int \theta_t^{jT} f^j(w|x)\, h(x|C_i = j, D_i = 1, t)dx$$

represents how employers value immigrants from the same characteristic distribution as natives, holding labor force participation unchanged. The difference in earnings distribution,

$$g_{0t}^{j}(w|D_i = 1) - g(w|C_i = 0, D_i = 1, t = base\ year) \quad (13)$$

shows the difference in the density for each earnings level w for natives and immigrants with the same distribution of measured characteristics. When immigrants and natives with the same measured characteristics are perfect substitutes from the employers' points of view, the earnings assignment function for immigrant cohort j, $f^j(w|x)$, is identical to the earnings assignment function of natives. The earnings distribution of natives and the adjusted earnings distribution of the immigration cohort should be identical. Assimilation is defined as a convergence of the earnings assignment function of immigrants to that of natives. People who alter their characteristics in the U.S., for example, obtaining an advanced education degree, cannot contribute to estimated assimilation based on this measure.

Unlike Abramitzky, Boustan, and Erikson (2014), who use occupational improvement to measure the assimilation process[7], this paper focuses on earnings.

[7] They use occupation scores to proxy one's earnings. In their analysis, earnings improvement only occurs when an immigrant moves into an occupation with higher average earnings.

Whether or not they are associated with occupational changes is irrelevant. The improvements in communication skills, cultural knowledge, and other skills may or may not lead to occupation changes. For example, a math professor from Russia may receive higher earnings when his English writing and speaking skills improve to describe his ideas more clearly to editors and his students, even if there is no change to his occupational designation.

This non-parametric method is more flexible than OLS regression. The assimilation process is not assumed to be linear. I can also specify any characteristics distribution as the base distribution. A corresponding wage distribution would be created. I choose the characteristics distribution of natives in accord with assimilation theory. This method asks employers what earnings they would offer to a native group and an immigrant group with an identical observed characteristics distribution?

V. Results

A. Conventional Estimates of Assimilation

This section presents four regression estimates of assimilation. The results call into question the role of cohort bias and the outmigration bias in the assimilation estimation. Table 5 shows the results of regression based on equations (2) using repeated cross-sectional data.

Table 5

Regression Results

	(A)	(B)	(C)	(D)
Year since migration	0.013***	0.006*	0.012***	0.002
	(0.001)	(0.003)	(0.002)	(0.002)
Cohort		✓		✓
year & age	✓	✓	✓	✓
other control			✓	✓

Data from the 1990 and 2000 Censuses, and the 2010 ACS. The standard error is robust and clustered by cohorts. Cohort indicator: immigrants arrived before 1970 or, 1970-74, 1975-79, 1980-84, 1985-1990, 1991-95, 1996-2000, 2001-05, 2006-10. Age is controlled as a quadratic function. Other control variables include gender, race, marital status, educational attainment, and region. Educational attainment is classified as less than high school, high school or equivalent, some college, bachelor's degree, and advanced degrees. The region is classified as New England Division, Middle Atlantic Division, East North Division, West North Central Division, South Atlantic Division, East South Central Division, West South Central Division, Mountain Division, and Pacific Division. Individual weights based on the survey design are used for all regression. *** p<0.01, ** p<0.05, * p<0.1

Column A shows the simple regression controlling only years since migration, calendar year fixed effects, and a quadratic age function. The estimated assimilation coefficient is 1.3 percent, implying that, for two otherwise identical immigrants, the one who stayed one more year has 1.3 percent higher earnings in addition to the earnings growth predicted by age and year. Column B adds the cohort fixed effect. With cohort controlled, the assimilation estimate exploits the within-cohort variance only. The estimated coefficient drops to 0.6 percent. This result is similar to findings of Borjas (1985) and Abramitzky, Boustan, and Eriksson (2014). Their interpretation is that the unaccounted differences between cohorts account for half of the estimated assimilation in Column A. For example, since the regression does not control the race, the assimilation estimate is positively biased. An immigrant who came earlier is more likely to be non-Hispanic and has higher earnings, creating a positive bias.

Column C shows the regression without the cohort fixed effect but with more control variables. The extended set of control variables include a quadratic age function, year fixed effects, gender, race, educational attainment, marital status, and residence region. This set of variables includes all determinants that are generally used in studies of earnings. The estimated assimilation is 1.2 percent per year, similar to the previous estimate. The assimilation appears to have a relatively independent effect on earnings growth.

Column D shows the regression with both the cohort fixed effects and the extended set of controls. The estimated assimilation is statistically insignificant. This result calls into question the conclusion that the assimilation estimation is biased by dropping cohort qualities. Since the difference between Column A and B estimates is smaller than the difference between Column C and D estimates, the "bias" appears larger with more covariates. In contrast, the difference between cohorts should get smaller with more controls (see discussion in section IV.i).

Instead, a misspecification problem is plausible. The cohort fixed effect is similar to a step function of years since migration. The coefficient of years since migration captures the linear assimilation effect, while the cohort fixed effect may capture the nonlinear part of assimilation.

The patterns observed in Table 1 are also not consistent with the assertions of cohort bias. Most studies (Abramitzky, Boustan, & Eriksson, 2014; Borjas G. , 1985; Kim, 2009) using panel data conclude that there is no evidence of assimilation. Assuming the "true" assimilation effect is zero, I can infer the extent to which the estimated positive assimilation effects after controlling for cohort fixed effects are due to outmigration bias. Then, I can calculate the outmigration's size that would lead to the bias and compared it with the observed size of outmigration.

Table 1 shows that about 27.88 million immigrants live in the U.S. in 2000 who arrived before 1996 (we exclude immigrants who arrived between 1996 and 2000 due to incomplete count). The average annual earnings are around $27,580 per immigrant worker at that time. Assuming there is a constant proportion of people in the labor force and only people making zero dollars left, the smallest number of outmigrants needed to create a bias of 0.6 percent per year in the assimilation estimation using the data in 2000 and 2010. I also assume no real earnings improvement for either stayers or leavers to simplify the calculation. The labor force participation rate of immigrants is 67 percent in 2000. As a result, there are about 18.68 million immigrant workers, making 515 billion dollars per year. To make the average income increase by 6 percent solely by filtering out the lowest earners (0.6 percent per year for ten years), the number of immigrants in the labor force would decline to 17.62 million. The population size should be roughly 26.3 million (17.62 million / 67%) by 2010, implying 1.58 million outmigrants. The data show that outmigration is about 1.66 million immigrants, similar to our prediction. Notice the estimation is based on assuming all outmigrants make zero dollars per year, which is a very conservative assumption. Suppose the average earnings level of these short-term immigrants is assumed to be two-thirds of the average earnings. In that case, there must be 2.87 million emigrants to bias the assimilation estimation by 6 percent[8]. The argument of outmigration creates bias in the assimilation estimation is dubious.

B. The Nonparametric Method Controlling for Observed Characteristics

To explain this inconsistency, we need to know more about the assimilation process. This section proposes a function to unify the multivariate distribution of immigrants' characteristics across time and cohorts. Thus, I can compare the earnings distribution of different groups to learn about the assimilation process.

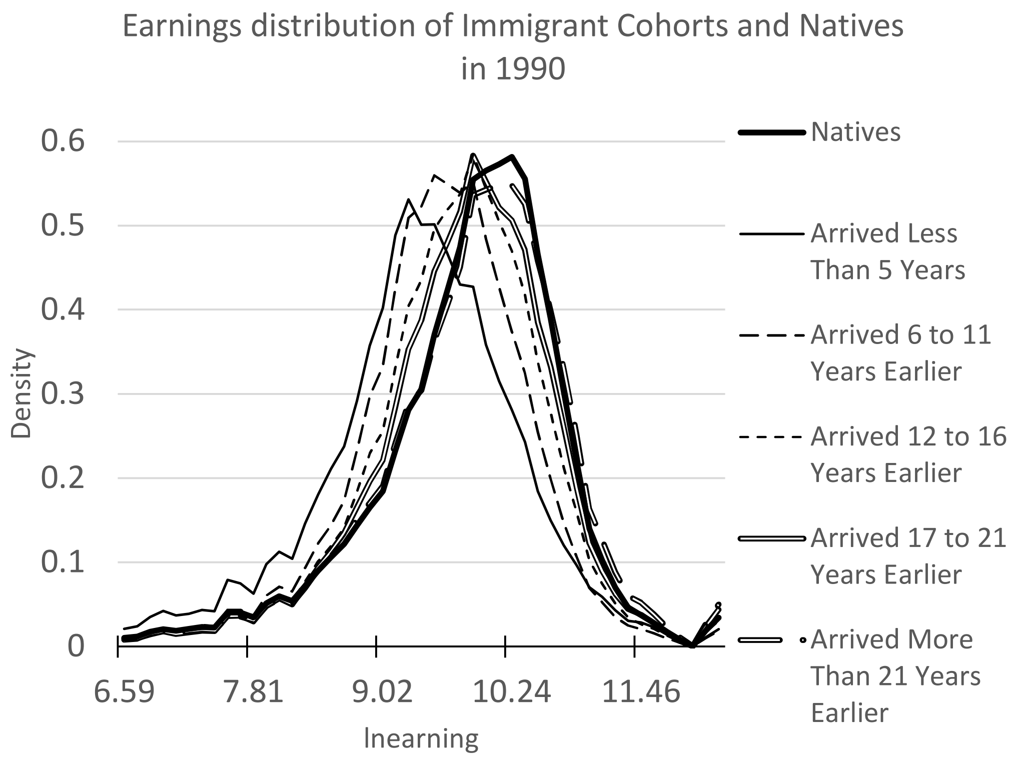
Figure 2 depicts the earnings distribution of the natives and all immigrant cohorts using the 1990 census data. The earnings distribution of older immigrant cohorts is to the right of the earnings distribution of newer immigrant cohorts, indicating a shift to higher earnings for more extended stays. The right shift is stable. Notice that the earnings distribution of workers who have stayed for 16 to 20 years is still on the left of the native earnings distribution. Immigrants who have stayed for more than 20 years have a similar earnings distribution to natives but with slightly more significant variation.

Figure 3 shows the differences in the distribution density between each immigrant cohort's earnings and the native earnings distribution. The thin black line stands for new immigrants who had arrived within the previous five years. When the line intercepts with the horizontal axis, the fraction of natives and immigrants at that earnings level are identical. When the line is above the horizontal axis, there are relatively more immigrants at this earnings level and vice versa. Immigrants are over-

[8] Since the most recent immigrant cohort is dropped due to the Census issue, which would have the highest outmigration rate, I predicted the outmigration separately, and the calculation shows a similar conclusion as the example here.

represented in the lower end of the distribution (below $22,026 per year[9]). Earlier cohorts have fewer low-income workers. The earliest cohort has a nearly identical distribution to that of the natives.

Figure 1



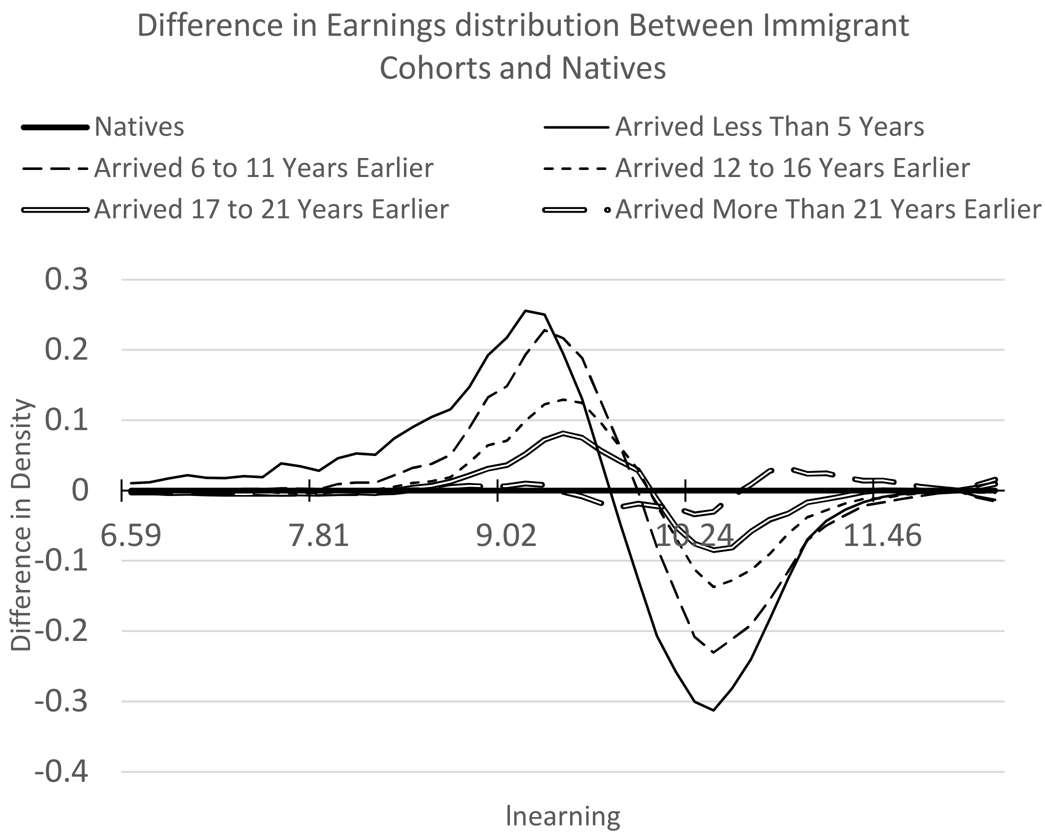
Note: The source of data is the U.S. census of 1990 from IPUMS. The width of kernel density is 0.1.

The distribution differences are partially due to cohort quality differences. As described in the data section, immigrant characteristics vary by cohorts. The most recent immigration wave is largely driven by migration from developing countries from Latin America and Asia. Immigrants who arrived before 1970 were mostly from Europe. Immigrants from different source countries may differ in terms of education, race, work experience, and other factors summarized as cohort quality. However, as

[9] All dollar values are in 1999 dollars.

noted in the data description section, decreasing cohort quality is not apparent. The observed differences in characteristics for immigrant cohorts are not large, especially for cohorts after 1970.

Figure 2

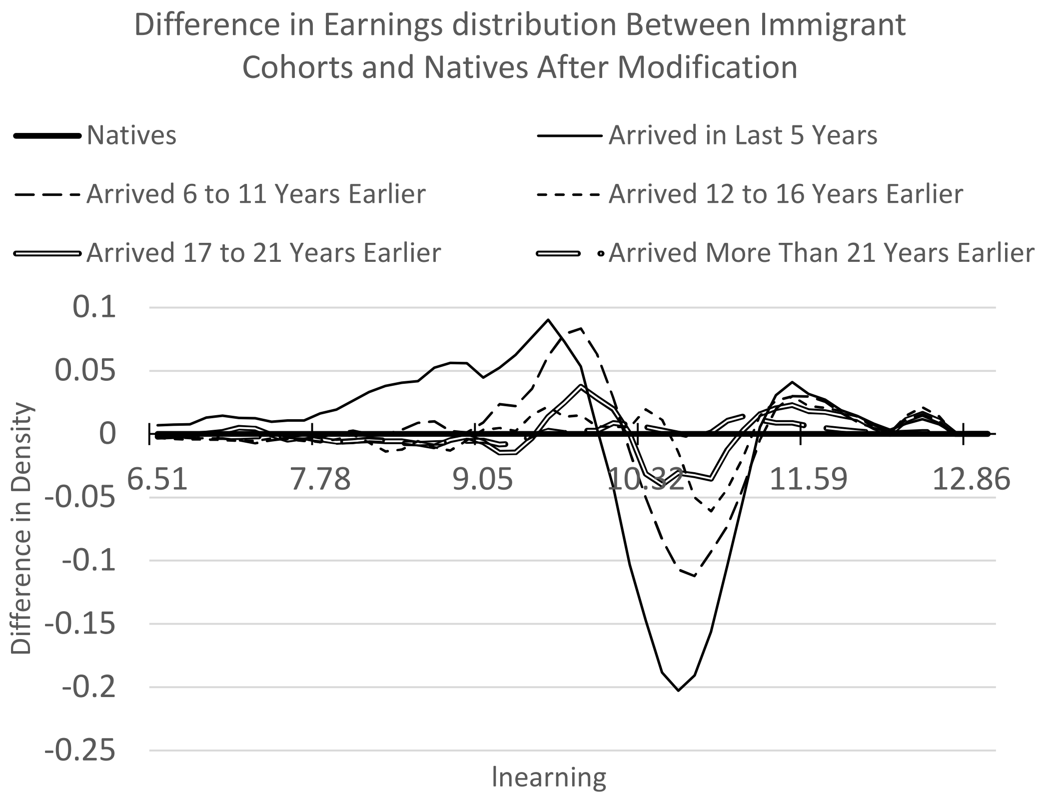


Note: The source of data is the U.S. census of 1990 from IPUMS.

Figure 4 depicts the earnings distribution after unifying the underlying characteristics distribution. The sample weights of all immigrants are adjusted with θ^j_{1990}, as in formula (9). The earnings distribution differences between immigration cohorts and the natives become smaller compare to Figure 3. Controlling characteristics reduces both the cohort bias and the outmigration bias. Notice the reduction varies by cohort. New immigrants, who have lived five years or less in the

U.S., are still over-represented in the low-income range. Immigrants who had arrived 5 to 10 years before the survey still make less money but with far fewer people at the bottom of the distribution. Immigrants who arrived more than ten years earlier have a similar earnings distribution to that of the natives.

Figure 3



Note: The source of data is the U.S. Census of 1990 IPUMS. The individual earnings is modified with θ_{jt}^{T} to conform to the distribution of natives in educational attainment, gender, race, marital status, residence region, and age structure.

Figure 4 indicates how employers assign earnings based on length of stay for individuals in each cohort with the same observed characteristics. There are more migrants than natives for the most recent immigrants making less than $16,000 per year (measured in 1999 dollars). The conventional argument is that many immigrants are young, with low education, and little experience, thus making less money. The results suggest that they receive lower earnings than natives and receive lower earnings than immigrants who arrived earlier, even with the same characteristics. For the cohort that stayed only five more years, far fewer people make meager earnings. That is, employers are willing to offer higher earnings, presumably reflecting improvements in factors not captured in our observed measures, like communication ability and job-related skills.

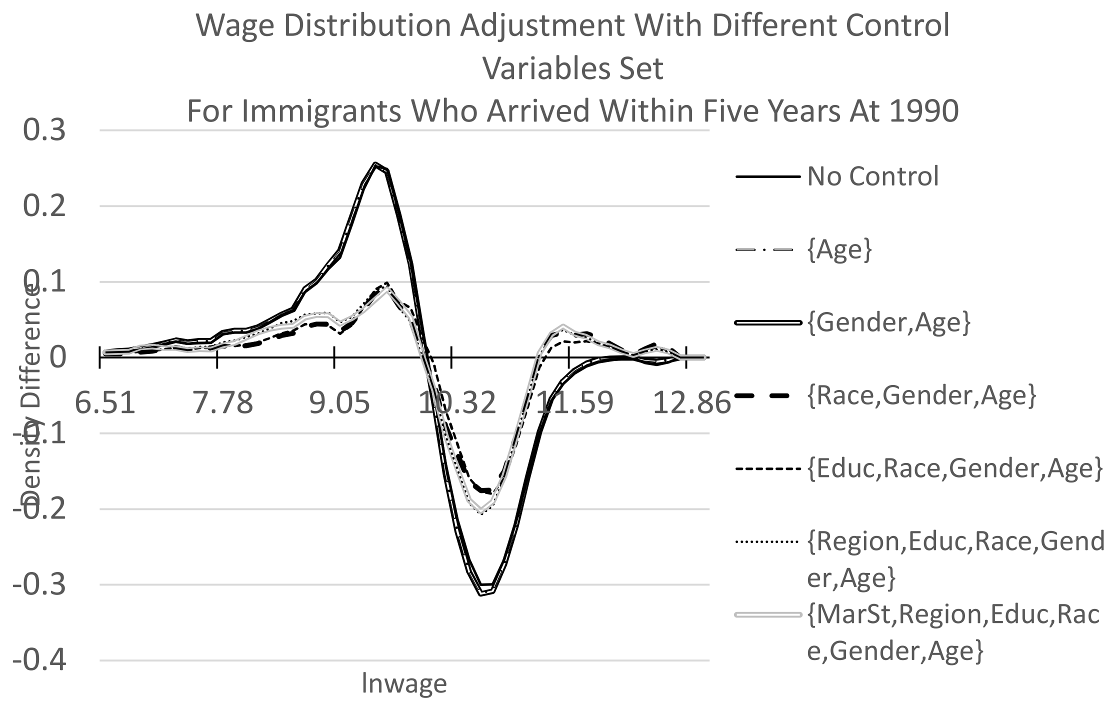
For immigrants who stayed for more than ten years they are under-represented relatives to natives for incomes between $26,600 and $73,900 per year. However, they are over-represented both below and above this range, indicating a larger variance in the earnings distribution than natives. The difference in variance also decreases as immigrants stay longer in the U.S. For the group that migrated 20 years earlier, their earnings distribution shows only small differences than the earnings distribution of natives.

Figure 4 also suggests that the assimilation process is not linear. Although the earnings distribution converges to that of the natives, the longer immigrants stay, the smaller the marginal improvement is. For immigrants who stayed more than 16 years, their earnings distribution is less than 0.05 density points from the natives' earnings distribution across all possible earnings levels.

In Figure 5, I use data on immigrants who arrived within five years before 1990 to explore how adding each control variable affects the earnings distribution. A variable must meet two conditions to impact the earnings distribution. First, there must be an identifiable difference between immigrants and natives in this variable. Second, the variable must affect earnings. That is, the difference in this variable must produce a perceptible difference in earnings. Given the description of data in section III, six variables are chosen, $s_i =$ {Age, Gender, Race, Education Attainment, Region, Marital Status}, since immigrants and natives are statistically different in these variables. The analysis is repeated seven times. In the i^{th} step, the first i variables are included. The unadjusted wage distribution of immigrants is step zero. The adjusted wage distribution of immigrants changes for each set of variables. The result from the ith step is then compared with the result from the previous $i-1$th step to determine how the inclusion of variable i affect the result.

The race is the most essential measure explaining the difference between the original earnings distribution and the adjusted earnings distribution. The race composition varies significantly based on immigration status. As a result, we can confirm that the wage assignment function, as described in equation (4), considers race an essential factor in determining wages. After adjusting for race composition, the difference in earnings distribution between immigrants and natives shrinks perceptibly, with far fewer immigrants are in the lower earnings ranges. Notice, the distribution adjustment does not include country of origin by construction, as the base distribution is natives. To some degree, the race may proxy for the country of origin. Another variable, that makes noticeable but moderate changes, is the composition of residence regions. Other variables in the sequence are relatively unimportant.

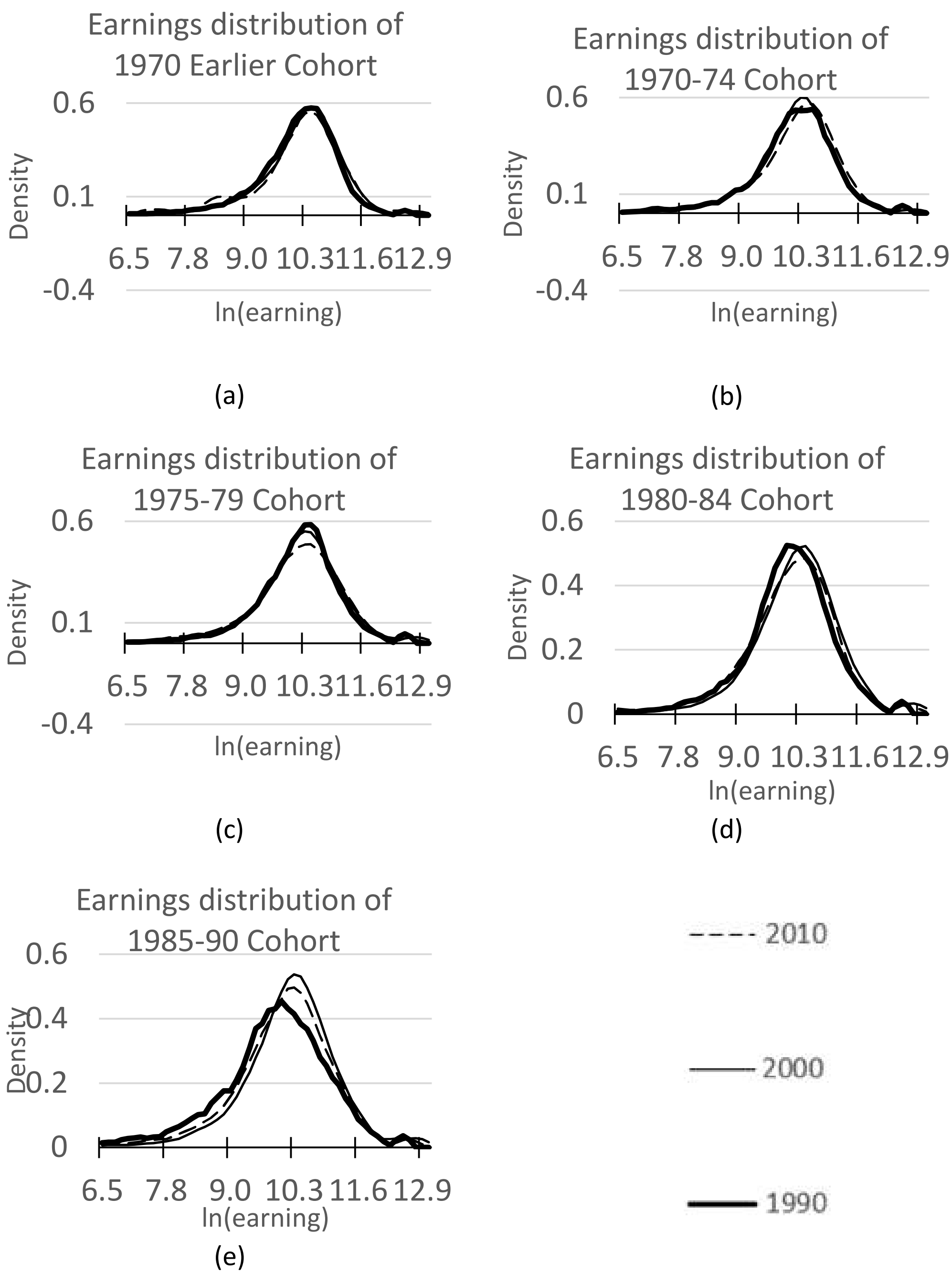
Figure 4 Adding Variables as Control



Note: The source of data is the U.S. census of 1990 from IPUMS.

The changes in the distributions for cohorts over time confirms our interpretation. Figure 6 presents changes in the earnings distribution of each cohort over time. With three censuses, changes in the earnings distribution of a cohort over 20 years can be evaluated. The thick solid line is the earnings distribution measured using the 1990 census. The thin solid line is the 2000 census. The dashed line is the 2010 census. Figure (7a) shows the cohort's earnings distribution arriving before 1970, observed in 1990, 2000, and 2010. For example, when immigrants had been in the country for over 20 years, over 30 years, and over 40 years. The earnings distribution change is small. Figure (7b) to (7d) do not show any obvious change in the earnings distribution. Figure (7e) represents the shift in the earnings distribution in the first 20 years after migration. These graphs suggest that the assimilation effect is only significant for the first ten years in the country.

Figure 5



The conclusion is that the assimilation process slows down after the first ten years of stay in the U.S. Using what learned from the above analysis, we return to equation (2) with a specification that allows for nonlinear assimilation process.

C. Revisiting Conventional Model

To capture the curvature of the assimilation process, I add a quadratic term of years since migration. Estimates for regressions that include such quadratic terms are shown in Table 6, along with linear regression coefficients (D) and (E) from Table 2 for comparison.

Table 6

	Regression Results				
	(A)	(B)	(C)	(D)	(E)
Ysm	0.024***	0.029***	0.024**	0.012***	0.002
	(0.002)	(0.004)	(0.008)		
ysm_sq	-0.00023***	-0.00039***	-0.00034**	-	-
	(0.00004)	(0.00008)	(0.00011)		
Cohort			✓		✓
1970-74			0.007		-0.033
			(0.026)		(0.029)
1975-79			0.005		-0.080*
			(0.043)		(0.038)
1980-84			-0.032		-0.159**
			(0.060)		(0.048)
1985-90			-0.071		-0.271***
			(0.084)		(0.055)
1991-95			-0.039		-0.288***
			(0.098)		(0.052)
1996-2000			-0.098		-0.426***
			(0.125)		(0.060)
2001-05			-0.098		-0.444***
			(0.141)		(0.063)
2006-10			-0.183		-0.617***
			(0.171)		(0.072)
year & age	✓	✓	✓	✓	✓
other control		✓	✓	✓	✓

Data from the 1990 and 2000 Censuses, and 2010 ACS. The standard error is robust and clustered by cohorts. Other controls include gender, race, educational attainment, marital status, and region. *** p<0.01, ** p<0.05, * p<0.1

Column (A) shows the regression without additional control variables. The initial assimilation rate is about 2.4 percent each year. The square term coefficient shows that the speed slows down by 0.045 percentage points each year. Column (B) shows that new immigrants who just arrived improve their earnings by 2.9 percent in the model with controls in the first year. The rate of earnings growth is slowing down by 0.078 percentage points each year as immigrants stay longer. The quadratic function implies that the assimilation effect reaches a maximum around the 37th year after arrival. The full assimilation effect implies a growth of about 42 percent in earnings in 30 to 40 years after arrival. In contrast to the linear specification, the regression with the quadratic term is not sensitive to controlling the cohort's fixed effects.

All estimates of cohort fixed effects become insignificant after adding the quadratic term for years since the migration. The reason is that the linear assimilation process with the cohort fixed effect assumes that the assimilation is an exact linear process across cohorts. The nonlinear assimilation effect is captured by the cohort fixed effects, given that cohorts differ in their distributions of years since migration. Also, because of the squared term, the linear years since migration measure reflects the initial earnings growth, instead of a fitted linear annual earnings growth as in the linear model. The linear years' coefficient since migration in the quadratic model shows that the initial assimilation rate is higher than the average subsequent earnings growth. The quadratic estimates of the assimilation process exhibit the same pattern as suggested in the method that adjusted for the distribution of migrant characteristics. Immigrants assimilate quickly initially, and the pace slows down after that, as shown in Figure 6.

Figure 6

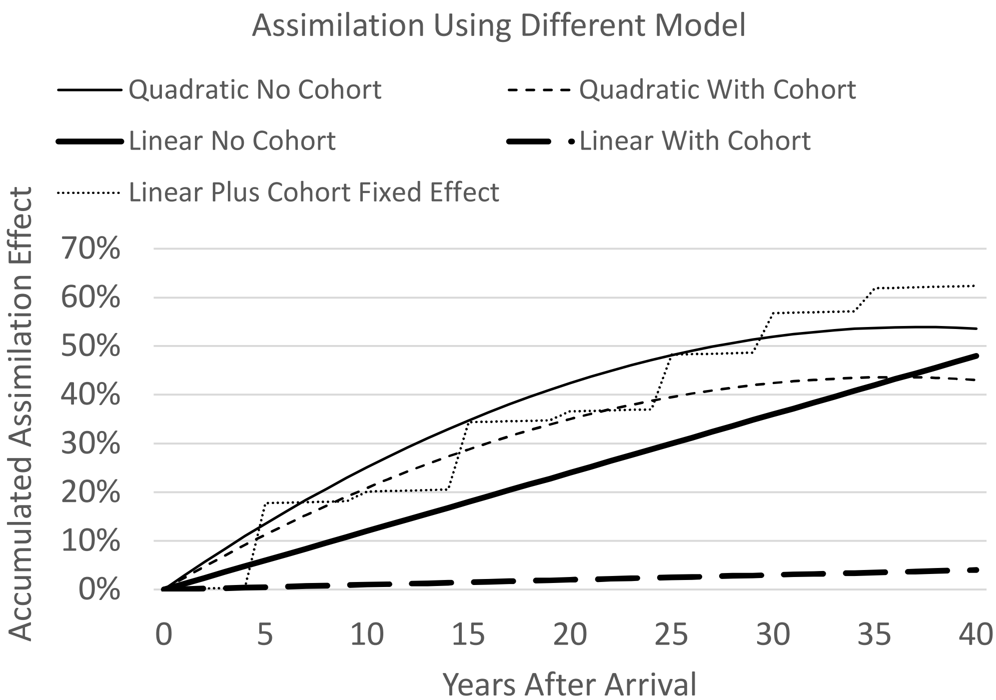


Figure 7 depicts the projected earnings growth after arrival due to assimilation using four different specifications. When using the quadratic term, the assimilation rate is much higher than that estimated based on the simple linear model. Adding cohort fixed effects does not lower the estimation of assimilation by much. The assimilation (thin dashed line) in the first ten years contributes 20 percent to earnings. This contribution is about 15 percent for the next ten years and 7 percent in the following decade. From the third decade to the fourth decade after arrival, assimilation increases earnings by less than 1 percent. The linear model without cohort fixed effects works like a linear estimation of the thin dashed line. The linear assimilation effect estimate drops significantly after adding the cohort fixed effect. The dotted line adds the cohort fixed effect to the assimilation estimates. The earnings growth of the step function is very similar to the nonlinear assimilation estimation.

Why does the speed of assimilation decline over time? Psychologists have argued that human beings have different learning abilities at different ages; a typical example is language learning (Lenneberg, 1967). Cheung, Chudek, and Heine (2011) discovered a sensitive period, less than 14.5 years old, for “acculturation.” The idea of acculturation is the cornerstone of assimilation, focused on learning new skills, knowledge, and apprehension about the host society. These psychologists find differences between acculturation rates for child immigrants and adult immigrants. They wonder if this difference is due to the length of exposure or exposure during the sensitive period, during which immigrants adapt to the new environment faster than they would have if they had arrived at older ages. They confirmed it is the latter reason that makes the difference.

Our sample does not include teenagers. Nevertheless, the idea that assimilation rate is sensitive to migration age can be examined with an interactive term between age and years since migration. The model used for assimilation can be modified as:

$$\ln\left(wage_{ijgt}\right) = \alpha + \beta_0 \cdot ysm_{ijgt} + \beta_1 \cdot ysm_{ijgt}^2 + \delta_0 \cdot age_g \cdot ysm_{ijgt} + \delta_1 \cdot age_g \cdot ysm_{ijgt}^2 + age_g + \mu_j + year_t + \gamma \cdot X_{ijt} + \varepsilon_{ijt} \quad (14)$$

In equation (14), age is grouped into five age groups. δ_0 and δ_1 are estimates of the interactive terms between years since migration and the age groups. These estimates indicate whether the assimilation effect varies by age at arrival. Table 7 shows the result of the estimation of equation (14). The assimilation effect of young immigrants, whose ages are between 25 and 34, are used as the reference group. The assimilation effect is slower for immigrants who arrived in their older years of age. Figure 8 shows the projection of the assimilation effect for the first ten years after arrival based on estimates in Table 7. The two lines at the top represent the assimilation of immigrants who came to the U.S. at ages 35 to 44 and 45 to 54. These two groups have intermediate assimilation effects, where the younger group has slightly more advantage in assimilation. In the meantime, both groups' earnings upon arrival are 15 percent higher than the reference group. As a result, they are the most successful group among all immigrants. In articles regarding emigration, for example, Borjas and Bratsberg (1996), it is common to assume that temporary immigrants migrate at the beginning of their career and stay for a fixed amount of time. The second assumption can be relaxed to include the endogeneity of time in the host country spent to optimize the lifetime utility. In the meantime, the first assumption is rarely critical since many immigrants are, indeed, very young. The finding of this paper indicates another possibility. That is, the optimal time to migrate in the first place may not be right after graduation. Of course, the conclusion cannot be generalized since immigrants arrived in later years are likely to be selected. It is still not clear if training all immigrants before migration will reduce the initial earnings gap and not slowing down the assimilation rate as in this figure.

Immigrants who arrived at age 55 to 64 show no difference in earnings than the reference group. Neither groups are the most preferred worker in the labor market. One group has little experience; the other is near the end of their careers. Surprisingly, immigrants in their late 50s and early 60s can still improve their earnings by almost 25 percent ten years after arrival by successfully assimilate to the host country. Not surprisingly, immigrants in their 20s and 30s are assimilating faster. Immigrants who arrived after age 64 had lower average earnings and showed mild assimilation.

The findings in this section are consistent with the idea of a “sensitive period” of acculturation. In the early years after arrival, when immigrants are generally young, their ability to learn new skills and embrace new culture leads to higher earnings growth rates. The rate of earnings growth is lower for immigrants who arrive at an older age. The findings are also consistent with Borjas’s application of Roy’s migration model (Roy, 1951). Borjas (1987) sheds light on migrants’ selection using Roy’s model, where people with a higher net return to migration leave their home countries. Using Roy’s model, potential immigrants make decisions by comparing the difference between the expectation of future income in their home countries and the host country with the migration cost. Findings in this paper suggest young immigrants can project higher rates of growth in their earnings than old immigrants.

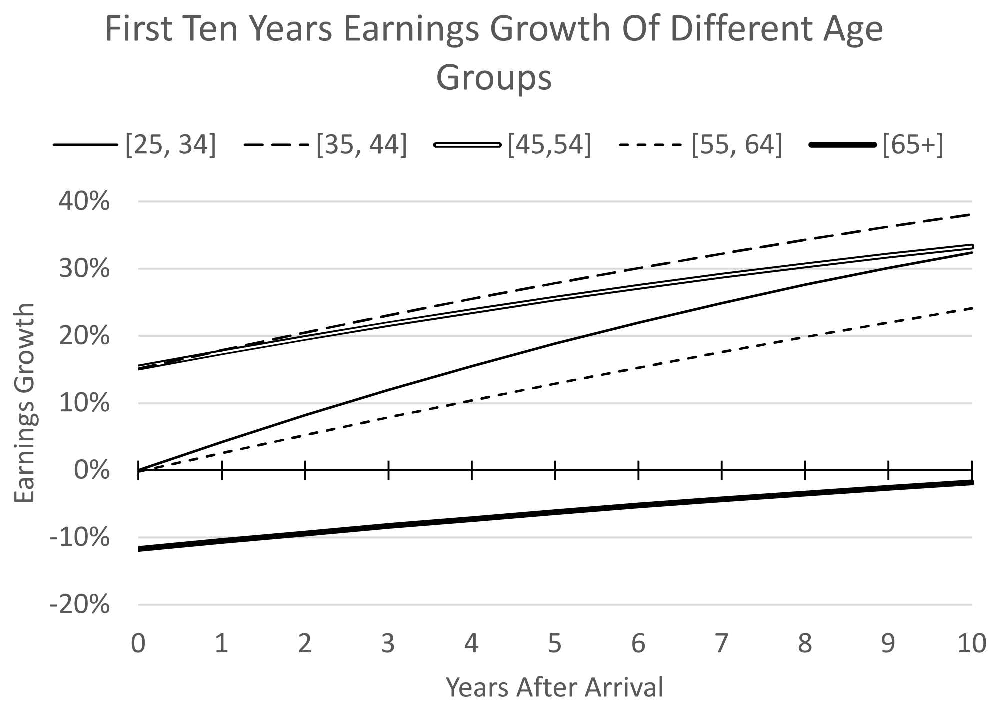
Table 7

Regression with Interactive Term

	Coef.		Std. Err.
Year Since Migration	0.043	***	(0.010)
Ysm Square	-0.00106	***	(0.000)
[35,44]	0.151	***	(0.022)
[45,54]	0.153	***	(0.037)
[55,64]	-0.002		(0.032)
[65+]	-0.117	*	(0.054)
Age Interactive Term			
[35,44]			
Year Since Migration	-0.015	***	(0.005)
Ysm Square	0.00056	***	(0.000)
[45,54]			
Year Since Migration	-0.020	***	(0.006)
Ysm Square	0.00056	**	(0.000)
[55,64]			
Year Since Migration	-0.015	**	(0.007)
Ysm Square	0.00069	***	(0.000)
[65+]			
Year Since Migration	-0.031	***	(0.006)
Ysm Square	0.00085	***	(0.000)
Control Variables	✓		
Number of Observations	1,310,535		
R Square	0.2528		

Note: Data from 1990 and 2000 Census and 2010 ACS. The standard error is robust and clustered in cohorts. Control variables include gender, race, educational attainment, marital status, and region. *** p<0.01, ** p<0.05, * p<0.1

Figure 7



Many studies (Feliciano, 2005; Kim, 2009; Abramitzky, Boustan, & Eriksson, 2014) find that the assimilation processes vary by country of origin. The findings from the modified distribution method also confirm the importance of racial composition[10].

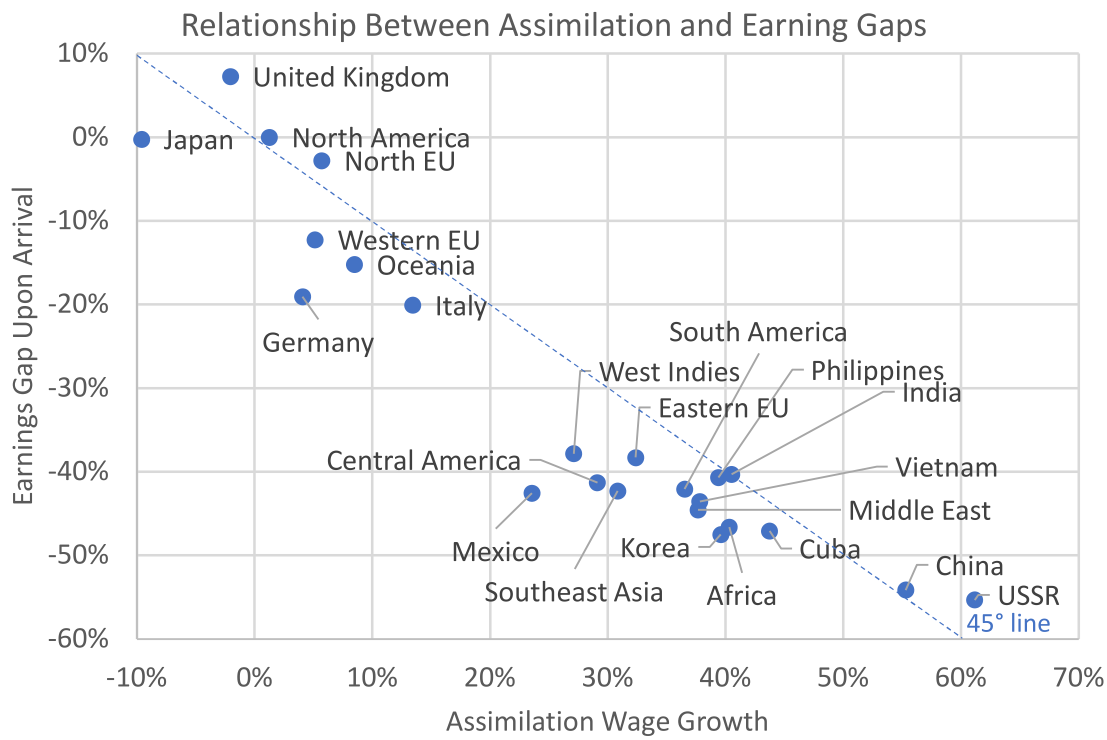
The regression results are listed in Table 8. The regression model is similar to equation (10), except we use immigrants' birthplace as the interactive term with assimilation terms. Notice that countries with relatively fewer immigrants are grouped with other countries within the same geographical area. For instance, North America, the reference group, immigrants from Canada with few people from St. Pierre and Miquelon.

[10] The race composition is used as a proxy for the country of original composition in the previous analysis.

Canadians are used as the reference group since they are similar to Americans both culturally and economically. Unsurprisingly, there is no evidence that Canadians are significantly different from Americans nor that they assimilated.

The first column in Table 8 shows the ethnic fixed effect compared to immigrants from Canada. No immigrants have higher average earnings as Canadians except immigrants from the UK. The ethnic fixed effect is transformed into a percentage difference and summarized in Figure 10. In the figure, these countries are classified into three groups by the average wage gaps. North America, Japan, and most European countries have less than 20% to no earnings disadvantage. People from China and former USSR territories have significant earnings disadvantages of more than 55 percent. The middle group includes Mexico, the Philippines, and Africa countries, with earnings disadvantages between 35 percent to 50 percent.

Figure 8



Note: Data are based on the regression in Table 8. The earning gaps use immigrants from North America as the baseline. The assimilation rate is calculated as the 10-year projected earning growth.

The second and third columns in Table 8 show different patterns of assimilation. Most estimates are statistically significant. The most common pattern is increasing earnings with decreasing speed as immigrants stay longer in the U.S. Six exceptions include Germany, North Europe, Oceania, and Western Europe that show no assimilation and Japan and UK that show negative assimilation. Among those countries with similar assimilation patterns, the initial assimilation rate and their assimilation curve's curvature are different from each other. Figure 10 also indicates a negative correlation between the assimilation rate for the first ten years of stay and the earnings gaps, as most countries are close to the 45-degree line. Figure 10 also indicates a negative correlation between the assimilation rate for the first ten years of stay and the earnings gaps, as most countries are close to the 45-degree line.

Among all immigrant groups, immigrants from Mexico have the lowest net earnings growths within the first ten years of stay. Mexican immigrants face 43 percent point initial earning gaps, with only 23 percent point assimilation in the first ten years. A report (Noe-Bustamante, Flores, & Shah, 2019) of the Pew Research Center states that Hispanic immigrants comprised the largest share of all immigrants in the U.S. in 2017. Most of them reside in the southern border states. The slower assimilation rate of Hispanics may be due to large Hispanic communities with many internal economic activities.

Among immigrants from developing countries, Indians, Chinese, and immigrants from former USSR countries can gain positive net earnings growth in the first ten years of arrival in the U.S. For different ethnic groups, the reason for success is probably different. For Indian, one possible reason is that many Indians speak English even though no language is granted the national language status in India. Chinese immigrants have considerably higher levels of educational attainment than all immigrants from developing countries. However, their lack of language skills might disguise their productivity in the first place.

Table 8

Assimilation Regression by Country

Reference					**Ysm**			**Ysm_Sq**		
					Coef.		**SE**	**Coef.**		**SE**
North America					0.002		(0.006)	-0.00006		(0.00007)
	Ethnic				**Ysm**			**Ysm_Sq**		
Country	**Coef.**		**SE**		**Coef.**		**SE**	**Coef.**		**SE**
Africa	-0.628	***	(0.059)		0.038	***	(0.005)	-0.00054	***	(0.00011)
Central America	-0.533	***	(0.033)		0.027	***	(0.003)	-0.00030	***	(0.00006)
China	-0.780	***	(0.050)		0.050	***	(0.007)	-0.00068	***	(0.00014)
Cuba	-0.636	***	(0.092)		0.039	***	(0.007)	-0.00044	***	(0.00011)
Eastern EU	-0.483	***	(0.086)		0.031	***	(0.006)	-0.00039	***	(0.00010)
Germany	-0.211	***	(0.018)		0.002		(0.003)	0.00009		(0.00006)
India	-0.516	***	(0.045)		0.038	***	(0.007)	-0.00051	**	(0.00017)
Italy	-0.224	***	(0.058)		0.013	***	(0.003)	-0.00012	**	(0.00004)
Japan	-0.003		(0.075)		-0.015	**	(0.006)	0.00039	**	(0.00012)
Korea	-0.645	***	(0.049)		0.036	***	(0.004)	-0.00044	***	(0.00009)
Mexico	-0.555	***	(0.036)		0.022	***	(0.002)	-0.00022	***	(0.00005)
Middle East	-0.590	***	(0.058)		0.035	***	(0.006)	-0.00042	***	(0.00012)
North EU	-0.029		(0.029)		0.005		(0.003)	-0.00007		(0.00006)
Oceania	-0.165	***	(0.020)		0.007	**	(0.003)	-0.00005		(0.00006)
Philippines	-0.523	***	(0.051)		0.037	***	(0.006)	-0.00054	***	(0.00011)
South America	-0.546	***	(0.072)		0.034	***	(0.005)	-0.00044	***	(0.00009)
Southeast Asia	-0.550	***	(0.045)		0.029	***	(0.004)	-0.00032	**	(0.00009)
United Kingdom	0.070	***	(0.019)		-0.004	*	(0.002)	0.00007		(0.00004)
USSR	-0.806	***	(0.083)		0.054	***	(0.010)	-0.00073	***	(0.00019)
Vietnam	-0.572	***	(0.054)		0.036	***	(0.004)	-0.00048	***	(0.00010)
West Indies	-0.476	***	(0.054)		0.026	***	(0.004)	-0.00029	***	(0.00006)
Western EU	-0.131	***	(0.037)		0.004		(0.002)	-0.00001		(0.00004)
Number of Observations					1,240,006					
Squared					0.2642					

Note: Data from 1990 and 2000 Census and 2010 ACS. The standard error is robust and clustered in cohorts. Control variables include the year, gender, race, education, marital status, region, and age. Countries with relatively fewer immigrants are grouped with other countries within the same geographical area.

VI. Conclusion

Economic assimilation can be estimated with repeated cross-sectional data. The model that assumes linear assimilation and cohort fixed effect are misleading since cohort indicators capture the curvature in the assimilation process when the only measured assimilation effect is constrained to be linear. Researchers may falsely conclude that the difference between estimates from cross-sectional data and longitudinal data is due to cohort and outmigration biases. This paper shows neither decreasing cohort quality nor immigrants' return can account for the huge "bias" between cross-sectional and longitudinal estimates.

A potential simple check to run with longitudinal data analysis is to calculate the average years since migration of the longitudinal sample and the cross-sectional data used. When the longitudinal data is limited to long-term immigrants, the estimated assimilation process will appear less significant.

This paper also sheds some light on the heterogeneity of the assimilation process. The assimilation process is sensitive to age at arrival and country of origin. The larger the difference is between the home country and the host country, the more significant the assimilation effect will be. Roughly, the first ten years after arrival is sufficient for an immigrant to close most of the gap.

Based on the findings of this paper, I propose the following conjecture. The assimilation process is a projection of immigrants' learning curve of accumulating a package of skills for accommodating the host country's labor market. The assimilation path is determined by the level of skills they have attained before immigration and the shape of their learning curves after arrival.

VII. References

Abramitzky, R., & Boustan, L. (2017). Immigration in American Economic History. *Journal of Economic Literature, 55*(4), 1311-45.

Abramitzky, R., Boustan, L. P., & Eriksson, K. (2014). A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration. *Journal of Political Economy, 122*(3), 467-506.

Ahmed, B., & Robinson, G. J. (1994). Estimates of Emigration of the Foreign-Born Population: 1980–1990. *US Bureau of the Census Population Division Working Paper 9*.

Beaujot, R., & Rappak, P. J. (1989). The Link Between Immigration and Emigration in Canada, 1945-1986. *Canadian Studies in Population, 16*(2), 201-216.

Bleakley, H., & Chin, A. (2010). Age at Arrival, English Proficiency, and Social Assimilation among US Immigrants. *American Economic Journal: Applied Economics, 2*(1), 165-92.

Borjas, G. (1985). Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants. *Journal of Labor Economics, 3*(4), 463-89.

Borjas, G. J., & Bratsberg, B. (1994). Who leaves? The outmigration of the foreign-born. *National Bureau of Economic Research*(No. w4913).

Borjas, G. J., & Bratsberg, B. (1996). Who Leaves? The Outmigration of the Foreign-Born. *The Review of Economics and Statistics, 78*(1), 165-76.

Borjas, J. G. (1987). Self-Selection and the Earnings of Immigrants. *American Economic Review, 77*, 531-553.

Cheung, B. Y., Chudek, M., & Heine, S. J. (2011). Evidence for a Sensitive Period for Acculturation: Younger Immigrants Report Acculturating at a Faster Rate. *Psychological Science, 22*(2), 147-52.

Chiquiar, D., & 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), 239-281.

Chiswick, B. R. (1978). The Effect of Americanization on the Earnings of Foreign-Born Men. *Journal of Political Economy, 86*(5), 897-921.

Constant, A., & Massey, D. S. (2003). Self-Selection, Earnings, and Out-Migration: A Longitudinal Study of Immigrants to Germany. *Journal of Population Economics, 16*(4), 631-53.

Duleep, H. O., & Dowhan, D. J. (2002). Insights from Longitudinal Data on the Earnings Growth of U.S. Foreign-Born Men. *Demography, 39*(3), 485-506.

Dustmann, C., Fadlon, I., & Weiss, Y. (2011). Return Migration, Human Capital Accumulation and the Brain Drain. *Journal of Development Economics, 95*(1), 58-67.

Edin, P.-A., Fredriksson, P., & Aslund, O. (2003). Ethnic Enclaves and the Economic Success of, Evidence from a Natural Experiment. *Quarterly Journal of Economics, 118*(1), 329-57.

Edin, P.-A., LaLonde, R. J., & Aslund, O. (2000). Emigration of Immigrants and Measures of Immigrant Assimilation: Evidence from Sweden. *Working Paper*.

Feliciano, C. (2005). Educational Selectivity in US Immigration: How Do Immigrants Compare to Those Left Behind? *Demography, 42*(1), 131-152.

Hu, W.-Y. (2000). Immigrant Earnings Assimilation: Estimates from Longitudinal Data. *American Economic Review*, 368-372.

Kim, S. (2009). Economic Assimilation of Foreign Born Workers in the United States: An Overlapping Rotating Panel Analysis. *No. UWEC-2008-19*.

Lenneberg, E. H. (1967). *The Biological Foundations of Language* (Vol. 2). Hopital Practice.

Lubotsky, D. (2007). Chutes and Ladders? A Longitudinal Analysis of Immigrant Earnings. *Journal of Political Economy, 115*(5), 820-67.

Lubotsky, D. (2011, Feburary). The Effect of Changes in the US Wage Structure on Recent Immigrants' Earnings. *The Review of Economics and Statistics, 93*(1), 59-71.

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy, 66*(4), 281-302.

Noe-Bustamante, L., Flores, A., & Shah, S. (2019, September 16). *Facts on Hispanics of Mexican origin in the United States, 2017*. Retrieved from Pew Research Center: https://www.pewresearch.org/hispanic/fact-sheet/u-s-hispanics-facts-on-mexican-origin-latinos/

Qian, Z., & Lichter, D. (2007). Social Boundaries and Marital Assimilation: Interpreting Trends in Racial and Ethnic Intermarriage. *American Sociological Review, 72*(1), 68-94.

Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers, 3*(2), 135-146.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., & Sobek, M. (n.d.). *IPUMS USA: Version 9.0[dataset]. Minneapolis, MN: IPUMS, 2019*. Retrieved 2019, from https://doi.org/10.18128/D010.V9.0

Waters, M. C., & Jimenez, T. R. (2005). Assessing Immigrant Assimilation: New Empirical and Theoretical Challenges. *Annual Review of Sociology, 31*(1), 105-125. doi:10.1146/annurev.soc.29.010202.100026

Appendix. Data Description Tables

	Census 1990					
	Native	**1970 Earlier**	**[1970, 1975)**	**[1975, 1980)**	**[1980, 1985)**	**[1985, 1990]**
Observation (000)	141,430	7,409	1,939	2,264	2,757	2,714
AGE						
[25,44]	50.56%	30.45%	63.91%	73.83%	77.41%	77.57%
[45,64]	29.47%	38.07%	29.68%	20.28%	17.47%	16.92%
64+	19.97%	31.48%	6.41%	5.89%	5.12%	5.51%
GENDER						
Male	47.33%	44.10%	48.07%	50.62%	51.91%	50.21%
Female	52.67%	55.90%	51.93%	49.38%	48.09%	49.79%

RACE						
White	84.48%	57.52%	21.22%	20.85%	15.73%	21.39%
American Indian	0.73%	0.20%	0.22%	0.25%	0.24%	0.28%
Hispanic	1.97%	18.13%	23.40%	19.09%	21.11%	19.38%
Black	10.97%	4.39%	9.73%	8.28%	10.34%	7.73%
Chinese	0.12%	2.62%	5.25%	6.71%	7.87%	9.02%
Japanese	0.27%	1.30%	1.26%	0.76%	0.77%	2.87%
Other Asia	0.20%	4.19%	16.83%	22.06%	22.65%	20.75%
Other Race	1.27%	11.65%	22.08%	22.00%	21.29%	18.57%
EDUC						
<High School	22.83%	39.38%	43.14%	42.69%	42.82%	38.07%
High School	31.25%	22.65%	18.05%	17.01%	18.11%	18.48%
Some College	25.58%	20.51%	18.57%	18.84%	18.71%	17.27%
Bachelor Degree	13.28%	9.73%	11.74%	13.05%	12.26%	15.05%
> Bachelor Degree	7.07%	7.73%	8.51%	8.41%	8.10%	11.12%
REGION						
New England	5.41%	7.48%	5.31%	4.22%	4.34%	5.49%
Middle Atlantic	14.64%	26.89%	24.57%	18.32%	21.42%	22.62%
East North Central	17.66%	11.73%	8.99%	8.16%	6.53%	7.73%
West North Central	7.70%	2.12%	1.17%	1.56%	1.43%	1.86%
South Atlantic	18.31%	15.66%	13.60%	11.70%	14.85%	14.84%
East South Central	6.62%	1.11%	0.75%	0.75%	0.63%	0.89%
West South Central	10.56%	6.43%	8.84%	10.93%	10.12%	7.18%
Mountain	5.49%	4.04%	3.25%	3.40%	2.98%	3.28%
Pacific	13.61%	24.54%	33.52%	40.96%	37.69%	36.11%
MARST						
Married, Present	63.58%	61.94%	68.71%	67.49%	61.77%	56.41%
Married, Absent	1.26%	2.03%	3.24%	3.61%	5.17%	9.28%
Separated	2.48%	2.71%	3.82%	3.88%	4.33%	3.47%
Divorced	10.32%	8.85%	7.63%	6.37%	5.78%	3.99%
Widowed	9.09%	14.96%	4.18%	3.85%	3.62%	4.10%
Single	13.27%	9.50%	12.43%	14.81%	19.32%	22.76%

Note: Data from the 1990 Census. The sample includes people over 24 years old.

	Census 2000							
	Natives	1970 Earlier	[1970,1975)	[1975,1980)	[1980,1985)	[1985,1990]	(1990,1995]	(1995,2000]
Observation (000)	55,474	5,950	2,285	2,815	3,524	4,965	3,799	3,370
AGE								
[25,44]	45.31%	17.59%	41.15%	54.99%	64.75%	74.40%	74.87%	77.95%
[45,64]	34.52%	43.87%	49.06%	37.51%	28.77%	20.15%	19.11%	17.27%
64+	20.17%	38.54%	9.80%	7.50%	6.48%	5.46%	6.02%	4.78%
GENDER								
Male	47.62%	43.67%	47.98%	50.15%	51.64%	50.91%	48.22%	50.96%
Female	52.38%	56.33%	52.02%	49.85%	48.36%	49.09%	51.78%	49.04%
RACE								
White	81.64%	52.86%	22.34%	19.26%	13.08%	13.91%	19.29%	22.12%
American Indian	0.80%	0.29%	0.42%	0.41%	0.44%	0.47%	0.38%	0.38%
Hispanic	2.41%	20.24%	21.62%	17.85%	20.88%	22.30%	20.55%	21.15%
Black	11.62%	4.26%	8.43%	7.10%	8.78%	7.38%	6.79%	6.69%
Chinese	0.14%	2.57%	4.52%	5.97%	6.92%	6.58%	6.58%	6.43%
Japanese	0.24%	1.22%	1.01%	0.65%	0.54%	0.59%	0.93%	2.14%
Other Asia	0.28%	4.57%	15.38%	21.15%	21.06%	17.03%	19.46%	16.98%
Other Race	2.88%	13.99%	26.29%	27.62%	28.31%	31.74%	26.02%	24.10%
EDUC								
<High School	16.57%	33.10%	35.99%	36.83%	40.36%	42.31%	38.44%	34.01%
High School	30.20%	22.52%	18.72%	17.56%	18.40%	19.25%	19.20%	18.10%
Some College	28.78%	22.72%	21.33%	20.91%	19.36%	17.49%	16.87%	15.13%
Bachelor Degree	15.81%	11.67%	14.53%	15.49%	13.61%	12.27%	14.11%	18.38%
> Bachelor Degree	8.65%	10.00%	9.44%	9.20%	8.26%	8.69%	11.38%	14.37%
REGION								
New England	5.17%	6.70%	5.02%	4.09%	3.93%	4.37%	4.26%	5.17%
Middle Atlantic	13.53%	23.56%	21.29%	16.48%	18.95%	20.31%	20.76%	18.30%

East North Central	17.22%	10.83%	8.83%	8.36%	6.69%	7.39%	9.22%	10.23%
West North Central	7.58%	2.12%	1.61%	1.84%	1.71%	1.82%	2.64%	3.24%
South Atlantic	19.24%	18.44%	15.50%	13.18%	15.96%	14.96%	17.62%	20.17%
East South Central	6.89%	1.41%	1.12%	0.95%	0.86%	0.87%	1.32%	1.96%
West South Central	10.88%	7.25%	9.75%	11.57%	11.21%	9.18%	10.15%	10.53%
Mountain	6.41%	5.28%	4.79%	4.93%	4.57%	5.17%	5.46%	6.08%
Pacific	13.08%	24.41%	32.09%	38.60%	36.13%	35.92%	28.56%	24.31%
MARST								
Married, Present	60.02%	60.93%	64.80%	65.06%	64.65%	64.20%	61.20%	55.59%
Married, Absent	2.41%	2.51%	3.40%	3.61%	4.34%	5.25%	6.62%	10.42%
Separated	2.30%	2.57%	3.75%	3.62%	4.00%	3.83%	3.57%	2.97%
Divorced	12.30%	11.52%	9.94%	8.25%	7.76%	6.25%	5.33%	4.38%
Widowed	8.32%	15.00%	5.02%	4.16%	3.87%	3.28%	3.37%	3.16%
Single	14.65%	7.47%	13.09%	15.29%	15.38%	17.19%	19.91%	23.48%

Note: Data from 2000 Census. The sample includes people over 24 years old.

	ACS 2010									
	Natives	1970 Earlier	[1970,1975)	[1975,1980)	[1980,1985)	[1985,1990]	(1990,1995]	(1995,2000]	(2000,2005]	(2005,2010]
Observation (000)	167,709	4,770	2,119	2,654	3,695	5,437	4,346	5,587	4,611	3,418
AGE										
[25,44]	38.32%	3.40%	20.63%	28.46%	35.78%	50.63%	61.50%	71.01%	74.52%	71.67%
[45,64]	40.77%	43.66%	55.76%	56.83%	53.08%	41.27%	30.60%	24.00%	20.50%	20.31%
64+	20.90%	52.94%	23.61%	14.72%	11.15%	8.10%	7.91%	4.98%	4.97%	8.02%
GENDER										
Male	48.07%	44.34%	47.26%	49.05%	50.83%	50.05%	47.61%	49.50%	47.43%	48.22%
Female	51.93%	55.66%	52.74%	50.95%	49.17%	49.95%	52.39%	50.50%	52.57%	51.78%
RACE										
White	78.71%	50.16%	22.97%	20.65%	14.45%	13.84%	16.55%	16.33%	15.51%	18.36%
American Indian	0.81%	0.30%	0.33%	0.39%	0.36%	0.44%	0.37%	0.33%	0.46%	0.41%
Hispanic	4.27%	25.43%	30.14%	25.43%	28.51%	31.50%	29.21%	32.37%	30.82%	24.55%
Black	12.31%	4.48%	8.61%	7.30%	9.59%	8.45%	7.59%	8.27%	9.10%	8.86%
Chinese	0.21%	2.62%	4.36%	5.55%	6.49%	6.14%	6.18%	5.76%	5.29%	7.64%
Japanese	0.22%	1.31%	0.84%	0.60%	0.46%	0.54%	0.56%	0.44%	0.65%	1.56%
Other Asia	0.54%	5.20%	15.78%	22.56%	21.85%	17.91%	19.70%	15.89%	17.76%	23.58%
Other Race	2.92%	10.51%	16.96%	17.52%	18.30%	21.18%	19.84%	20.60%	20.41%	15.05%
EDUC										
<High School	10.83%	26.54%	30.82%	29.22%	30.30%	33.05%	32.70%	32.97%	31.59%	27.32%
High School	29.76%	24.23%	20.95%	19.47%	22.14%	23.05%	22.76%	23.65%	23.59%	20.78%
Some College	30.93%	24.11%	22.24%	22.39%	22.10%	20.25%	18.92%	16.92%	16.57%	14.58%
Bachelor Degree	18.14%	13.47%	15.61%	17.20%	16.08%	14.93%	15.31%	15.41%	16.28%	22.16%
> Bachelor Degree	10.33%	11.65%	10.38%	11.73%	9.38%	8.72%	10.30%	11.04%	11.97%	15.16%
REGION										
New England	4.87%	6.34%	5.06%	3.79%	3.86%	4.06%	4.24%	4.33%	5.13%	5.29%
Middle Atlantic	12.60%	20.82%	19.41%	15.23%	16.97%	18.37%	19.20%	16.83%	16.69%	17.96%

East North Central	16.57%	9.51%	8.95%	8.18%	6.51%	6.65%	8.10%	9.19%	8.33%	8.88%
West North Central	7.53%	2.23%	1.57%	2.07%	2.12%	1.95%	2.82%	3.00%	2.90%	3.27%
South Atlantic	19.83%	20.48%	16.97%	14.27%	17.68%	16.89%	17.94%	20.88%	22.59%	21.56%
East South Central	6.94%	1.53%	1.21%	1.32%	1.14%	0.98%	1.15%	1.91%	2.27%	2.81%
West South Central	11.34%	8.30%	10.00%	11.65%	12.04%	10.42%	11.18%	12.49%	11.69%	11.83%
Mountain	7.17%	6.30%	5.80%	6.15%	5.75%	5.77%	5.90%	6.41%	6.05%	5.41%
Pacific	13.15%	24.50%	31.03%	37.33%	33.94%	34.91%	29.48%	24.95%	24.34%	22.99%
MARST										
Married, Present	54.81%	56.62%	62.36%	63.33%	59.93%	60.00%	61.18%	60.67%	56.61%	51.59%
Married, Absent	1.71%	2.22%	2.90%	2.73%	3.84%	4.29%	4.45%	4.97%	6.36%	9.51%
Separated	2.42%	2.37%	3.65%	3.45%	3.87%	3.97%	3.77%	3.63%	3.34%	2.36%
Divorced	14.07%	13.83%	12.53%	11.74%	10.85%	9.03%	8.00%	6.50%	5.87%	4.46%
Widowed	7.66%	17.12%	8.06%	5.92%	4.76%	3.83%	3.70%	2.76%	2.77%	4.62%
Single	19.33%	7.84%	10.51%	12.82%	16.75%	18.88%	18.90%	21.48%	25.05%	27.46%

Note: Data from 2010 ACS. The sample includes people over 24 years old.