Ethnic Neighborhoods and Naturalization in the United States

A Senior Honors Thesis for the Department of Economics,

Manasvini Baba

Tufts University, 2014

Abstract

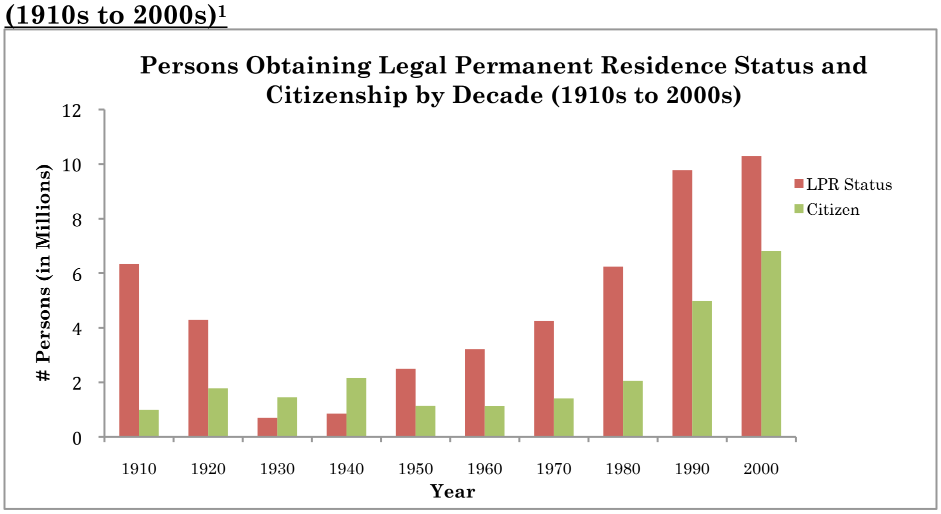
This thesis seeks to understand the naturalization choices of Chinese and Indian immigrants in the United States. I focus in particular on the association between ethnic neighborhoods (or enclaves) and naturalization outcomes. Using microdata from the 2011 American Community Survey, I explore the impact of individual and neighborhood attributes on (a) the probability of an immigrant being naturalized and (b) how soon an eligible immigrant chooses to naturalize. My analysis produces mixed results with regards to the neighborhood variable: depending on the model specification and immigrant group in question, the coefficient of living in or near an ethnic enclave is positive but not always significant. The individual attribute variables confirm the findings of previous literature. One exception is the dummy variable of having children, which I find to be negatively correlated with the probability of naturalization.

Table of Contents

Chapter 1: Introduction

Contemporary levels of immigration to the United States are rising: as Figure 1 illustrates, immigrant inflows have more than tripled from about 3.2 million in the 1960s to 11.3 million in the first decade of the millennium. Though this increased inflow has sparked a contentious debate over immigration reform and border control in the United States, immigrant integration is a subject of equal import. Integration may be measured across multiple dimensions, but this paper will focus on naturalization, the process by which an immigrant attains citizenship in the host country, as the metric of choice.

Figure 1: Persons Obtaining Legal Permanent Resident Status and Citizenship by Decade (1910s to 2000s)[1]



Source: Department of Homeland Security, Yearbook Statistics 2011, *Persons Obtaining Legal Permanent Resident Status: Fiscal Years 1820 to 2011*

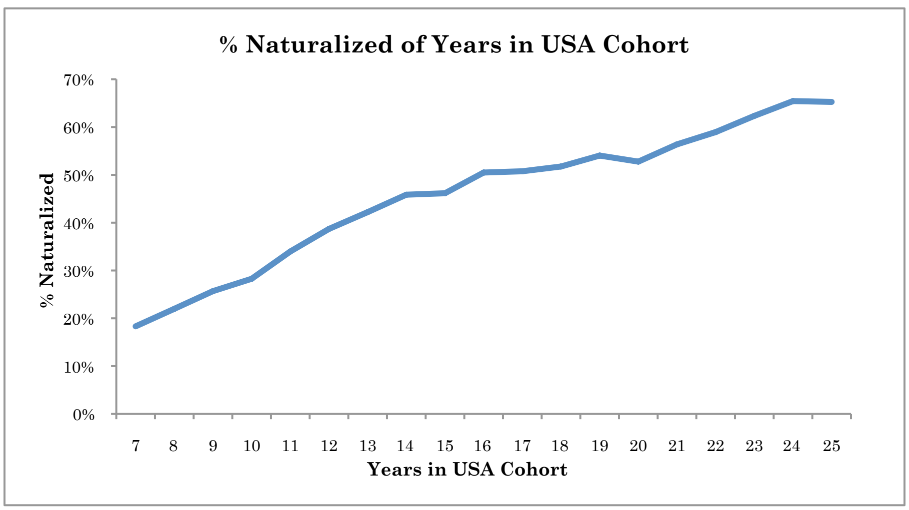
Naturalization is a topic of great interest to political scientists, sociologists, and economists alike. Political scientists view naturalization as a critical "pathway through which immigrant minorities obtain the right to vote" (Hainmueller, 2013); sociologists consider naturalization as an essential part of how "immigrant minorities have advanced collectively and gained recognition from the dominant society" (Portes and Curtis, 1987); and economists are interested in the labor market impact of naturalization and whether "naturalization leads to higher wages, either immediately or by accelerating wage growth" (Bratsberg, Ragan and Nasir, 2002). Despite such cross-disciplinary pertinence, sociologists have published most of the papers to date on this topic, with economists only recently

[1] Legal permanent residents are immigrants that are eligible to naturalize after five years of attaining LPR status. Chapter 2 further explains LPRs and the U.S. naturalization process.

contributing. By analyzing naturalization decisions through a rational choice and social interactions framework, I seek to contribute to the growing body of literature studying naturalization through an economic lens.

Figure 1 illustrates that the volume of naturalization petitions has risen steeply over the past three decades, but these petitions only represent a fraction of the pool actually eligible to naturalize. Estimates from the Office of Immigrant Statistics reveal that there were 8.8 million legal permanent residents eligible to naturalize in 2012, but the United States Citizenship and Immigration Services only received naturalization petitions corresponding to 10% of that number (899,162) in the same year (Batalova and Auclair, 2013). Although a portion of the stock of 8.8 million eligible immigrants may simply naturalize in subsequent years, Figure 2 shows that even after 25 years in the United States, 30% of immigrants have still not naturalized.[2,3]

Figure 2: Percent Naturalized of Years in USA Cohort



Source: *IPUMS USA 5%,* American Community Survey 2011

Such trends prompt this paper's core research questions: what are the factors most salient in explaining (a) the probability of an immigrant being naturalized and (b) the speed

[2] The Census Bureau does not collect information on the legal status of immigrants, so eligibility is determined simply by limiting the sample to the foreign born who have been in the United States for 7+ years. Chapter 2a reviews this eligibility concern in more detail.

[3] 85% to 90% of later cohorts (i.e. 1950s and 1960s arrivals) are naturalized; however, since earlier immigrants are very different from contemporary ones (e.g. in terms of national origin) and smaller in sample size, I limit my analysis to the post-1986 cohort of immigrants (i.e. after the Immigration Reform and Control Act: a critical turning point in immigration policies, which legalized illegal immigrants and put pressures on employers and borders to stem further illegal immigration).

at which an immigrant naturalizes? Building from the existing body of literature, I use a 3% sample of the 2011 American Community Survey to study a similar range of individual, family, and locational variables. However, I distinguish my analysis from my predecessors in two key ways. First, I utilize recently available information on the timing of naturalization to study the factors influencing the rates at which immigrants naturalize. Second, I adopt a more rigorous approach in specifying locational influences by exploiting data ranging from the metropolitan to census tract level.[4]

I base my analysis on a sample of immigrants born in Mainland China and India, two of the fastest growing immigrant groups in the United States.[5] Using Public Use Microdata Areas as my measure of neighborhoods, I hypothesize that neighborhoods with a higher concentration of co-ethnics will be correlated with a higher probability of and faster time to naturalization.[6] I further examine how these effects vary according to the skill level of the immigrant and the socioeconomic status of the neighborhood.

The rest of the paper is organized as follows: Chapter 2 describes current U.S. naturalization policies and presents a profile of Chinese and Indian immigrants vis-à-vis other immigrant source groups. Chapter 3 provides a literature overview of the causes, consequences, and measures of immigrant enclaves and neighborhoods, and it also discusses some of the difficulties inherent in studying social interactions. Chapter 4 reviews the current naturalization literature and the ways in which this paper is situated among the existing studies. Chapter 5 discusses the economic theories underpinning this study. Chapter 6 presents the data and summary statistics. Chapter 7 introduces the empirical models, and Chapter 8 presents the results of the analysis. Finally, Chapter 9 concludes the paper and offers suggestions for further research.

[4] Refer to Chapter 4 for an overview of census spatial units
[5] According to *Yearbook 2011 Statistics* from the Department of Homeland Security: In the period 2002-2011, after Mexico, China was the second most prevalent source country of immigrants at 6.6% while India was the third at 6.3%
[6] See footnote 3

Chapter 2a: Naturalization Background: U.S. History & Policy

In this section, I document the evolution of naturalization policies and processes in the United States. Most of the Chinese and Indian immigrants in my sample immigrated to the United States in the 1980s onwards. Nonetheless, my sample still includes immigrants who arrived as far back as 1919. As such, it is important to examine how naturalization policies and the political climate have evolved since the early twentieth century in order to understand the costs, benefits, obstacles, and incentives confronting these immigrants at different points in time. Such an understanding provides a critical context for interpreting time to naturalization for different immigrant cohorts. I structure my discussion in five stages: (1) eligibility, (2) application steps, (3) processing time and external incentives, (4) current naturalization process, and (5) key conclusions.

Eligibility

Chinese and Indian immigrants in the first part of the twentieth century were ineligible to naturalize simply due to their race. The Chinese Exclusion Act of 1882, which barred immigration of Chinese laborers, was still in full force in the early twentieth century and was complemented by the Asiatic Barred Zone Act of 1917, which further banned immigration from the rest of the Asian continent (Vigdor, 2009). Naturalization bans went hand in hand with these immigration restrictions, and they were cemented by Supreme Court cases such as *United States vs. Bhagat Singh Thind* (1923), which ruled against the rights of Indian-Americans to naturalize (McMahon, 2001).[7] These citizenship restrictions were overturned for the Chinese in 1943 with the passage of the Magnuson Act and they were also repealed in 1946 for Indians with the passage of the Luce-Celler Act (Vigdor, 2009). It was not until the 1965 Hart-Celler Act, however, that the entire system of racially based immigration and naturalization restrictions was overturned (Vigdor, 2009).

Although race may no longer be a criterion in contemporary naturalization policies, a weak analog remains in the good moral character requirement. The good moral character requirement dates back to the 1790 Naturalization Act (Lapp, 2012). Up until the mid-twentieth century, good moral character was determined primarily by a review of an immigrant's criminal history and other character attributes deemed appropriate by each local court. However, the 1952 McCarran Walter Act explicitly listed qualities constituting good moral character, limiting the flexibility of local courts in making their own character assessments (Lapp, 2012). In addition to criminal history, factors to be considered included habitual drunkenness, polygamy, and illegal gambling (Lapp, 2012). To this day, the good moral character requirements have continued to evolve with each iteration of the naturalization petition form.

Last but not least is the residency requirement. Since the Naturalization Act of 1795, United States law has required an immigrant seeking citizenship to live continuously in the United States for at least five years before naturalizing (Lapp, 2012). The process was expedited for wives of U.S. citizens by the Cable Act of 1922 and for husbands by the Naturalization Act of 1934 (U.S.C.I.S, 2013). The 1952 McCarran Walter Act reduced the residency requirement for spouses of U.S. citizens to three years (USCIS, 2013). During World War I, immigrants in the armed forces could naturalize in military camps, and since then, the residency requirement has not been applied to those who serve or have served in the military (Family Search, 2010).

An important complication for research on naturalization eligibility is the fact that the Census Bureau does not collect information on the legal status of immigrants; rather, it asks if the respondent is foreign born, his or her country of birth if so, and his or her year of arrival.[8] It also asks about citizenship but only since 2007, the year of naturalization. Thus, the immigrants in my sample represent a mix of legal permanent residents, non-immigrant visa holders, and undocumented immigrants. It is important to distinguish between these classes of immigrants because only legal permanent residents are eligible to apply for citizenship and the residency requirement only applies to them.[9]

Before 1940, immigrants could count the five-year requirement starting from their time of arrival in the United States. In 1940, the Alien Registration Act required immigrants to first formally register with the federal government at their local post offices.

[8] The year of arrival is an ambiguously phrased question that consequently poses a lot of empirical challenges- respondents might record the year they arrived as a tourist, student, or temporary worker rather than their year of legal immigration. In the 2011 American Community Survey, the question was phrased as "**When did this person come to live in the United States?"** Further instructions for enumerators and respondents were: "**If the person came to live in the United States (that is, the 50 states and the District of Columbia) more than once, enter the latest year he or she came to live in the United States."** (IPUMS USA, 2011).

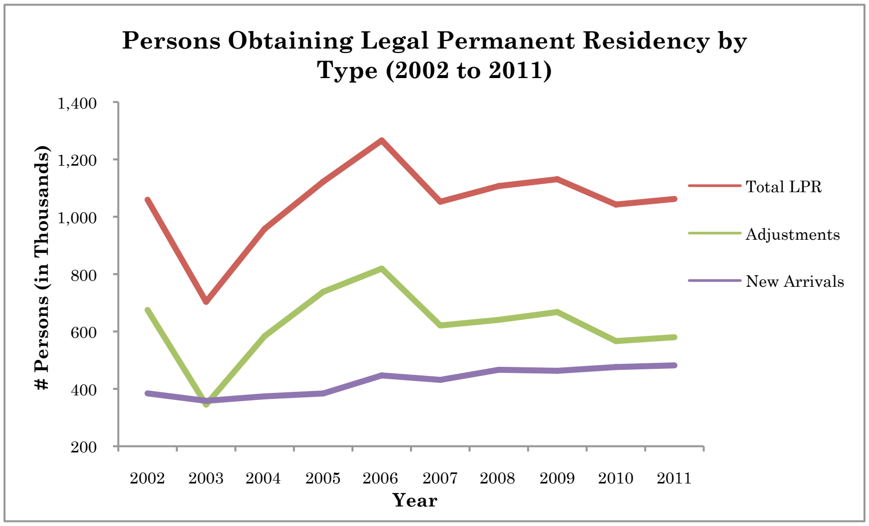
[9] Using the New Immigrant Survey- Pilot, Jasso et al. (2000) provide a detailed overview of the limitations of government data on immigration and also analyze and compare skill and earnings attributes of legal immigrants, legal non-immigrants, and illegal immigrants.

This registration was a pre-cursor to the “green card,” a formal alien identification card that was first issued in 1950 (Jaramillo, 2012). The five-year requirement subsequently counted only after the immigrant had secured a green card. The passage of the Hart-Celler Act in 1965 established a system of preferential visas based on categories such as family reunification and employment as opposed to national origin. This system created a distinction between immigrants, who were considered permanent residents upon arrival, and non-immigrants, who could only be considered a permanent resident after adjusting their status (Vigdor, 2009). Chinese and Indian immigrants entering on family or employment based immigration visas currently contend with long waitlists ranging from 3 to 8 years due to national caps on the amount of visas that can be issues (U.S. Dept. of State: Bureau of Consular Affairs, 2013).[10]

Common types of non-immigrant visas are employment-based, tourism-related, business-related, refugee-based, and education-related (U.S. Dept. of State: Bureau of Consular Affairs, 2014). In order for non-immigrants to adjust their status, a U.S. citizen, green-card holder, employer, or humanitarian organization typically must file a petition on the alien’s behalf (U.S.C.I.S, 2011). The processing time for adjusting statuses depends on the type of application in question: employment-based adjustments can take anywhere from 6 months to 2 years, while family based adjustments may take up to three or four years (Zhang and Associates, 2012). Depending on the employer in question or visa quotas, temporary visa holders may have to wait an additional few years before they can even apply to adjust their status (Zhang and Associates, 2012). Due to incredibly long immigration waitlists, non-immigrant visa acquisition is a popular pathway through which aliens can subsequently acquire legal permanent residency through adjustment of status. The number of immigrants in my sample that may be non-immigrants and may want to settle permanently in the United States is non-trivial. I estimate non-immigrants to account for 25% of the Indian sample and 13% of the Chinese sample[11], and Figure 3 below illustrates that a substantial portion (i.e. 59%) of recent legal permanent residents come from adjustment of status compared to new arrivals.

[10] See Bureau of Consular Affairs for details on quotas for different visa categories
[11] Appendix A includes full details of these estimates’ derivation

Figure 3: Legal Permanent Residency by Type



Source: Department of Homeland Security, Yearbook Statistics 2011, *Persons Obtaining Legal Permanent Resident Status by Type and Major Class of Admission: Fiscal Years 2002 to 2011*

The evolution of eligibility requirements yields a few key implications for the purpose of this study. On an empirical level, the lack of information on immigrants' legal status means that a prime challenge for this paper will be determining the eligibility of immigrants. In terms of expected time-to-naturalization trends, we may expect longer times to naturalization for early twentieth century immigrants. This is because they were unable to naturalize and discouraged from doing so due to the rampant discrimination of the era, even after being legally accorded the right to naturalize. On the other hand, immigrants in the post-1965 era would have naturalized more quickly for a few reasons. First, Chinese and Indian immigrants may have likely been more comfortable setting down roots in the United States, since the dissolution of national origin quotas made the system much less discriminatory and more welcoming to them. Second, since many Chinese and Indian immigrants were arriving through family reunification visas, these immigrants would be more likely to naturalize with a family united in one place. Third, particularly for the most recent set of immigrants, there may be a selection effect at play: since attaining legal permanent residency and even non-immigrant visas is steeply rising in difficulty, the immigrants who choose to weather the process may likely be the ones who desire naturalization and permanent settlement in the United States the most.

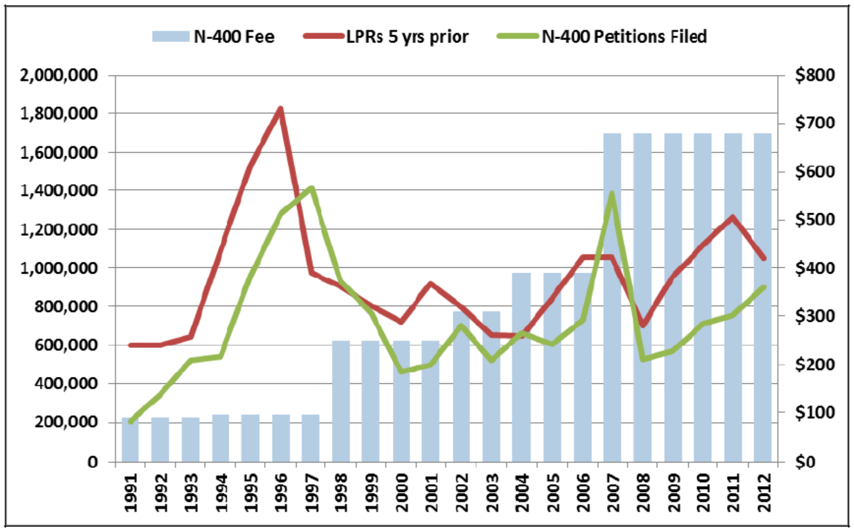
Application Steps

Once eligible, the immigrant must complete a series of steps in order to naturalize. Up until 1952, immigrants had to first declare their intention to naturalize and then file a petition two years later, though they could declare their intention prior to living in the United States for the full five years. After 1952, a declaration was no longer necessary, and immigrants simply had to submit photographs, a set of fingerprints, and a petition in order to naturalize. Over time, the naturalization process has become increasingly standardized and less subject to the impulses of local courts. In 1990, for instance, federal bodies appointed the naturalization examiners in charge of approving immigrants' petitions, thus eliminating the potential biases of local court officials (U.S.C.I.S, 2013). The clear-cut description of the good moral character requirement is another example of such standardization.

After filing a petition, immigrants must next pass an English and civics test. The English requirement has been a part of the naturalization process since the Immigration Act of 1917; in 1990, the requirement was waived for immigrants over the age of 50 years (U.S.C.I.S., 2013). The civics test has also always been a part of the process, though it was formalized with the passage of the 1952 McCarran-Walter Act. The civics test was reformed in 2006 to be less trivia-based and more centered on American values of democracy. Prior to 2006, applicants could buy free study materials for $8.50, but with this change, free flashcards were published online and made more accessible to applicants (U.S.C.I.S., 2013). Though the content of the civics and English test has likely varied over time, the changes discussed above are the only ones that are clearly documented. One may speculate, however, that the human capital costs of naturalizing are lower in contemporary times: since the content is more standardized and predictable, immigrants can more efficiently invest their time in preparing for the tests and can also avail themselves of the growing number of immigrant service groups, such as "English as a Second Language" centers, that cater to immigrants who may struggle with the requirements.

A third element to discuss is the naturalization fee. The earliest information available on naturalization fees dates back to 1991. As illustrated by Figure 4, there have been significant fee increases over the past two decades. The increases were particularly high in 1998 and 2007, with the current price now standing at $680. Correspondingly, the number of N-400 petitions appeared to drop before the price increase in 1998 while it first soared and then descended before and after the change in 2007.

Figure 4: Petition Volume and Fees for N-400 Naturalization Forms



Source: CRS analysis of USCIS data and Office of Immigration Statistics data.

Lastly, since neither China nor India permits dual nationality, Chinese and Indian immigrants must by law forfeit their origin country citizenship upon naturalizing in the United States. There are both monetary and non-monetary costs involved in doing so.

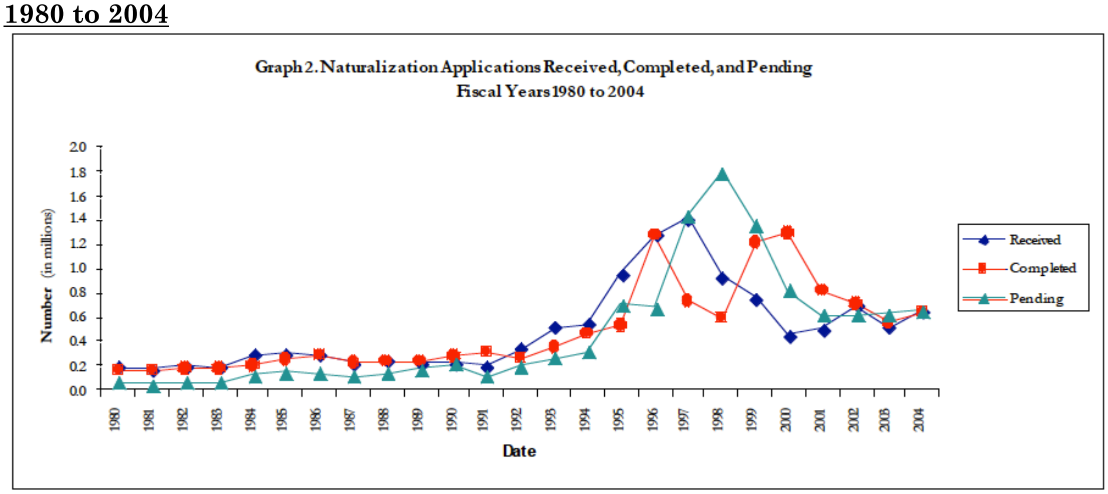
The monetary costs are small. China requires a modest fee of $75 to surrender Chinese citizenship, and India charges $25 before May 31, 2010 and $175 after this date to surrender Indian citizenship (GovHK, 2014; Immihelp, 2014). The non-monetary costs, however, are quite substantial, since naturalized citizens lose all political rights they once enjoyed in their country of origin and may face certain psychological costs born from formally breaking ties with their home country. These losses are more severe for Chinese immigrants than they are for Indian immigrants. Once they relinquish their Chinese citizenship, Chinese immigrants may return to China for an extended stay through employer sponsorship. However, it is much harder for retirees to return, since China has not developed visas catering to extended home or family visits. On the other hand, as of 2005 for the cost of $275, Indian immigrants can opt for a lifelong visa, the Overseas Citizen of India (OCI). The OCI visa allows India immigrants to stay and work in India for an indefinite period of time, own land, and operate and manage Indian bank accounts and investments. Prior to the OCI visa, immigrants could also opt for a Person of Indian Origin visa for $25, which entitled them to similar rights but expired after 15 years (NRI Realty News, 2014).

On the whole, studying the evolution of the application process provides insight into the different costs confronting immigrants throughout the naturalization process. The civics and English tests have always been a part of the naturalization process, but the standardization of their content and the availability of study materials have likely reduced the human capital costs of these exams for contemporary migrants. On the other hand, the monetary costs of the process have dramatically increased over time. Though immigrants who are financially strapped can apply for a fee waiver, the financial burden may still be too heavy for some.

Processing Time & External Incentives

In addition to individual eligibility requirements and application steps, policies and bureaucratic pressures of the time also significantly inform patterns in the time to naturalize for different immigrant cohorts. By comparing the amount of petitions received to the next year's amount completed in Figure 5, it appears that the processing time in the 1980s was around a year. However, in the 1990s, there was a surge in the number of pending petitions.

<u>Figure 5: Naturalization Applications Received, Completed, and Pending: Fiscal Years 1980 to 2004</u>



Source: *Office of Immigration Statistics, Department of Homeland Security, G-22.3 Naturalization Summary Report. Total completed is the sum of approved and denied. Data are current as of October 2004. Note that counts of the last three years may fluctuate upwards or downwards due to revisions.*

The exceptional trend of the 1990s can be attributed to a couple of factors. First, the 1986 Immigration and Reform Control Act legalized all undocumented migrants who arrived in the United States after 1982, partially contributing to the quadrupling of naturalization petitions from 385,000 to 1.5 million between 1992 and 1997.

Other contributing factors included the 1992 Green Card Replacement Program, which required long-term permanent residents to replace their green card by 1996 and declared that green cards must be renewed every ten years at a fee of $450. Since these immigrants would have to face some sort of costly administrative process in the near future, this program is likely to have motivated long-term immigrants to choose to naturalize instead of renewing their green cards (Mazzolari, 2011). Moreover, Proposition 187 in 1994 and the Personal Responsibility and Work Reconciliation Act in 1996 further incentivized naturalization, particularly for lower income immigrants, by restricting access to certain welfare benefits to citizens only.

To tackle this surge in naturalization petitions, Clinton's Administration announced the launch of Citizenship USA in 1995, a program aiming to reduce the backlog and process applications "from start to citizenship within six months, by the end of FY 1996" (Jenks, 1997). It succeeded in reducing a backlog of 500,000 naturalization petitions but was swiftly dismantled after being heavily criticized for letting immigrants with criminal backgrounds slip through the cracks (On the Issues, 2014). It appears with the transition from the INS to the USCIS in 2003, however, the administrative apparatus for processing naturalization applications became more streamlined and efficient. Nonetheless, security concerns stemming from 9/11 increased the processing time of applications, and, at the same time, are likely to have motivated immigrants to naturalize in order to safeguard their right to stay in the United States.

In summary, the mid-1980s cohort of immigrants as well as post-9/11 immigrants may have taken longer to naturalize than their predecessors due to administrative growing pains, policy reforms, and the political climate of the period.

Current Naturalization Process

With this evolution in mind, the current naturalization requirements are as follows: [12]

[12] There are certain exceptions to these stipulations according to the immigrant's age, his duration of stay, if he's married to a U.S. citizen, and if he has served in the armed forces.

Figure 6: Naturalization Requirements	
Requirement Type	**Details**
Age Eligibility	Be 18 years or older at time of filing
Residency Eligibility	Be green-card holder of at least five years
	Live in the same state for at least three months prior to filing
	Not have travelled outside U.S.A for more than 180 days/year since receiving green card and live in U.S.A for at least 30 months of five year period
Moral Character Eligibility	Be of good moral character and not have a politically controversial background.[13]
Human Capital Requirements (i.e. income, language, and education sensitive requirements)	Pay application fee of $680, covering the cost of the application ($595) and the biometrics process[14] ($85): • Filing Form I-912 allows the immigrant to waive the fee if he or she is receiving means-tested benefits or can demonstrate financial hardship.
	Read, speak, write, and understand English and pass a civics exam
Commitment Requirement	Be willing and able to take an oath of allegiance to the United States

Source: *United States Citizenship and Immigration Services*, "Citizenship through Naturalization"

Key Conclusions

This section has illustrated that history matters in the naturalization outcomes of immigrants. First, early Indian and Chinese immigrants could not naturalize due to exclusionary policies and potential biases from local courts, even after such policies were eliminated. Second, the time it takes to naturalize varied historically as a product of both eligibility requirements as well as administrative backlogs. Third, the 1965 Hart-Celler Act and the series of reforms in the 1990s played a particularly crucial role in stimulating the demand for naturalization. Fourth, immigrants have faced different costs of naturalizing over time, related to the application fees, the investment of time, and the investment of human capital necessary to pass the exams.

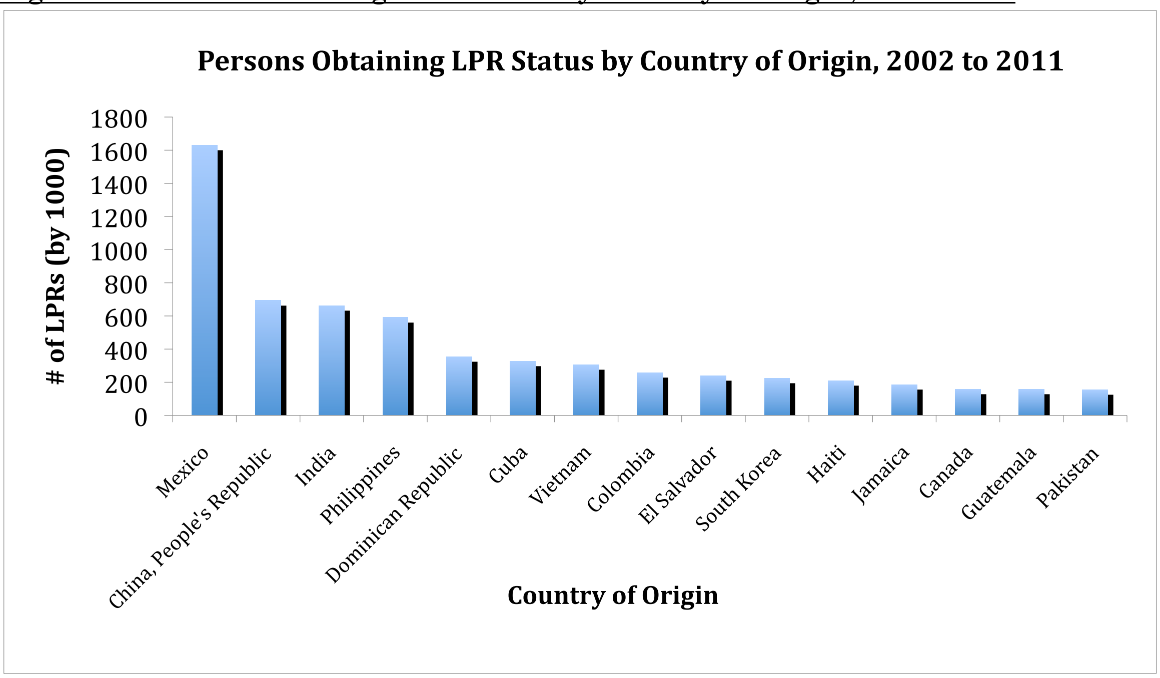
[13] This requirement refers to a range of questions on the N-400 pertaining to affiliations with Nazi, communist, or terrorist parties, for example.
[14] The biometrics process involves attending a scheduled appointment at the local USCISC Application Support Center and having one's fingerprints taken

Chapter 2b: Profile of Chinese and Indian Immigrants

Immigrants tend to significantly differ not only according to personal characteristics but also according to national origin group. As a result, it is important to understand the differences between Chinese and Indian immigrants and the larger immigrant pool in the United States, and how these differences may influence their comparative assimilation outcome.

Starting with the size of different immigrant groups, Figure 7 shows that Chinese and Indian immigrants constitute two of the largest groups obtaining legal permanent residency in the United States, falling only behind Mexico.

Figure 7: Persons Obtaining LPR Status by Country of Origin, 2002-2011[15]

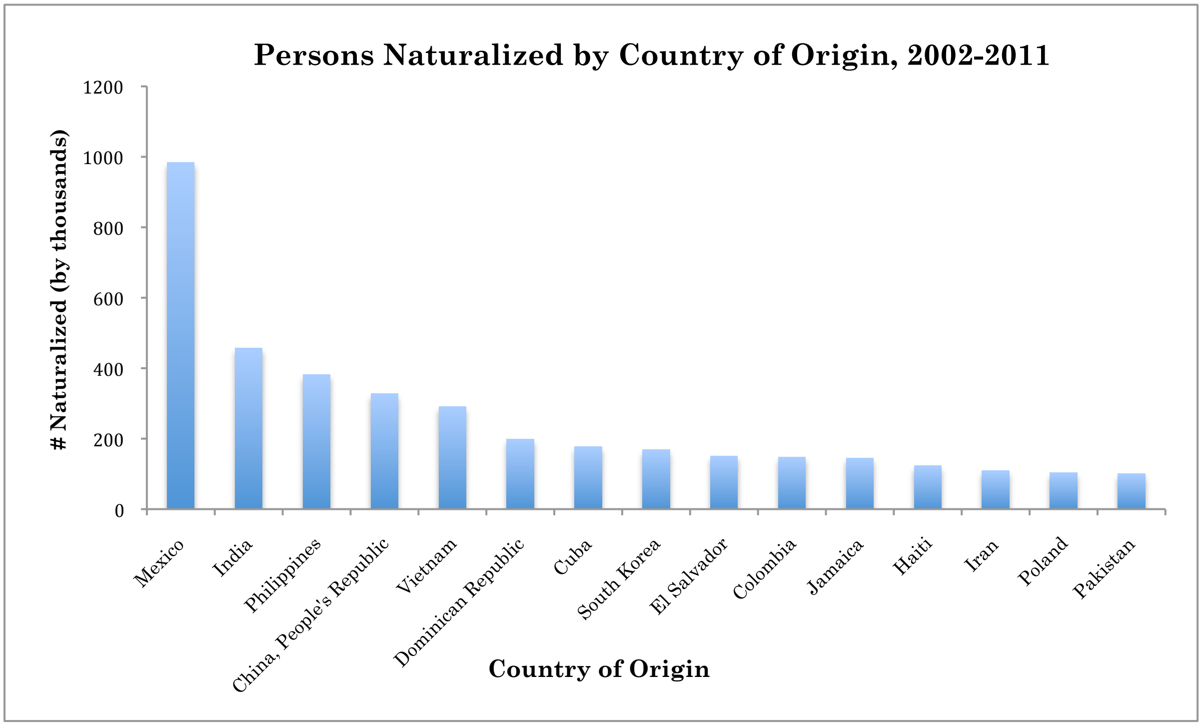


Source: *Department of Homeland Security,* Yearbook 2011 Statistics, "Persons Obtaining Legal Permanent Resident Status by Region and Country of Birth: Fiscal Years 2002 to 2011"

In the years 2002-2011, Chinese and Indian have also been among the most dominant origin groups naturalizing, with India coming in second and China coming in fourth. Since Mexico and the Philippines permit dual citizenship and India offers an emigrant-friendly OCI visa, China's relatively lower naturalization rate could be a product of its citizenship policies.

[15] The Department of Homeland Security only releases country of origin information for 2002-2011, though a comparison aggregating post-1965 flows would be more telling.

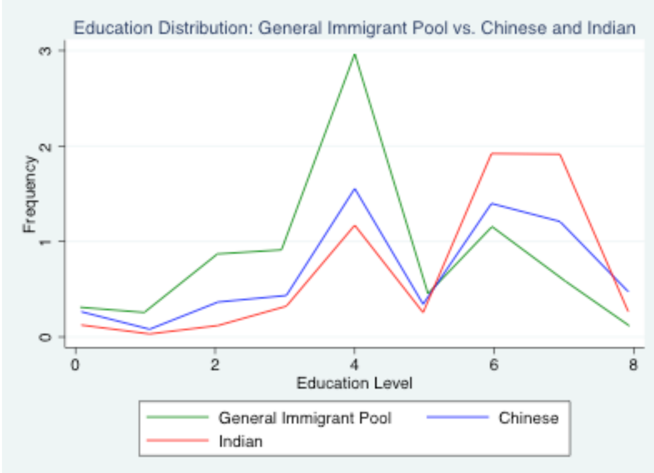
Figure 8: Persons Naturalized by Country of Origin, 2002-2011



Source: *Department of Homeland Security,* Yearbook 2011 Statistics, "Persons Naturalized by Region and Country of Birth: Fiscal Years 2002 to 2011"

Many reports suggest that Chinese and Indian immigrants are endowed with more human capital vis-à-vis other immigrant groups (Pew Research Social and Demographic Trends, 2012; Nowrasteh, 2012). The following graphs assess this assertion using a sample from the ACS 2011 5%. In these graphs, "Chinese" and "Indian" respectively refer to individuals who were born in Mainland China or India. "General Immigrant Pool" refers to all individuals born outside of the United States, excluding those born in China or India. The measures of human capital subsequently explored are education, English ability, occupational prestige, and income.

Figure 10: Education: General Immigrant Pool vs. Chinese and Indian Foreign Born

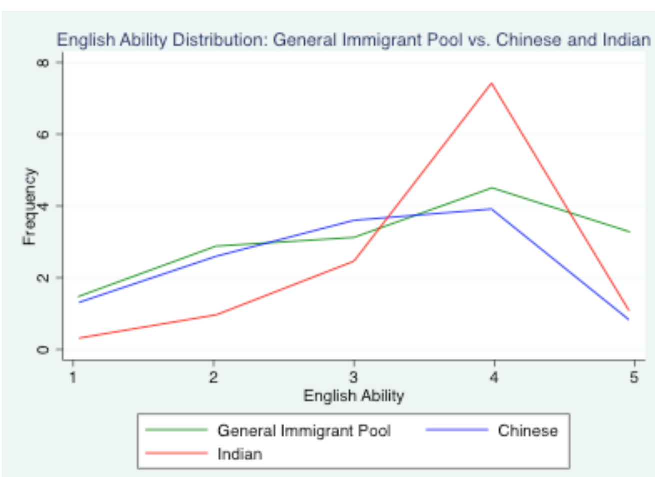


Level	Meaning
0	No education
1	Nursery – Grade 4
2	Middle School
3	Some high school
4	High school degree
5	Associate's degree
6	Bachelor's degree
7	Master's degree
8	PhD

Source: *IPUMS USA 5%,* American Community Survey 2011

Figure 10 clearly shows that the Chinese and Indian groups have many more highly educated (i.e. bachelor's degree and beyond) immigrants compared to the rest of the immigrant pool. Concurrently, there are less Chinese and Indian immigrants who are low skilled (i.e. high school degree and lower) compared to the overall immigrant pool. These results are a product of the 1965 Hart-Celler Act, which encouraged the critical mass of high-skilled workers in China and India to migrate through immigrant and non-immigrant channels.

Figure 11: English Ability Distribution: General Immigrant Pool vs. Chinese and Indian

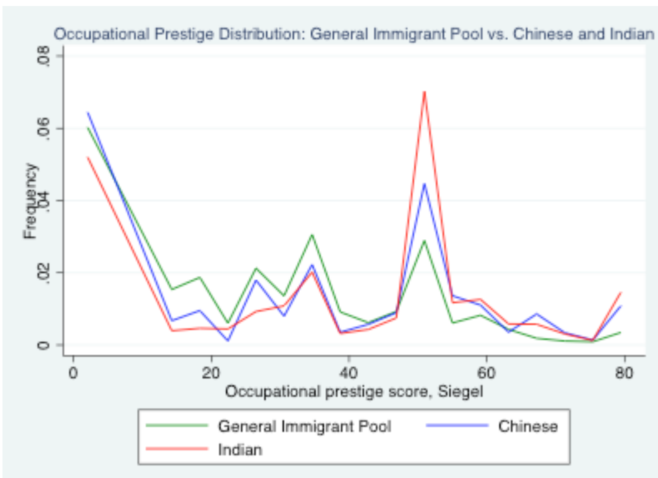


Level	Meaning
1	Speaks no English
2	Speaks English, but poorly
3	Speaks English well
4	Speaks English very well
5	Speaks only English

Source: *IPUMS USA 5%,* American Community Survey 2011

Moving to English ability, Figure 11 clearly illustrates that Indians far exceed the general immigrant pool in their English ability, while the Chinese mirror the general pool. It should be noted that low frequencies for level 5 are to be expected, since India and China are home to a multitude of other languages, and immigrants are likely to speak other languages while still being fluent in English.

Figure 12: Occupational Prestige Distribution: General Immigrant Pool vs. Chinese and Indian[16]

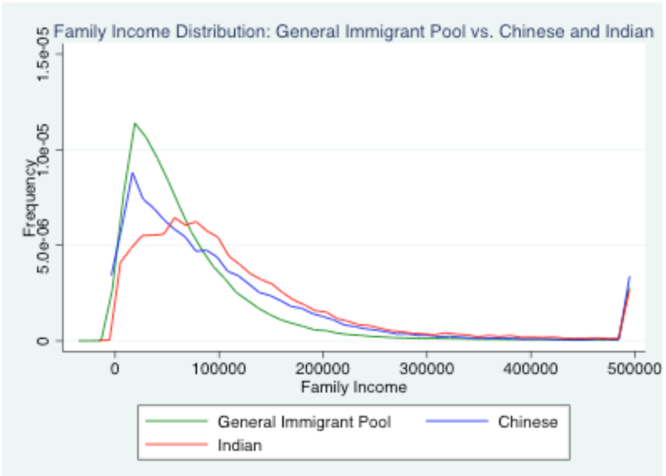


Source: *IPUMS USA 5%,* American Community Survey 2011

The occupational prestige score used in Figure 12 is the Siegel Prestige score measure, based on a 1960s evaluation conducted by the National Opinion Research Center. The distribution of Chinese and Indian professionals seems to roughly mirror the distribution of the general immigrant pool, with exceptions in the 15-20 and 45-60 ranges. The 15-20 ranges represent professions such as attendants, laundry and dry cleaning professionals, vehicle washers, construction workers, garbage collectors, etc. The 45-60 ranges correspond with professions such as engineers, health practitioners, computer specialists, accountants etc. (General Social Survey, 2012).

[16] For a detailed breakdown of prestige classifications, consult: http://publicdata.norc.org/GSS/DOCUMENTS/BOOK/GSS_Codebook_AppendixF.pdf

Figure 13: Family Income Distribution: General Immigrant Pool vs. Chinese and Indian



Label	Meaning
500,001	All income > $500,001

Source: *IPUMS USA 5%,* American Community Survey 2011

Figure 13 shows that Chinese and Indian immigrants earn more than the general immigrant pool, given their higher frequency rate in the $100,000 + income bracket. Concurrently, there are fewer Chinese and Indian immigrants that have a family income below $75,000.

The data support the notion that Chinese and Indian immigrants do on average have a higher level of human capital than the general immigrant pool. Indians particularly score the best, while Chinese immigrants perform better but also more closely resemble the general immigrant pool compared to Indians. Since human capital is positively associated with the propensity and ability to naturalize, one would accordingly expect naturalization rates to be higher for Chinese and Indian immigrants.

Chapter 3: Assimilation Literature

This chapter provides an overview of the literature focused on (a) the influence of ethnic concentrations on assimilation outcomes, such as wage convergence and language acquisition, and (b) the overall determinants of naturalization.

Chapter 3a: Ethnic Concentrations and Assimilation

Economists have traditionally been interested in immigrant labor market outcomes as a metric of assimilation. In their studies, many have found ethnic concentrations to be a significant determinant, but the direction of this variable's effect has been contested.

Certain studies illustrate a trade-off between the cultural comfort of an enclave and earnings (Gonzalez 1998; Borjas, 2000; McManus, 1990), while other studies illustrate that immigrants in enclaves are more likely to make human capital investments that improve their economic livelihood (Aydemir 2009). Another metric of interest is language acquisition, since the ability to speak English opens doors to more employment opportunities for immigrants. Results consistently show that living in concentrated enclaves in America is associated with a lower likelihood of attaining proficiency in English, particularly among Hispanic immigrants (Chiswick 1992; Beckhuser et al, 2012).

Studies emphasize that the influence of ethnic concentrations vary according to the profile of both the immigrant and neighborhood. Borjas (2000) stresses the importance of neighborhood quality (i.e. average human capital) in particular, finding that immigrants in poor quality neighborhoods have worse assimilation outcomes than those in higher quality neighborhoods. Moreover, Cutler and Vigdor (2007) find that the impact of segregation on certain outcomes, such as English ability or level of idleness, is positive for well-educated immigrants but negative for less-educated ones. A key issue underpinning these studies is endogeneity, since immigrants select into certain neighborhoods and not into others. Some studies exploit natural experiments through refugee resettlement programs to tackle this issue (Aydemir, 2009; Edin et al, 2003). Other studies develop instruments based on average neighborhood characteristics or include regional fixed effects (Beckhuser et al, 2012; Cutler and Vigdor, 2007).

On the whole, this body of literature has illustrated the significant but ambiguous role ethnic neighborhoods play in economic assimilation outcomes, and it sets the stage for analyzing the role of ethnic concentrations in naturalization outcomes.

Chapter 3b: Naturalization Literature Overview

In this section, I review the existing body of naturalization literature and emphasize the most oft-cited papers. I organize my review according to both the chronology of these studies as well as the main types of explanatory variables explored. These variables are: (1) human capital and personal characteristics, (2) country of origin characteristics, (3) family and home ownership characteristics, and (4) location and ethnicity characteristics. I include a table following each of these sections that summarize the findings, data used, and modeling technique employed by each study.

Human Capital & Personal Characteristics

Human capital, or the skills and experience of an individual, is the oldest type of variable to be studied in the naturalization literature. The most common measures of human capital are English ability and educational background. Age and years spent in the United States additionally proxy for immigrants' level of experience. Furthermore, occupational prestige and income levels can be viewed as manifestations of an individual's level of human capital. Lastly, though not a measure of human capital, gender is another personal attribute introduced in later studies of naturalization. We expect higher levels of human capital to be positively correlated with naturalization rates.

Sociologists Gavit (1922) and Bernard (1936) were two of the earliest social scientists to formally study naturalization outcomes. Due to data constraints, their studies draw conclusions from cross-tabular analyses rather than regression analyses. At the time of their work, immigration policies were highly racialized and contrasts were constantly drawn between "old" (i.e. Western Europe) and "new" (i.e. Southern or Eastern Europe) immigrants. Both authors use human capital as a way of illustrating that "old" immigrants are not innately superior due to their race or ethnicity: rather, since naturalized immigrants exhibit higher levels of income, English ability, and occupational prestige, they argue that naturalization outcomes are more closely related to human capital rather than racial attributes. Indeed, Bernard (1936) demonstrates that these human capital factors explain much of the within group variation in naturalization outcomes. In the latter half of the twentieth century, studies continued to employ the same measures of human capital but conducted regression rather than cross-tabular analyses. These papers study human capital on an individual level rather than group level.

The literature as a whole indicates that human capital is positively correlated to naturalization rates. This result is consistent across different datasets and modeling techniques. For example, studies have modeled years in the United States as dummies for every five-year interval (Jasso and Rosenzweig, 1986; Yang, 1994; Liang, 1994; DeVoretz and Pivnenko, 2005) or in continuous terms (Evans, 1988; Chiswick and Miller, 2008; Duncan and Waldorf, 2009). Additionally, education has been modeled as a series of dummy variables (Bernard, 1936; Liang, 1994; DeVoretz and Pivnenko, 2005) and continuous terms

(Chiswick and Miller, 2008; Yang, 1994). English ability has been modeled as a dummy variable according to whether an immigrant speaks it well (Duncan and Waldorf, 2009; Evans, 1988), multiple levels of fluency (Yang, 1994; Liang, 1994), and a unique Knowledge of English exam (Portes and Curtis, 1987). Lastly, income is modeled with the greatest degree of diversity, for every study measures it in a different way. Throughout these variations, human capital indicators have retained a significantly positive coefficient and thus appear to be robust to several specification styles.

Two exceptions, however, exist: Yang (1994) finds that additional years of education after high school reduce the odds of naturalization; he suggests that this result is a product of highly educated immigrants being more aware of prejudices and discrimination against minorities in U.S. society. Using data from the New Immigrant Survey to study intent to naturalize, Massey & Akresh (2006) find that immigrants with a higher level of income are less likely to want to naturalize. In contrast to Yang, they argue that since returns to skills and education have been increasing around the world, immigrants with a high degree of human capital are less likely to naturalize because "the grass may well seem greener in other national pastures."

Some studies show insignificant results for one or more of these variables. For example, Portes and Curtis (1987) only find a significantly positive correlation with English ability. However, much of this divergence can be attributed to their unique panel dataset (i.e. surveys they conducted on Mexican immigrants) as well as the short time span analyzed (1972- 1979). For other studies, collinearity issues rather than peculiarities in modeling technique or dataset, likely explain this outcome.

In terms of gender, females have a higher propensity to naturalize than males (Grebler, 1966; Jasso & Rosenzweig, 1986; Chiswick and Miller, 2008; Yang, 1994; Liang, 1994; Duncan and Waldorf, 2009).

Figures 10a summarizes these studies and findings on the influence of human capital.

Figure 10a: Human Capital and Personal Characteristics: Experience & Gender				
Variable	**Relationship**	**Authors**	**Data**	**Modeling**
Years in USA	Positive	Grebler (1966)	INS, 1959-1965, focused on Mexicans	Cross-tabular
		Gavit (1922)	INS, 1913-1914; 26,284 petitions	Cross-tabular
		Bernard (1936)	1930s sample of foreign-born: New Haven, CT	Cross-tabular
		Jasso & Rosenzweig (1986)	1970 PUMS Census	Dummy for 5 year buckets
		Yang (1994)	1980 PUMS Census	Dummy for 5 year buckets
		Liang (1994)	1980 PUMS Census	Dummy for 5 year buckets
		DeVoretz & Pivnenko (2005)	1996 Census of Canada	Dummy for 5 year buckets
	Positive and diminishing	Chiswick & Miller (2008)	2000 Census	Squared term
		Duncan & Waldorf (2009)	2005 ACS 5%	Squared term
		Evans (1988)	1981 Australian Census	Parametric depictions
	None	Portes & Curtis (1987)	1973-1979 unique panel study of Mexican migrants	Continuous
Age	Positive	Chiswick & Miller (2008)	2000 Census	Continuous
		Massey & Akresh (2006)	New Immigrant Survey (2002)	Continuous
		DevVoretz & Pivnenko (2005)	1996 Census of Canada	Continuous
	Positive and diminishing	Yang (1994)	1980 PUMS Census	Squared Term
	Negative	Jasso & Rosenzweig (1986)	1970 PUMS Census	Continuous
		Liang (1994)	1980 PUMS Census	Continuous
	None	Grebler (1966)	INS, 1959-1965, focused on Mexicans	Cross-tabular
		Portes & Curtis (1987)	1973-1979 panel study of new Mexican migrants	Continuous
Female	Positive	Grebler (1966)	INS, 1959-1965, focused on Mexicans	Dummy variable
		Jasso & Rosenzweig (1986)	1970 PUMS Census	
		Chiswick & Miller (2008)	2000 Census	
		Yang (1994)	1980 PUMS Census	
		Liang (1994)	1980 PUMS Census	
		Duncan & Waldorf (2009)	2005 ACS 5%	

	None	Evans (1998)	1981 Australian Census	

*SEI score= Socioeconomic Index Score= a composite of occupational prestige, income, and education

Figure 10a: Human Capital: Common Measures and Manifestations

Education	Positive	Gavit (1922)	INS, 1913-1914; 26,284 petitions	Dummy variables per education level.
		Bernard (1936)	1930s sample of New Haven, CT	Dummy variables per education level
		Chiswick & Miller (2008)	2000 Census	Continuous
		Liang (1994)	1980 PUMS Census	Dummy per education level
		DeVoretz & Pivnenko (2005)	1996 Canadian Census	Dummy variables for bachelor's, bachelor's plus, and PhD.
	Positive, then negative after h.s.	Yang (1994)	1980 PUMS Census	Squared Term
	None	Portes & Curtis (1987)	1973-1979 panel study of new Mexican migrants	Continuous
		Evans (1988)	1981 Australian Census	Squared term
		Massey & Akresh (2006)	New Immigrant Survey (2002)	Dummy variables per education level
Occupation	Positive	Grebler (1966)	INS, 1959-1965, focused on Mexicans	Cross-tabular buckets
		Bernard (1936)	1930s sample of New Haven, CT	Cross-tabular buckets by skill and occupation
		Yang (1994)	1980 PUMS	SEI index*
		Liang (1994)	1980 PUMS Census	SEI index
	None	Portes & Curtis (1987)	1973-1979 panel study of new Mexican migrants	Dummy variable (skilled=1)
English ability	Positive	Portes & Curtis (1987)	1973-1979 panel study of new Mexican migrants	KEI scores- continuous
		Yang (1994)	1980 PUMS	Dummy per level of English
		Liang (1994)	1980 PUMS Census	Dummy per level of English
		Duncan & Waldorf (2009)	2005 ACS 5%	Dummy (speaks well=1)
		Evans (1988)	1981 Australian Census	Dummy (speaks only English at home=1)
	None	Duncan & Waldorf (2009)	2005 ACS 5%	Dummy (speaks at home=1)
		Massey & Akresh (2006)	New Immigrant Survey (2002)	Dummy (speaks well=1)
Income	Positive	Grebler (1966)	INS, 1959-1965, focused on Mexicans	Cross-tabular earnings
		Gavit (1922)	1913-1914 court petitions	Weekly earnings cross tabular
		Duncan & Waldorf (2009)	2005 ACS 5%	Continuous % of poverty threshold
	None	DeVoretz & Pivnenko (2005)	1996 Census of Canada	Logged earning

		Evans (1988)	1981 Australian Census	Family total income
	Negative	Massey & Akresh (2006)	New Immigrant Survey (2002)	Years of U.S. income

Country of Origin Characteristics

While Gavit (1922) and Bernard (1936) challenged the old-new dichotomy, the passage of the 1965 Hart-Celler Act made this approach obsolete. Nonetheless, differences in national origin still matter, because certain attributes of origin countries may uniformly affect the propensity of different immigrant groups to naturalize. Studies have only recently begun to consider such attributes: the earliest paper to do so dates back to 1994. With the exception of one paper, all studies that have included a characteristic of some sort have found it to be statistically significant, though the direction of significance may be disputed. Apart from the

Country of origin factors that are positively correlated to naturalization include distance between origin country and the United States (Jasso and Rosenzweig, 1986; Chiswick and Miller, 2008), absence of civil liberties or presence of socialist regimes (Jasso and Rosenzweig, 1986; Yang, 1994; Chiswick and Miller, 2005), dual citizenship (Jones-Correa, 2001; Chiswick and Miller, 2008; Mazzolari, 2009), and coming from an English-speaking country (Jasso & Rosenzweig, 1986; Chiswick and Miller, 2008).

Though Yang (1994) finds a negative relationship between dual citizenship and naturalization, theory and empirical evidence provide no justification for this finding. He also finds a negative relationship between English-speaking country and naturalization outcomes but once again is unable to justify such a result. A country of origin factor that is clearly negatively correlated to naturalization outcomes is national income (Jasso and Rosenzweig, 1986; Yang, 1994; Chiswick and Miller, 2005).

Rather than identifying specific origin country attributes, other studies have opted to employ country fixed effects instead (Yang, 1994; Liang, 1994; DeVoretz and Pivnenko, 2005; Duncan and Waldorf, 2009; Massey & Akresh, 2006), finding significance in some dummies and none in others.

Figure 10b summarizes these studies and findings on the effects of origin country characteristics.

Figure 10b: Country of Origin Characteristics				
Variable	**Relationship**	**Authors**	**Data**	**Modeling**
Income	Negative	Jasso & Rosenzweig (1986)	1970 PUMS Census	GNP/capita
		Yang (1994)	1980 PUMS	
		Chiswick & Miller (2008)	2000 Census	GDP/capita
Distance	Positive	Jasso & Rosenzweig (1986)	1970 PUMS Census	Continuous (in terms of miles)
		Chiswick & Miller (2008)	2000 Census	
Socialist political regime/ Absence of Civil Liberties	Positive	Jasso & Rosenzweig (1986)	1970 PUMS Census	Dummy (1=Centrally planned economy)
		Yang (1994)	1980 PUMS	Dummy (1= Socialist/refugee sending countries)
		Chiswick & Miller (2008)	2000 Census	Index based on level of civil and political liberties
Literacy	Positive	Jasso & Rosenzweig (1986)	1970 PUMS Census	Continuous
English-speaking	Positive	Jasso & Rosenzweig (1986)	1970 PUMS Census	Dummy (1= English-speaking country)
	Mixed	Chiswick & Miller (2008)	2000 Census	
	Negative	Yang (1994)	1980 PUMS	
Dual Citizenship	Positive	Chiswick & Miller (2008)	2000 Census	Dummy (1= dual citizen permitting country)
		Jones-Correa (2001)	INS & Census 1965 – 1997	
		Mazzolari (2009)	1990 and 2000 Census	
	Negative	Yang (1994)	1980 PUMS	
Country Fixed Effect	None	Duncan & Waldorf (2009)	2005 ACS 5%	Country fixed effect variables
	Mixed	DeVoretz & Pivnenko (2005)	1996 Census of Canada	
		Massey & Akresh (2008)	New Immigrant Survey (2002)	
	Significant	Liang (1994)	1980 PUMS Census	

Family Characteristics and Home Ownership

As a result of data constraints, the early twentieth century studies could not include family characteristics and home ownership in their analysis. Specific variables include whether the immigrant has a spouse, if the spouse is an American citizen, whether the immigrant has children, and whether the immigrant owns a home. These variables are predicted to positively correlate with naturalization rates, because they anchor the immigrants in the United States. The results on the whole fall in line with these predictions.

Having a spouse increases the probability of naturalization (Jasso and Rosenzweig, 1986; Yang, 1994; Liang, 1994; Duncan and Waldorf, 2009), having a spouse who is a U.S. citizen is an even stronger predictor (Portes and Curtis, 1987; Liang, 1994; Chiswick and Miller, 2008), having children increases the probability of naturalization (Portes and Curtis, 1987; Liang, 1994; Yang, 1994; DeVoretz and Pivnenko, 2005), and lastly home ownership is positively correlated as well (Portes and Curtis, 1987; Yang, 1994; Liang, 1994; DeVoretz and Pivnenko, 2005).

The use of non-U.S. Census data rather than modeling technique explains the results that deviate from the aforementioned findings. Indeed, most of the variables are modeled very similarly as dummies. Exceptions include Jasso and Rosenzweig (1986), who model having a spouse as part of the multiplier effect of family reunification. In addition, children are modeled based on the number of children (Portes and Curtis, 1987; Liang, 1994; Massey and Akresh, 2008), if the immigrant has any children at all (Yang, 1994), and an interaction term with marriage (DeVoretz and Pivenko, 2005).

Figure 10c below summarizes these studies and findings on the effects of family characteristics and home ownership.

Figure 10c: Family Characteristics and Home Ownership				
Variable	**Relationship**	**Authors**	**Data**	**Modeling**
Spouse	Positive	Jasso & Rosenzweig (1986)	1970 PUMS Census	Effect as a family reunification multiplier
		Yang (1994)	1980 PUMS	Dummy (married=1)
		Liang (1994)	1980 PUMS Census	
		Duncan & Waldorf (2009)	2005 ACS 5%	
	None	Chiswick & Miller (2005)	2000 Census	
		Portes & Curtis (1987)	1973-1979 panel study of new Mexican migrants	
	Negative	DeVoretz & Pivenko (2005)	1996 Census of Canada	
		Massey & Akresh (2006)	New Immigrant Survey (2002)	
Spouse_U.S.	Positive	Portes & Curtis (1987)	1973-1979 panel study of new Mexican migrants	Dummy (1=married to US citizen)
		Liang (1994)	1980 PUMS Census	
		Chiswick & Miller (2005)	2000 Census	
Children	Positive	Portes & Curtis (1987)	1973-1979 panel study of new Mexican migrants	Number
		Yang (1994)	1980 PUMS	Dummy (1=if any)
		Liang (1994)	1980 PUMS Census	Number
		DeVoretz & Pivenko (2005)	1996 Census of Canada	Interacted with marriage
	None	Massey & Akresh (2008)	New Immigrant Survey (2002)	Number
Home Ownership	Positive	Portes & Curtis (1987)	1973-1979 panel study of new Mexican migrants	Dummy (1= home owner)
		Yang (1994)	1980 PUMS	
		Liang (1994)	1980 PUMS Census	
		DeVoretz & Pivenko (2005)	1996 Census of Canada	
	None	Evans (1988)	1981 Australian Census	
	Negative	Massey & Akresh (2006)	New Immigrant Survey (2002)	

Location and Ethnic Characteristics

Location and ethnic characteristics are the newest type of variable to enter the naturalization literature. An immigrant's decision to naturalize does not occur in a vacuum and may be significantly conditioned by their contextual surroundings. As such, co-ethnics can shape immigrant naturalization outcomes by influencing their desire and ability to do so. Although the relationship may be significant, its direction is ambiguous.

Some studies find a negative relationship between ethnic concentration and naturalization outcomes. Portes and Curtis (1987) were the first to include a measure of neighborhood traits in their analysis, and they specifically examine the influence of neighborhoods that are "Anglo" vs. "not Anglo".[17] Based on a unique panel data set, they find that immigrants from Anglo neighborhoods are less likely to *want to* become citizens, but they are more likely to *actually* become citizens when eligible. They hypothesize that this change of heart is due to the acquisition of place-specific human capital. Chiswick's and Miller's (2008) study compares states with varying minority language concentrations and similarly finds that immigrants who live in states where many others share their mother tongue are less likely to naturalize.

On the other hand, Yang (1994) finds a positive relationship between co-ethnic concentrations and naturalization outcomes. Her co-ethnic concentration measure, however, is on a national level and detached from any neighborhood level of analysis. Nonetheless, she does analyze urban vs. rural residencies and finds that immigrants in urban areas have a higher propensity to naturalize. Since urban areas are highly populated by the foreign born, this finding adds some spatial context to the discussion of ethnic influences.

Other studies show evidence of a mixed but significant relationship between ethnic neighborhoods and naturalization outcomes. Liang (1994) studies the effects of co-ethnics through a social contact and social capital framework. She examines the degree to which immigrants may interact with the native-born using the P* interaction index: this index is calculated based on ethnic information on census tracts aggregated at the metropolitan area level. She finds that higher co-ethnic concentrations increase the

[17] Their study does not explicitly define the parameters of the neighborhood or how exactly the Anglo determination is made.

propensity to naturalize for Chinese immigrants but reduce the propensity for Mexican, Cuban, Colombian, and Korean immigrants.

Duncan and Waldorf (2009) employ PUMAs as their spatial unit of analysis and analyze the naturalization patterns of Caribbean immigrants in the New York tri-state area. They create neighborhood (i.e. PUMA) variables for (a) the *number* (as opposed to percentage) of co-ethnics in a PUMA and (b) the proportion of immigrants naturalized within a PUMA. They find that well-assimilated (i.e. neighborhoods with a greater proportion naturalized) PUMAs increase the probability of naturalization and that the ethnic concentration has a positive effect in well-assimilated enclaves and a negative effect in poorly assimilated neighborhoods, echoing Borjas (2000). They further find that the relationship between well-assimilated enclaves and the propensity to naturalize are stronger for highly educated immigrants more so than for poorly educated ones.

Several factors prompt these contrasting results. First, each study focuses on different immigrant groups: Liang (1994) and Duncan and Waldorf (2009) only focus on a few groups, while the balance of the studies analyze the whole foreign born population. Second, each study employs a different spatial unit of analysis, ranging from none at all (Yang, 1994), to the state (Chiswick and Miller, 2008), the metropolitan area (Liang, 1994), the Public Use Microdata Area (Duncan and Waldorf, 2009), and a non-Census definition of neighborhood (Portes and Curtis, 1987). Third, concentrations of these spatial units are measured in different ways as well. Measures include co-ethnic concentrations (Yang, 1994; Duncan and Waldorf, 2009), native-born vs. foreign-born dominant neighborhoods (Portes and Curtis, 1987), minority language concentration (Chiswick and Miller, 2008), and interaction indexes as opposed to general concentration levels (Liang, 1994).

Figure 10d summarizes these studies and their findings.

Figure 10d: Location and Ethnic Characteristics				
Variable	**Relationship**	**Authors**	**Data**	**Modeling**
Ethnic Stock	Positive	Yang (1994)	1980 PUMS	Ln (size of 1975 immigrant ethnic community)
Urban	Positive	Yang (1994)	1980 PUMS	% urban population
Enclave	Negative	Portes & Curtis (1987)	1973-1979 panel study of new Mexican migrants	Dummy (1=anglo)
		Chiswick & Miller (2008)	2000 Census	Concentration of minority language in state
	Mixed	Duncan & Waldorf (2009)	2005 ACS 5%	Co-ethnic concentration in PUMA; % of PUMA that has naturalized
		Liang (1994)	1980 PUMS Census	P* index based on census tract and metro area by Massey and Denton (1987)*

***** P* = interaction index

Conclusions

As a whole, human capital, countries of origin, and family characteristics have a consistent effect throughout the literature with a few exceptions. On the other hand, the effect of location and ethnic concentrations are much more ambiguous and volatile depending on the immigrant group in question, spatial unit of analysis, and definition of ethnic concentration.

I contribute to the current pool of studies by providing a more intensive focus on the location and ethnic characteristic variable. Specifically, I seek to interrogate the mechanisms behind the ethnic neighborhood effect, since the current studies tend to justify their results on an ad-hoc basis. Additionally, I offer a more granular analysis of neighborhoods by leveraging information on the census tract, PUMA, and metropolitan area level.

I lastly contribute to the naturalization literature by studying time to naturalization in addition to probability of naturalization. This analysis is made possible by leveraging new information on the year an immigrant naturalizes.

Chapter 4: Studying Neighborhood Effects

Chapter 3 demonstrates that there is a high degree of ambiguity surrounding the role of ethnic neighborhoods in assimilation outcomes. I argue that much of this ambiguity is prompted by (a) definitional challenges and (b) differences in data and empirical strategies employed in each study. In this chapter, I demonstrate that issues of selection bias, endogeneity, and correlated effects additionally complicate our understanding of ethnic neighborhood effects.

Figures 9a and 9b show that immigrants are not randomly spread across the United States. Since immigrants can choose the neighborhoods in which they want to live, selection bias is a critical concern in the study of neighborhood effects. Naturalization raises less endogeneity concerns in comparison to other assimilation metrics. For example, immigrants may choose to live in certain neighborhoods based on the wages they can earn, but they are not as likely to choose neighborhoods based on naturalization criteria. However, they still may select into neighborhoods based on the mechanisms that promote naturalization.[18]

One mechanism is network benefits: immigrants may choose to live among co-ethnics to avail themselves of connections to opportunities that improve their assimilation prospects. For example, Andersson et al. (2009) conclude that ethnic enclaves play a positive role in improving employment outcomes and earnings for new immigrants, and that enclave members disproportionately tend to work with other enclave members. From a sociological lens, Portes (1987) also illustrates how the formation of ethnic institutions such as the Latin Chamber of Commerce and Latin Builder's Association in Miami encourage social and economic networking that consequently draw Cuban immigrants into certain areas of Miami. Network benefits typically promote economic assimilation, but since the same networks likely promote other assimilation outcomes, such as naturalization, endogeneity concerns are valid.

A second mechanism at work is cultural amenities: immigrants may derive a certain utility from living with co-ethnics because of shared culture and preferences as well as access to tangible ethnic goods and services. Gonzalez (1998) underscores the importance

[18] I discuss these mechanisms in greater detail in Chapter 5

of cultural amenities by arguing that immigrants sacrifice higher earnings and lower land rents to live in areas with a sizable Mexican population. Chiswick and Miller (2005) specify what these cultural amenities may be and introduce the concept of "ethnic goods." Ethnic goods are a set of market and non-market goods and services that can range from ethnic grocery stores to marital markets. The higher the ethnic concentration, the more prevalent and lower the cost of these ethnic goods. Consequently, immigrants may choose to live in areas with high ethnic densities to maximize their ethnic consumption.

However, since immigrants may not always be able to perfectly sort themselves into a neighborhood of their choosing, ethnic neighborhoods may still exert an exogenous amenities effect. Specifically, immigrants who live in neighborhoods that cater to their ethnic consumption preferences may have a greater degree of satisfaction with life in the United States and thus be more willing to permanently settle and naturalize. The exogenous and endogenous components of the amenities effect thus illustrate the problem endogeneity poses in detecting causal relationships.

Correlated effects are another issue that leads to biased estimates. In the context of this paper, this issue rises when ethnic composition of the neighborhood is highly correlated to factors that are important in determining naturalization outcomes. To identify possible correlated factors, I turn to the literature on immigrant location choice. Immigrants, particularly those who are on employment-based visas, select locations based on the employment conditions of the metropolitan area. Such conditions include the type of industry and its labor demand (Hall et al, 2011), general wage levels and unemployment rates (Jaeger, 2002), and the education level and availability of high-paying jobs (Scott et al, 2005). Employment correlations could be particularly concerning for metropolitan areas such as New York or San Francisco. Both areas have a high density of Asians as well as a growing number of attractive high skilled jobs; as a result, Chinese and Indian immigrants in these areas may want to naturalize because of job satisfaction rather than ethnic concentrations. Including a metropolitan area dummy and clustering standard errors by metropolitan area may easily remedy such a correlated effect.

Certain studies emphasize the correlation between low levels of human capital and residence in ethnic neighborhoods. Bartel (1989) shows that the ethnic concentration of metropolitan areas is a less significant pull factor for immigrants with higher levels of education. Concurrently, Borjas (1998) argues that low levels of income, parental skills, and ethnic capital silos immigrants into ethnic enclaves. Poor English ability is another important factor forcing immigrants to live near co-ethnics (Bauer et al., 2005; Chiswick et al, 2001).

Nonetheless, while these studies speak for the overall immigrant pool at large, studies that focus on the Asian subset illustrate that the human capital story is not as clear-cut. For instance, in his study of residential segregation in San Francisco, Bayer (2003) found that factors such as education, income, language, and immigration status only explained 50% of segregation for Asians, while they explain 95% of segregation for Hispanic households. Though his study is not conducted on a neighborhood level, Kerr (2010) also illustrates that 81% of Chinese and Indian ethnic inventors are increasingly concentrating in major metropolitan areas compared to 73% of the general inventor pool. Lastly, Li (1998) uses the San Gabriel Valley, CA as a case study to illustrate that Asian densities (30% and higher) are persisting outside of central cities and in affluent suburbs. As a result, the classic human capital explanation of ethnic enclaves is not as applicable for contemporary Asian immigrants.

Chapter 3 established human capital as a critical vector of variables explaining naturalization outcomes. However, since the literature indicates that a range of human capital levels may define ethnic concentrations, Chinese and Indian concentrations are not necessarily highly correlated with human capital. Concerns of correlated effects can further be addressed through the addition of neighborhood "quality" variables in regressions.

This chapter has demonstrated the difficulty inherent in studying causal relationships between ethnic neighborhoods and naturalization. The concerns of correlated effects are mitigated by (a) the inclusion of neighborhood quality variables, (b) evidence of mixed relationships between human capital and naturalization outcomes, and (c) the addition of metropolitan level dummies to control for unobservable differences between metropolitan areas. However, there may be other unobserved properties of ethnic neighborhoods that may still bias the ethnic neighborhood estimate. Moreover, the endogeneity between ethnic neighborhoods and the mechanisms prompting naturalization is a material concern and hard to avoid without an adequate instrument. As a result, the relationships between ethnic neighborhoods and naturalization very much represent correlations rather than causality.

Chapter 5: Theory

Under a rational choice framework, an immigrant will choose a state, citizen or non-citizen, which maximizes his or her utility. Utility may be driven by both pecuniary and non-pecuniary factors; more specifically, immigrants may choose a state that not only maximizes their income and economic well being but also their political and social opportunities. Additionally, since neither China nor India permit dual citizenship, it is important that the utility of being an American citizen net of naturalization costs exceeds that of being both a legal permanent resident as well as an origin country citizen. An immigrant will thus naturalize if:

$$U(E_c, P_c, S_c) - C > U(E_{lpr}, S_{lpr}) + U(E_{bpl}, P_{bpl}, S_{bpl})$$

In this model, *c, lpr,* and *bpl* respectively represent three states: American citizen, American legal permanent resident, and citizen of place of birth (i.e. China or India). *E, P,* and *S* respectively refer to the economic, political, and social benefits of the different states, and *C* refers to both the monetary and non-monetary costs specific to the naturalization process.

As non-citizens, immigrants to the United States enjoy benefits both as legal permanent residents and as foreign nationals of their country of origin. First, as foreign nationals, immigrants still retain their right to vote in, own property in, and easily return to their country of origin. Second, as legal permanent residents, immigrants face minimal employment restrictions and are entitled to a range of public benefits including Social Security and Medicare. However, the price of these benefits is double taxation both in the country of origin as well as in the United States.

As naturalized citizens, immigrants enjoy an array of additional economic, political, and social benefits. On an economic level, several studies have documented that immigrants who acquire citizenship benefit from higher levels of wages and economic assimilation (Bratsberg, Ragan, Nasir, 2002; Pastor and Scoggins, 2012; DeVoretz and Pivenko, 2006). This assimilation is partially driven by access to more jobs in the public sector, law enforcement, and defense domains. The economic benefits were much more salient in the 1960s, when many more professions (81 specifically) were restricted to citizens, such as accounting, architecture, and dentistry, among others (Sanders, 1968).

Additionally, employers may be biased towards hiring citizens, since their investment in employees will have a higher payoff if they can be sure an immigrant plans to stay in the United States for the long term. Secondly, citizenship grants the poorest segment of immigrants full access to various welfare schemes and benefits. Lastly, immigrants are no longer required to pay taxes in their country of origin.

Immigrants have no political rights as legal permanent residents, but once naturalized, they can exercise their right to vote and can enjoy the recognition and full rights accorded to all Americans under the United States Constitution. However, for Indian and Chinese nationalities, these rights come at the price of losing their political rights in their country of origin. A social benefit is that citizens can sponsor family members for legal permanent residency, including unmarried children and parents as well as married children and siblings, who they could not otherwise sponsor before becoming citizens. Moreover, the processing times of these requests are expedited when one is a citizen. Naturalized immigrants are also freed from the travel limits imposed on them as green card holders. As a result, citizens can spend more time in their country of origin if desired.

Nonetheless, there are significant costs associated with forfeiting origin country citizenship and naturalizing. In addition to modest fees for surrendering Chinese and Indian citizenships, naturalized citizens lose all political rights they once enjoyed in their country of origin and may face certain psychological costs born from formally breaking ties with their home country. As explained in Chapter 2a, these losses are more severe for Chinese immigrants than they are for Indian immigrants, since China does not have an emigrant-friendly visa.

In terms of monetary costs for naturalizing, the current naturalization fee is $680, though, as Chapter 2a illustrates, it has been as low as $100. In addition, since legal permanent residents must pay $450 every decade to renew their green card, in a way, immigrants who expect to permanently settle in the United States recover the naturalization fee within two decades. Other naturalization costs may include consulting fees for lawyers. Non-monetary costs of naturalization include the investment of time and human capital. The naturalization process can last up to a year (or up to three years for petitions filed in the 1990s) and requires filing an application, attending a biometrics appointment, preparing for and passing the civics and English exams, and taking an oath. Nonetheless, the human capital costs by way of exam preparation are likely negligible for the majority of the Chinese and Indian pool, since, as Chapter 2b illustrates, they are by and large well educated and proficient in English.

As established in Chapter 3, the extent or applicability of the costs and benefits detailed above vary according to the individual and his or her profile. Immigrants with a higher degree of human capital face lower costs than those with less. In particular, higher levels of education, language, income, and years spent in the United States likely lower the cost of naturalizing since immigrants are more ready and capable of handling the naturalization process. On a similar note, individuals with more anchors, such as home ownership, American spouses, and children, in the United States face higher costs of return migration and thus are also more likely to naturalize.

Neighborhoods with a high co-ethnic concentration can also play a crucial role in affecting the costs and increasing the benefits of naturalization. There are many competing hypotheses explaining neighborhoods' effect. One hypothesis is the "ethnic enclosure hypothesis," which suggests that "the more within-group interactions immigrants have, the more likely their ethnic identity will be reinforced and the less likely they will become citizens" (Liang, 1994). This paper advances the alternative "ethnic resilience hypothesis," which suggests on a general level that other immigrants can play a critical role in the adjustment process of new arrivals. I argue that ethnic neighborhoods mitigate costs and increase the benefits of naturalization in two manners.

First, following Venturini and Faini (2001), I posit that immigrants have a home bias. That is, they derive utility not only from the aforementioned economic, political, and social factors (*b*), but also from amenities specific to their country of origin (*f*). As a result, neighborhoods with a higher co-ethnic concentration may cater to this home bias by offering a number of ethnic goods and services (e.g. ethnic grocery stores, temples and religious circles, etc.) that serve the immigrant's unique consumption habits. This higher level of satisfaction may consequently increase the propensity of immigrants to formally settle and naturalize, an idea I will refer to as the "amenities hypothesis." The naturalization condition can hence be further expressed as:

$$U(b_{USA}, f_{USA}) > U(b_o, f_o)$$

where USA refers to the immigrants' state as an American citizen and O refers to their state as a citizen of their origin country. Nonetheless, as I explain in Chapter 4, the amenities hypothesis poses empirical challenges since immigrants who have strong preferences for ethnic amenities may, to some degree, be able to sort into neighborhoods that provide them.

In addition to the amenities hypothesis, neighborhoods may also affect naturalization outcomes due to social interactions, which provide the "social capital," referred to in the naturalization literature (Liang, 1994). Neighborhoods with a higher concentration of immigrants or co-ethnics likely have more informational resources to guide the immigrant through the naturalization process. These resources come in the form of relationships with friends and families who have naturalized, and, on a more formal level, the presence of non-profits specifically promoting immigrant integration. A simple payoff function of the decision to naturalize accounting for the influence of social interactions can be modeled as (Soetevant et al, 2007):

$$V(y_i, \mathbf{x}_i, \mathbf{y}_{-i}, \varepsilon_i(y_i)) = u(y_i, \mathbf{x}_i) + S(y_i, \mathbf{x}_i, \mathbf{y}_{-i}) + \varepsilon_i(y_i)$$

In this model, i refers to individual, y refers to the choice of naturalizing or not naturalizing, $\mathbf{x}$ is a vector of observable explanatory variables, and ε is a random error term. The first term accordingly accounts for private utility, the second term social utility, and the third random utility. As discussed in Chapter 4, the study of neighborhood social interactions also poses endogeneity and correlated effects concerns.

One of the challenges in studying neighborhood effects is that we do not know at what spatial scale interactions and amenities may be significant. Chapter 3 indicated that previous studies have focused on a range of geographies, from states, to metropolitan areas, to PUMAs. There are merits to examining neighborhood effects on all of these levels. On the one hand, all of the aforementioned areas are too big to really understand how day-to-day interactions with an immigrant's community may influence the individual's decision to naturalize. On the other hand, neighborhoods do not exist in vacuums, and it is important to consider the interaction of the neighborhood with the larger surrounding areas. For instance, an immigrant may not live in an ethnic cluster in San Francisco, but San Francisco's overall status as an immigrant gateway may still offer immigrants a certain degree of access to social capital and amenities. On the other hand, the role of an ethnic cluster in a state like Idaho may be larger, since immigrant culture, social capital, and amenities would be harder to overall come by. In addition to spatial unit considerations, qualities of neighborhoods must also be taken into account. Borjas (1992) advances the argument that poorer quality enclaves induce poor assimilation and higher quality enclaves promote better integration, where quality is a reflection of neighborhood levels of human capital.

This chapter highlights the potential role that human capital, commitment variables, and destination characteristics may play in shaping the cost-benefit analysis of a rational immigrant. In particular, I emphasize that the neighborhood variable must be carefully considered and constructed across many dimensions. First, to encompass a range of spatial unit sizes, I propose exploring ethnic concentrations on a metropolitan area, PUMA, and tract level. Since the PUMS 3% sample contains no tract level information, I will include a distributional measure reflecting tract characteristics for each PUMA. Second, I propose considering the quality of the neighborhood by including PUMA-level information on linguistic isolation, proportion of people below the poverty line, and proportion of foreign-born naturalized. This framework prompts the following hypotheses:

General Hypothesis:

1) Immigrants in PUMAs with high levels of co-ethnic concentrations will be more likely to have naturalized. This effect will be intensified when co-ethnics cluster together in the PUMA.

Selection Hypothesis

2) Differences across metropolitan areas may diminish the magnitude of the co-ethnic concentration coefficient, but hypothesis (1) should still hold.

Amenities Hypothesis

3) Hypothesis (2) should still apply to high-skilled immigrants, providing weak support for the amenities hypothesis. Since high-skilled migrants are more capable of navigating the naturalization process by themselves, they will not be influenced by the social capital mechanism.

Social Capital Hypothesis

4) The effect of ethnic concentrations on the propensity to naturalize will be diminished in linguistically isolated and poor neighborhoods and enhanced in neighborhoods with a high proportion naturalized. This effect will be particularly salient for low-skilled immigrants, who are more influence by social capital externalities.

Evidence of hypothesis (1) would provide support for the ethnic resilience hypothesis and would also highlight the importance of studying neighborhoods on a range of spatial levels. Hypothesis (2) controls for differences between metropolitan areas and eliminates some but not all selection effects. The social capital and amenities mechanisms behind this neighborhood effect cannot be entirely parsed, but hypotheses (3) and (4) offer some suggestive evidence of their importance to different classes of immigrants. Evidence of hypothesis (3) would offer some support for the amenities hypothesis, since high-skilled immigrants may not rely on social capital as much to navigate the naturalization process. Evidence of hypothesis (4) would offer some support for the social capital hypothesis, since low quality neighborhoods are characterized by low degrees of social capital.

Chapter 6: Data and Summary Statistics

I use data from the 3% sample of the 2011 American Community Survey (ACS) to test this paper's hypotheses. Conducted on an annual basis, the ACS surveys a random sample of American households each year on matters concerning demographics, housing, health, and migration. I create my sample through extracts from both the Integrated Public Use Microdata Series (IPUMS) database of the Minnesota Population Center as well as American Fact Finder. Specifically, I use the IPUMS data to construct individual-level variables, and I use tables from American Fact Finder to construct the neighborhood feature variables. These tables are "DP02: Selected Social Characteristics in the United States," "DP03: Selected Economic Characteristics," and "DP05: ACS Demographic and Housing Estimates."

The following characteristics shape the sample. First, I include only immigrants whose place of birth was India or Mainland China. Second, to ensure that citizenship was acquired autonomously rather than through parents, I restrict the sample to immigrants who migrated at the age of 18 or older.[19] Third, I limit the sample to immigrants who have been in the United States for either over five or seven years. I apply the five-year limit to immigrants who are married to a U.S. citizen, since these immigrants are eligible to naturalize after three years of having a green card. The seven-year limit is applied to the balance of the immigrant pool, since these immigrants must wait five years after having received their green card to naturalize. The additional two years on each limit are conservative estimates of the time it may take naturalization petitions to be processed and for immigrants to actually receive their green card. This selection criteria results in a sample size of 54,429 immigrants, composed of 26,674 Chinese and 27,755 Indians.

For my survival analysis model, I additionally limit the sample to immigrants

[19] There is no age at migration variable, so this limit is applied based on the difference between years in USA from the reported age.

who naturalized between 2008 and 2011. This restriction increases the likelihood that the information reported (e.g. income, language ability, neighborhood) in the ACS 2011 actually played a role in shaping the naturalization outcome. Since the pool of non-naturalized immigrants is the same both before and after this restriction, I randomly reduce the size of the non-naturalized pool by a proportional amount to maintain parity between the ratio of naturalized and non-naturalized citizens. This selection criteria results in a sample size of 7,681 immigrants, composed of 3,067 Chinese and 4,614 Indians.

Due to the large sample size, randomized collection, and extensive information provided on individual, household, and neighborhood specific characteristics, this dataset makes the countrywide analysis of local and individual impacts conducted in this paper feasible. Nonetheless, the IPUMS data does have a few notable limitations.

First, as noted in Chapter 2a, the Census Bureau does not collect data on the legal status of immigrants: thus, immigrants who are eligible to naturalize (i.e. green-card holders) cannot be distinguished from temporary visa holders or undocumented immigrants. I estimate the proportion of temporary visa holders to be 25% in the Indian sample and 13% in the Chinese sample, illustrating the materiality of this limitation.[20]

Second, due to privacy concerns, neighborhoods can only be broadly identified in microdata as the Public Use Microdata Area (PUMA) in which an immigrant resides. PUMAs are defined as geographically contiguous areal units with a population of 100,000 or more. As a spatial unit, PUMAs may be too large to study the desired day-to-day social interaction effects and neighborhood properties.

Third, although the most recent 3-year sample of the American Community Survey includes information on the year of naturalization, this information is usable in my model only if I know the corresponding attributes of immigrants at the time of naturalization.

Nonetheless, I control for these data shortcomings to some extent throughout my analysis. First, although I cannot distinguish temporary visa holders from green card holders, I at least control for eligible green card holders by leveraging information on the years an immigrant has been in the United States. Second, I add to the information provided by the PUMA variable by using macro data on the census tract level to identify PUMAs with segregated ethnic concentrations. Lastly, although I cannot use information

[20] Appendix A explains in detail how these proportions were calculated

on the year of naturalization in my logistic regression model, I apply this information in a Cox proportional hazard model to study the time to naturalization for recently naturalized immigrants (i.e. 2008-2011): this window ensures that the explanatory variables are relevant.

In accordance with the existing body of naturalization literature, I use a similar set of variables pertaining to personal characteristics, home ownership and family characteristics, and locational characteristics in my empirical analysis. Since regressions are run separately for the Chinese and Indian samples, I refrain from including any country of origin variables. The definition and construction of variables are summarized in Figures 11a and 11b.

Figure 11a: Individual Variable Descriptions and Constructions

Variable	**Definition**	**Modeling**	**Expectation**
Citizen	Coded 1 if naturalized, 0 otherwise	Dummy	N/A
Age	Age of immigrant	Continuous and squared term	Positive and diminishing
Years in USA	Number of years in USA since arrival	Continuous and squared term	Positive and diminishing
Female	Coded 1 if female, 0 if male	Dummy	Negative
Education Level	Highest level of education completed by immigrant, according to the following levels: (0) up to kindergarten (1) elementary school, middle school, and some high school (2) high school degree or GED or associate's degree (3) bachelor's degree (4) post-bachelor degrees, including masters and PhD	Dummy	Negative for below high school levels, positive for above high school levels
English Ability	Level of fluency in English, with (0) no ability (1) beginner's ability (2) proficient (3) fluent	Dummy	Positive for higher levels of English
Family Income	Annual income of all family members; all amounts over $1.5 million are censored as $1.5 million	Continuous and squared term	Positive and diminishing
Spouse’s Citizenship	Coded 1 if spouse is naturalized or American-born, 0 otherwise	Dummy	Positive
Child	Coded 1 if immigrant has any children, 0 otherwise	Dummy	Positive
Homeowner	Coded 1 if homeowner, 0 otherwise	Dummy	Positive

Figure 11b: Neighborhood Variable Descriptions and Constructions			
Variable	**Definition**	**Modeling**	**Expectation**
*_PUMA concentration	Proportion of PUMA's population with Chinese or Indian ancestry, excluding biracial population	Continuous	Positive
i.metarea	Coded 0 if not in a metro area and accords a dummy for over 200 metropolitan areas	Dummy	None
*_Cluster	Coded 1 if dissimilarity index score of PUMA is higher than 75th percentile value (0.6) for both Chinese and Indian, 0 otherwise. *_Diss is the index score included in summary statistics	Interaction term with *_PUMAperc	Negative
*_% naturalized PUMA	Percentage of PUMA's total population that is naturalized. Coded 1 if percentage is above 75th percentile value (i.e. 17%), 0 otherwise	Interaction term with *_PUMAperc	Positive
*_% bad Eng. PUMA	Percentage of PUMA that speaks "other language and bad English". Coded 1 if percentage is above 75th percentile value (i.e. 20%), 0 otherwise	Interaction term with *_PUMAperc	Negative
*_ poor PUMA	% of all people in PUMA that fall below the poverty line. Coded 1 if percentage is above 75th percentile value (i.e. 15%), 0 otherwise	Interaction term with *_PUMAperc	Negative

*_ refers to Indian or Chinese ancestry

Figures 12, 13, and 14 provide summary statistics for the variables employed in the logistic regression, with Figure 12 including the full sample and Figures 13 and 14 reporting the statistics for the Chinese and Indian sub-samples. To conserve space, I include the corresponding summary statistics for the survival analysis sample in Appendix B. I also include a correlation matrix for the independent variables in Appendix C.

Figure 12: Summary Statistics for Full Sample

Variable	N	Mean	Std. Dev.	Median	Min	Max
Citizen	54,429	0.66	0.47	1.00	0.00	1.00
Age	54,429	52.00	14.88	50.00	24.00	95.00
Years in USA	54,429	20.12	11.05	18.00	5.00	75.00
Female	54,429	0.51	0.50	1.00	0.00	1.00
Education Level	54,429	2.72	1.22	3.00	0.00	4.00
English Ability	54,429	2.16	1.01	3.00	0.00	3.00
Family Income	54,429	$125,405	$156,911	$92,000	-$17,800	$1,500,000
American Spouse	54,429	0.55	0.50	1.00	0.00	1.00
Child	54,429	0.64	0.48	1.00	0.00	1.00
Homeowner	54,429	0.76	0.43	1.00	0.00	1.00
Indian PUMA Concentration	54,421	0.03	0.04	0.02	0.00	0.22
Chinese PUMA Concentration	54,421	0.07	0.09	0.02	0.00	0.40
Indian Dissimilarity	54,421	0.51	0.15	0.48	0.00	0.99
Chinese Dissimilarity	54,421	0.46	0.16	0.43	0.00	0.99
% naturalized PUMA	54,421	0.13	0.09	0.11	0.00	0.37
% bad Eng. PUMA	54,421	0.15	0.11	0.12	0.00	0.57
% poor PUMA	54,421	10.82	6.37	9.14	1.89	44.18

Figure 13: Summary Statistics for Chinese Sub-Sample

Variable	n	Mean	Std. Dev.	Median	Min	Max
Citizen	26,674	0.67	0.47	1.00	0.00	1.00
Age	26,674	55.01	15.44	52.00	24.00	95.00
Years in USA	26,674	20.92	11.51	19.00	5.00	75.00
Female	26,674	0.55	0.50	1.00	0.00	1.00
Education Level	26,674	2.37	1.30	2.00	0.00	4.00
English Ability	26,674	1.69	1.06	2.00	0.00	3.00
Family Income	26,674	$104,106	$159,356	$67,083	-$8,741	$1,500,000
American Spouse	26,674	0.51	0.50	1.00	0.00	1.00
Child	26,674	0.59	0.49	1.00	0.00	1.00
Homeowner	26,674	0.72	0.45	1.00	0.00	1.00
Chinese PUMA Concentration	26,674	0.10	0.11	0.05	0.00	0.40
Chinese Dissimilarity	26,674	0.43	0.16	0.41	0.00	0.99
% naturalized PUMA	26,674	0.15	0.09	0.14	0.00	0.37

% bad Eng. PUMA	26,674	0.19	0.13	0.16	0.00	0.57
% poor PUMA	26,674	11.99	6.68	10.54	1.89	44.18

Figure 14: Summary Statistics for Indian Sub-Sample

Variable	n	Mean	Std. Dev.	Median	Min	Max
Citizen	27,755	0.66	0.48	1.00	0.00	1.00
Age	27,755	49.11	13.70	47.00	24.00	94.00
Years in USA	27,755	19.35	10.54	16.00	5.00	65.00
Female	27,755	0.47	0.50	0.00	0.00	1.00
Education Level	27,755	3.05	1.03	3.00	0.00	4.00
English Ability	27,755	2.62	0.71	3.00	0.00	3.00
Family Income	27,755	$145,874	$151,737	$111,258	-$17,800	$1,500,000
American Spouse	27,755	0.58	0.49	1.00	0.00	1.00
Child	27,755	0.68	0.47	1.00	0.00	1.00
Homeowner	27,755	0.79	0.41	1.00	0.00	1.00
Indian PUMA Concentration	27,755	0.04	0.04	0.02	0.00	0.22
Indian Dissimilarity	27,755	0.50	0.15	0.47	0.00	0.99
% naturalized PUMA	27,755	0.11	0.07	0.09	0.00	0.37
% bad Eng. PUMA	27,755	0.11	0.08	0.09	0.00	0.55
% poor PUMA	27,755	9.70	5.83	7.74	1.89	43.36

The summary statistics in Figures 12 to 14 offer a few preliminary insights into the sample. First, Chinese and Indian immigrants are comparable across most dimensions. Both immigrant groups on average are above the working age, have lived in the United States for a considerable amount of time at 20 years, and exhibit an equal gender split. Indicators such as English ability, education, income, and proportion of neighborhood below the poverty line illustrate a substantial degree of human capital and affluence for both groups. Nonetheless, Indians tend to fare slightly better across each of these four metrics than Chinese, echoing the results of Chapter 2b. On the other hand, there are some distinctions between both samples. For example, the average degree of ethnic neighborhood concentration in the Chinese sample (0.10) is much higher than that of the Indian sample (0.04). Additionally, a higher percentage of Indians are married (58%) and have children (68%) compared to the percentage of Chinese married (51%) with children (59%).

I include scatter plots in Appendix D depicting the correlations between PUMA concentrations and dissimilarity scores. The scatter plots illustrate that as the concentration of Indian or Chinese ethnic groups in a PUMA increases, their degree of segregation decreases.

Figure 16 documents the spread of the individual and household level variables among the naturalized pool of Chinese and Indian immigrants. The values are interpreted in the following manner: 28.37% of the 2,094 Chinese immigrants between the ages of 24 to 35 are naturalized; 37.39% of the 10,333 Chinese immigrants who have been in the United States for 5 to 15 years are naturalized, etc.

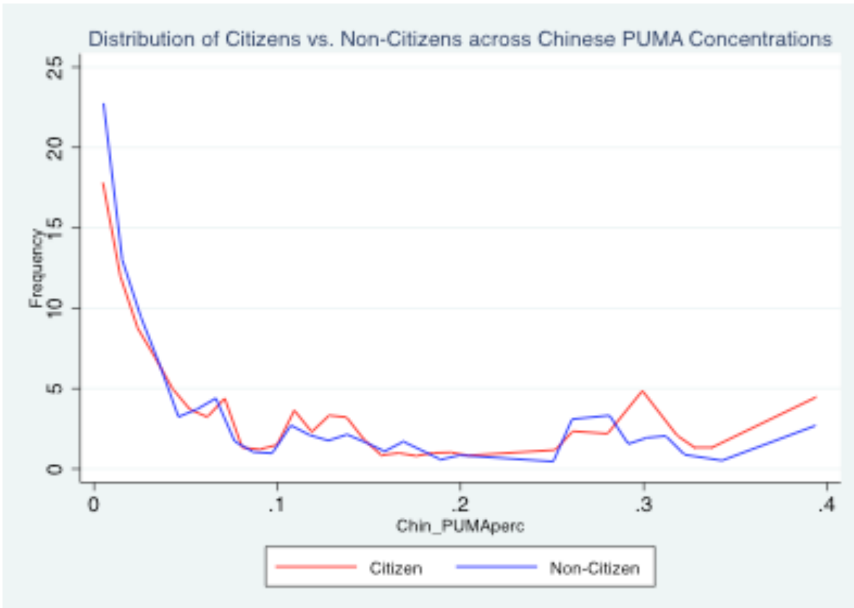
Figure 16: % Distribution of Naturalized Citizens by Individual Variables: Sub-Samples				
	Chinese		**Indian**	
Variable	**Base**	**% Naturalized**	**Base**	**% Naturalized**
Citizen	26,674	66.66%	27,755	65.53%
Age				
24 to 35	2,094	28.37%	4,961	29.07%
36 to 45	6,383	51.12%	8,026	58.41%
46 to 55	6,668	68.36%	5,865	78.36%
56 to 65	4,807	79.28%	5,028	85.28%
66 to 75	3,368	82.75%	2,813	84.64%
76 to 85	2,426	82.03%	885	77.06%
86+	928	83.94%	177	62.71%
Years in USA				
5 to 15	10,333	37.39%	13,288	42.05%
16 to 25	9,239	79.19%	7,184	81.85%
26 to 35	3,811	91.97%	4,452	90.81%
36 to 45	2,187	95.52%	2,437	95.08%
46 to 55	757	92.07%	362	92.54%
56 to 65	317	90.22%	32	81.25%
66 to 75	30	83.33%	0	0.00%
Female	14,670	69.08%	13,160	66.88%
Education Level				
No Education	2,141	67.21%	584	53.08%
Some High School	5,319	66.80%	1,882	62.22%
High School	7,414	70.35%	4,968	76.19%
Bachelor's	4,007	73.07%	8,332	66.33%
Post-bachelor's	7,793	59.62%	11,989	61.68%
English Ability				
No English	4,274	54.82%	621	37.20%
Speak Eng. poorly	7,572	70.69%	1,847	64.59%
Speak Eng. well	7,006	69.78%	5,067	72.75%
Speak Eng. fluently	7,822	66.44%	20,220	64.68%
Family Income				
<$20,000	4,634	69.14%	1277	64.06%
$20,000-$49,999	5,982	63.72%	3374	68.61%
$50,000-$99,000	6,792	65.92%	7227	62.42%
$100,000-$199,000	6,417	64.86%	10541	62.81%
$200,000-$300,000	1,814	72.33%	3037	70.96%
>$300,000	1,035	78.74%	2299	76.90%
American Spouse	13,555	84.93%	16,118	89.07%
Child	15,764	62.99%	18,926	63.85%
Homeowner	19,235	70.80%	21,931	72.38%
LowChin_State	535	51.03%	-	-
LowInd_State	-	-	222	58.11%

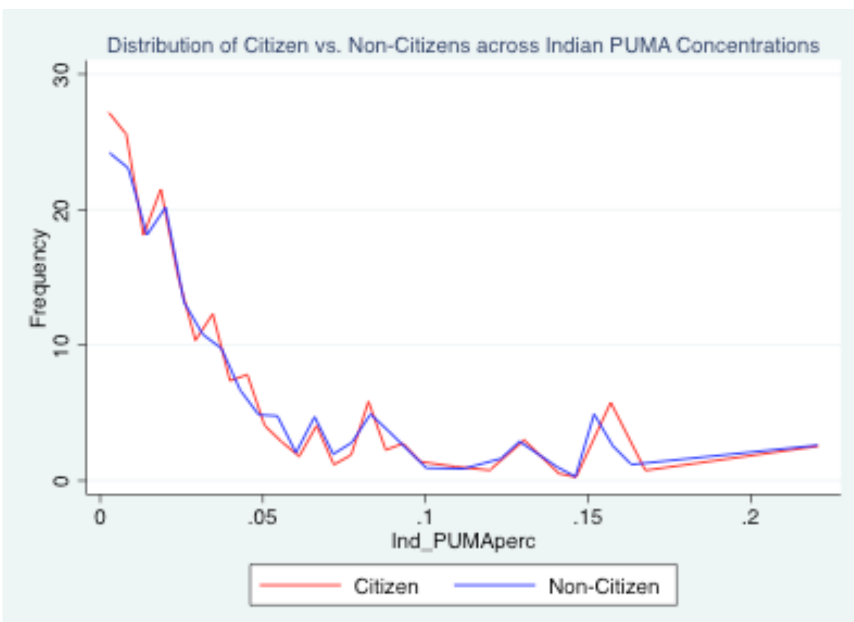
The initial relationships between each variable and the proportion naturalized confirm most but challenge some of the expectations detailed in the theory section in

Chapter 5. First, as expected, there is a generally positive relationship between age and years in the United States with the proportion naturalized, though the proportions tend to dip for the oldest bucket, 86+. However, this dip can be partially attributed to the significant smaller sample size. Second, a higher proportion of females are naturalized in both samples. Third, there does not appear to be any particular relationship with the level of education or family income and the proportion naturalized. Fourth, I find little to no positive relationship between language ability and naturalization, with one exception: immigrants who speak no English at all in both samples are less likely to have naturalized. The family variables appear to be particularly strong. Over 80% of immigrants who have a spouse who is a citizen are naturalized, over 60% of immigrants who have at least one child are naturalized, and over 70% of immigrants who own a house are naturalized. Lastly, there does not appear to be any initial relationships between immigrants in states with a low co-ethnic population.

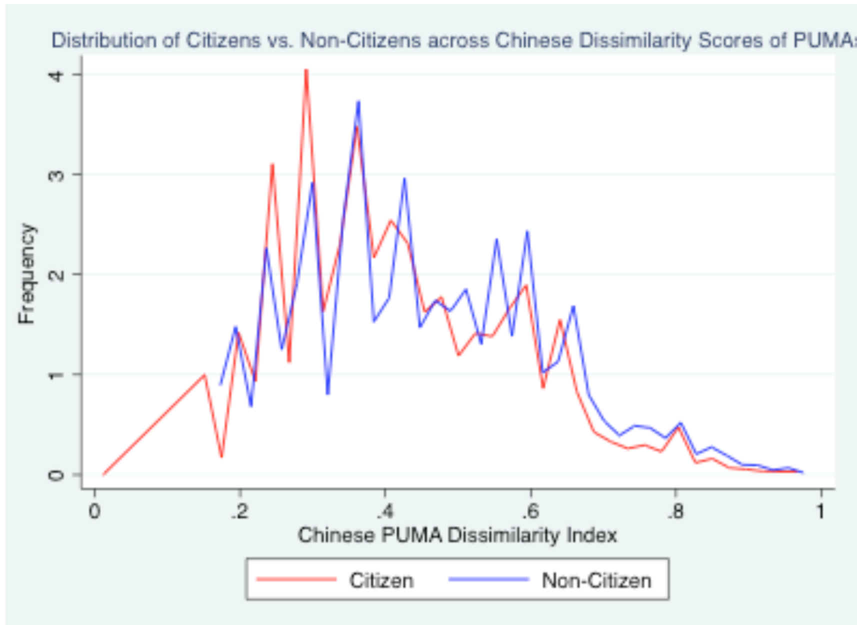
In the following series of figures, I continue to look at the distribution of naturalized vs. non-naturalized Chinese and Indian citizens across neighborhoods of differing: ethnic concentrations, dissimilarity scores, proportion of residents naturalized, levels of linguistic isolation, and poverty rates.

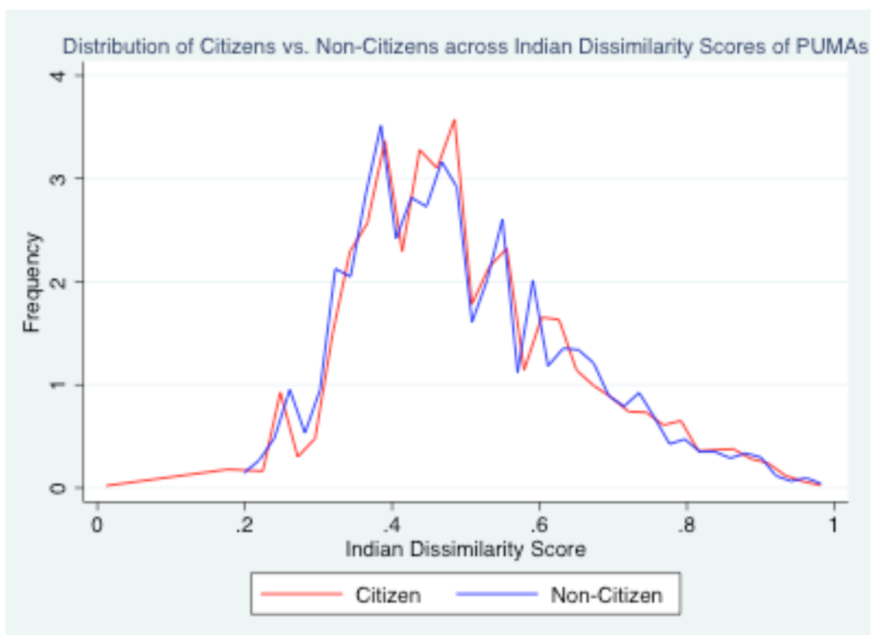
Figures 17a and 17b: Distribution of Citizens vs. Non-Citizens Across PUMA Co-Ethnic Concentrations



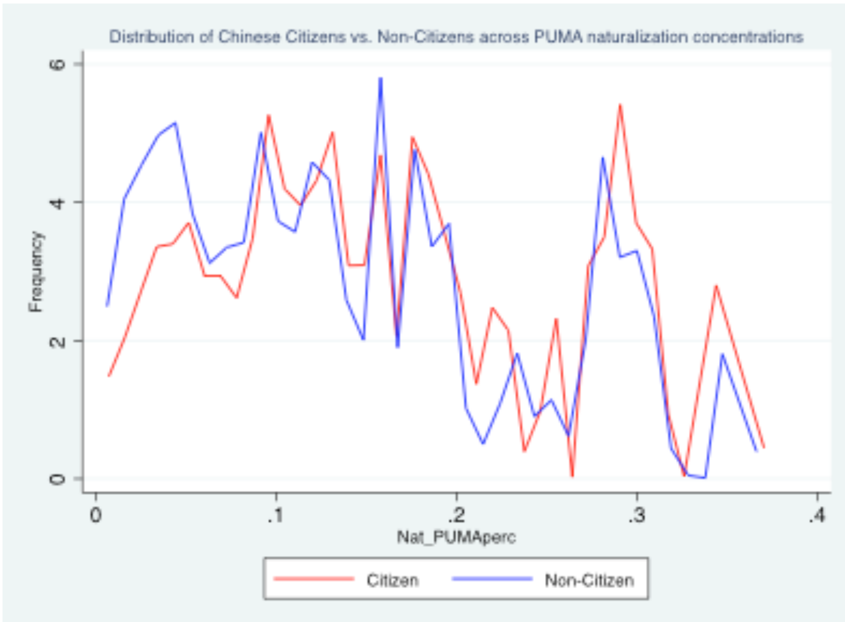


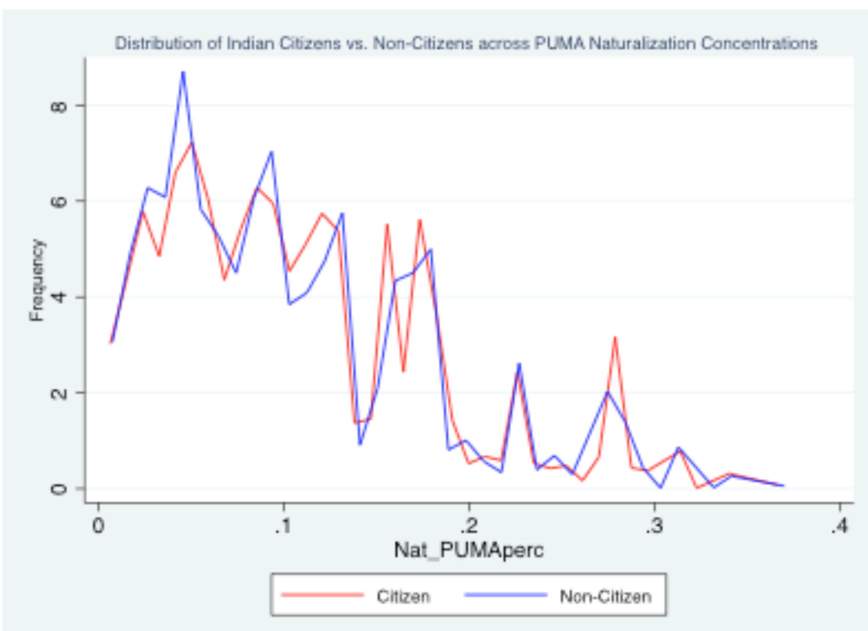
__**Figures 18a and 18b: Distribution of Citizens vs. Non-Citizens Across PUMA Dissimilarities**__



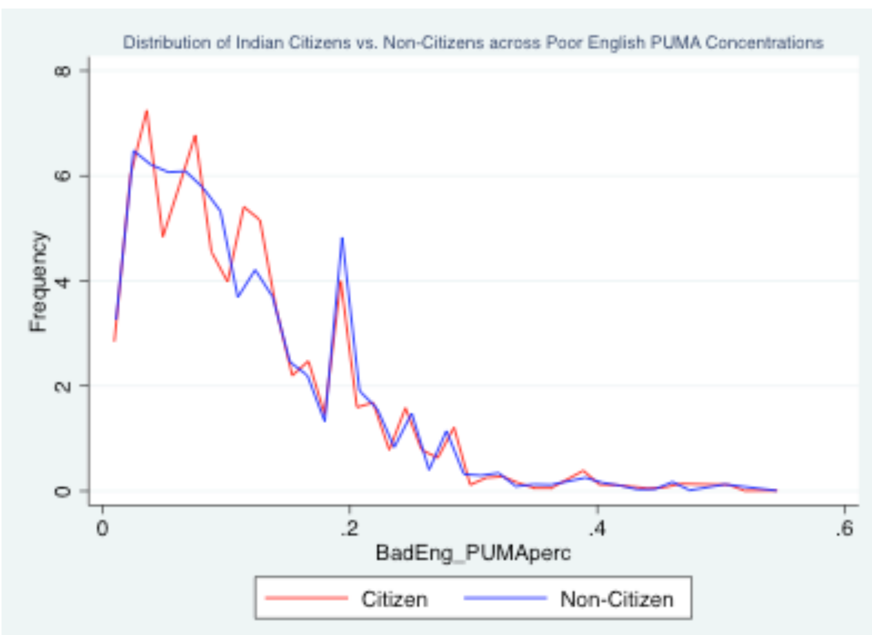


Figures 19a and 19b: Distribution of Citizens vs. Non-Citizens Across PUMA Naturalization Concentrations

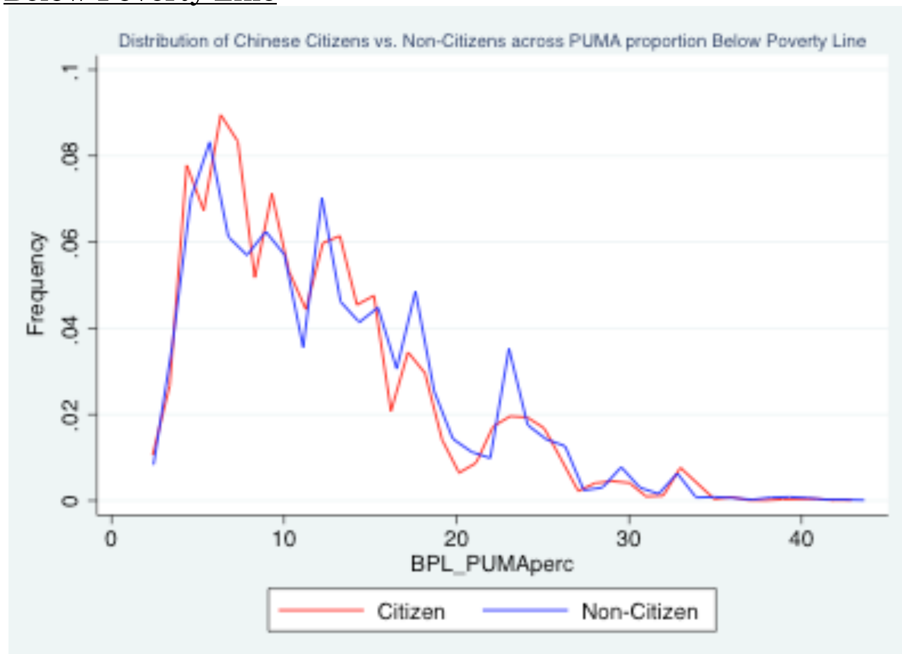


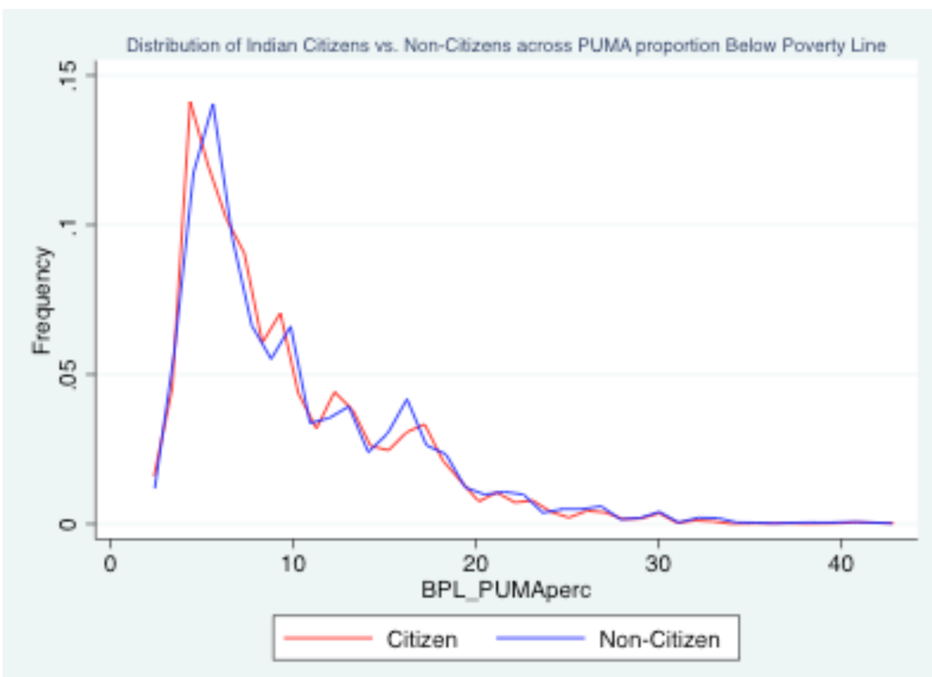


<u>Figures 20a and 20b: Distribution of Citizens vs. Non-Citizens Across PUMA Poor English Concentrations</u>



Figures 21a and 21b: Distribution of Citizens vs. Non-Citizens Across PUMA Proportion Below Poverty Line





By and large, Figures 17 to 21 illustrate that the distributional patterns of Chinese and Indians citizens across PUMA types closely resemble those of non-citizens, though there are slight differences at certain points on the PUMA spectrums. Starting with the co-ethnic PUMA concentrations, there are a higher number of Chinese non-citizens in PUMAs with very low co-ethnic concentrations, and a notably higher number of citizens in PUMAs with Chinese concentrations above 0.28. For Indian immigrants, there are a higher number of citizens in PUMAs with extremely low concentrations of co-ethnics (below 0.025), but the distributional patterns of Indian citizens overall matches that of non-citizens.

In terms of the PUMA dissimilarity scores, Chinese citizens are relatively concentrated in less segregated PUMAs of scores ranging from 0 to 0.3, while non-citizens are more clearly relatively concentrated in scores of 0.5 and higher. Indian citizens are clearly relatively more concentrated in less segregated PUMAs of dissimilarity scores below 0.2 and appear to be more numerous in scores of 0.4 to 0.5, but once again, their distributional pattern closely mirrors that of non-citizens.

Moving to the proportion of PUMA populations that are naturalized, there appear to be more Chinese and Indian non-citizens in PUMAs of up to 5% naturalized. However, beyond this 5% threshold, the distributions of non-citizens and citizens surprisingly mirror each other.

In terms of linguistically isolated PUMAs, there are slightly fewer Chinese citizens than non-citizens in areas with a high degree of poor English speakers (i.e. 0.4 to 0.6). However, there are also fewer Chinese citizens than non-citizens in PUMAs that are more linguistically integrated (i.e. 0 to 0.1). For Indian immigrants in general, very few immigrants, citizen and non-citizen alike, live in highly concentrated neighborhoods of poor English speakers. This trend makes sense, given that English is taught in the Indian educational system and spoken frequently throughout the country.

Lastly, there appears to be no significant variation in the citizen and non-citizen spread across neighborhoods with varying poverty levels. In general, very few Chinese and Indian immigrants in the sample live in very poor neighborhoods.

On the whole, there are no dramatic variations between distributions of citizens and non-citizens across different neighborhood properties. However, there are relatively more Chinese citizens in highly co-ethnic concentrated neighborhoods and relatively less Chinese citizens in neighborhoods that are highly segregated.

In Figure 22, I document the characteristics of immigrants who select into ethnic clusters. Ethnic clusters here are defined as dissimilarity index scores for Chinese and Indian PUMAs that are greater than 0.6. The values are interpreted in the following manner: 21.04% of the 4,961 Indian immigrants who are between the ages of 24 and 35 live in Indian clusters; 21.81% of the 13,288 Indian immigrants who have lived in the United States for 5 to 15 years live in Indian clusters, etc.

Figure 22: % Distribution of Chinese and Indian in Ethnic Cluster by Individual Variables

	Chinese		Indian	
Variable	**Base**	**% in Chin. Cluster**	**Base**	**% in Ind. Cluster**
Citizen	17,782	25.37%	18,188	23.45%
Age				
24 to 35	2,094	28.99%	4,961	21.04%
36 to 45	6,383	24.14%	8,026	20.70%
46 to 55	6,668	25.54%	5,865	23.10%
56 to 65	4,807	27.96%	5,028	27.21%
66 to 75	3,368	24.85%	2,813	26.84%
76 to 85	2,426	26.01%	885	30.51%
86+	928	23.28%	177	21.47%
Years in USA				
5 to 15	10,333	26.44%	13,288	21.81%
16 to 25	9,239	25.01%	7,184	22.83%
26 to 35	3,811	25.45%	4,452	26.35%
36 to 45	2,187	25.56%	2,437	26.71%
46 to 55	757	26.82%	362	31.49%
56 to 65	317	29.65%	32	46.88%
66 to 75	30	33.33%	0	0.00%
Female	14,670	25.50%	13,160	23.21%
Education Level				
No Education	2,141	17.66%	584	30.14%
Some High School	5,319	16.17%	1,882	30.23%
High School	7,414	16.56%	4,968	28.04%
Bachelor's	4,007	13.15%	8,332	21.01%
Post-bachelor's	7,793	15.06%	11,989	21.73%
English Ability				
No English	4,274	28.47%	621	27.86%
Speak Eng. poorly	7,572	28.75%	1,847	29.40%
Speak Eng. well	7,006	23.81%	5,067	25.00%
Speak Eng. fluently	7,822	23.23%	20,220	22.29%
Family Income				
<$20,000	4,634	28.85%	1277	35.47%
$20,000-$49,999	5,982	30.99%	3374	32.84%
$50,000-$99,000	6,792	29.31%	7227	27.09%
$100,000-$199,000	6,417	19.87%	10541	18.00%
$200,000-$300,000	1,814	11.69%	3037	15.38%
>$300,000	1,035	20.29%	2299	26.45%
American Spouse	13,555	24.94%	16,118	22.97%
Child	15,764	24.97%	18,926	21.29%
Homeowner	19,235	26.05%	21,931	23.17%

I define the ethnic cluster term at the 75th percentile score (i.e. 0.6) for both Chinese and Indian dissimilarity indexes; as such, a random distribution of the sample would mean that 25% of Chinese and Indian immigrants live in ethnic clusters. Overall, the values in Figure 22 do not depict many dramatic departures from this 25% benchmark.

There are some notable exceptions: only 11% to 20% of Chinese and Indian immigrants with family incomes between $100,000 and $300,000 cluster with co-ethnics. Concurrently, around 30% of immigrants with an income below $50,000 cluster together. Second, the elderly cohorts of Chinese and Indian immigrants tend to live close to co-ethnics: 33% of Chinese immigrants aged 66 to 75 cluster together and 47% of Indian immigrants aged 56 to 65 cluster together. Additionally, Chinese and Indian immigrants with a high school education or below cluster more than those with a bachelor's degree or higher. Lastly, though one would expect dramatic clustering for those with no English ability, the departure from 25% is only slight at 28% for both ethnic groups. Nonetheless, these results may be sensitive to the definition of an ethnic cluster. For instance, examining the distribution using general PUMA concentrations instead of dissimilarity scores may create a slightly different picture.

Chapter 7: Empirical Strategy

I employ two models to explore the determinants of naturalization. Logistic regression is the first type of model, used in cases where the dependent variable is discrete. With a binary dependent variable, one can no longer assume normality in the distribution of error terms, thereby reducing the efficiency of any results an Ordinary Least Squares regression would produce. The logistic regression is a more efficient model and can be expressed as:

$$\ln[\frac{p_i}{1-p_i}] = a + \sum b_i X_i + e$$

In this equation, i is individual {1, 2, 3…n}, p is the probability that Y=1 (i.e. immigrant is a naturalized citizen), a is a constant, X is a vector of explanatory variables with b as the vector coefficients, and e is the error term. The left hand-side ratio represents the odds of Y=1 (i.e. the immigrant is a naturalized citizen). The coefficients b in this estimation report an odds ratio. Values greater than 1 indicate a positive effect of the independent variable on the probability of naturalization by a magnitude of b-1; values less than 1 indicate a negative effect of the independent variable on the probability of naturalization by a magnitude of 1-b; and values of 1 suggest no effect. The null hypothesis is thus tested against a coefficient of 1 instead of 0, and p-values indicate the degree to which the beta value is statistically different from 1.

My second model uses a Cox Proportional Hazard Model, used in survival analysis when examining the time it takes for an event to occur (in this case, time to naturalize). It adopts the functional form:

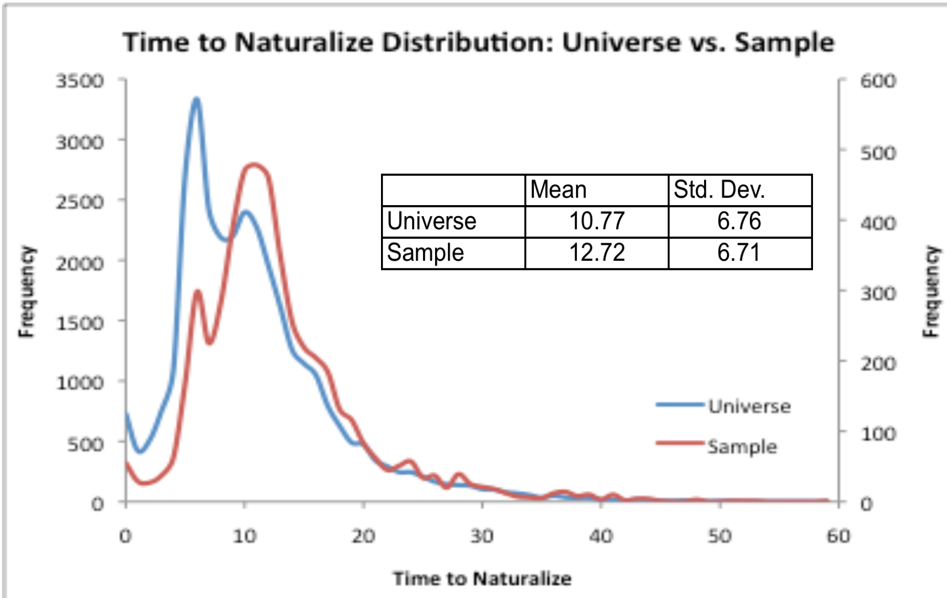
$$h(t,\mathbf{X}) = h_0(t)\exp(\sum b_i X_i)$$

In this model $h_0(t)$ is a baseline hazard dependent only on time, while X is a vector of time independent explanatory variables. The dependent variable is constructed based on the time to event (i.e. years to naturalization) and the occurrence of the event itself (i.e. if the immigrant is a citizen or not). The model is right-censored in this manner, because it includes observations where the event has not yet occurred.

The Cox Proportional Hazard model offers two specific benefits compared to other survival analysis models. First, its semi-parametric nature offers more flexibility in estimation than other survival analysis models, which are mostly parametric. Second, the reported hazard ratios as coefficients offer an easier interpretation of model results and can be read in a similar manner to the odds ratios of a logistic regression. For example, for values greater than 1, the time to naturalize is shorter by a percentage of $b - 1$, and for values less than 1, the time to naturalization is longer by a percentage of $1 - b$.

Since many of the independent variables are time varying, I reduce the sample for the survival analysis to include only those who naturalized in 2008 – 2011, and I randomly reduce the pool of non-naturalized citizens by a proportional amount. Such a reduction assumes that the most recent window of naturalizations is a random sample of the total universe. Figure 23 illustrates that this is likely not the case, as the mean time to naturalize for the recent window (10.77) is approximately two points higher than that of the universe (12.72). Nonetheless, given the data constraints, it is a necessary assumption to make.

Figure 23: Time to Naturalize: Universe vs. Sample Distribution



Chapter 8: Results

Regression 1 tests the first hypothesis: ethnic concentrations will be positively associated with the propensity to naturalize, and this effect will be intensified when the concentrations are clustered. The results for Chinese neighborhoods support this hypothesis: an incremental increase in the Chinese PUMA density more than triples the odds of a Chinese citizen being naturalized, and the presence of an enclave intensifies this effect four-fold. This latter effect, however, is only significant at the 10% level. On the other hand, the results for Indian neighborhoods are not statistically significant.

For both immigrant groups, years in USA, English ability, having a high school degree, spouse's citizenship, and home ownership are positive and statistically significant, in line with expectations and results in the literature. It should be noted that the years in USA variable captures both a learning effect as well as a cohort effect: older immigrants have acquired more U.S. specific human capital, and, as Chapter 2a illustrates, dealt with very different naturalization climates. As a result, this variable is limited in its ability to predict behaviors for newer cohorts. In terms of the other variables, the magnitude of the English ability variable interestingly does not increase with degree of fluency; rather, having some ability as opposed to none at all has a consistently strong effect throughout the levels of ability. Spouse's citizenship has the strongest effect in terms of both magnitude and statistical significance out of all explanatory variables. Surprisingly, having children is negatively correlated with the probability of being naturalized. Based on the previous literature as well as Figure 16 in Chapter 6, this result is unexpected and not intuitional.

Furthermore, having some primary or secondary schooling background and family income is insignificant for both immigrant groups. The first result is not surprising in suggesting that an incomplete schooling background is indistinguishable from no schooling at all. The second result may be due to the high degree of correlation shared between education and family income, as well as the magnitude of the units in question (i.e. thousands of dollars as opposed to percentages).

Being a female and having a bachelor's degree is positively correlated with the probability of being naturalized for Chinese immigrants but insignificant for Indians. Interestingly, having a graduate degree reduces the odds of being naturalized for Chinese immigrants. This result may be capturing the body of immigrants who are on skilled temporary visas and ineligible to naturalize. In addition, following Massey's and Akresh's (2008) line of reasoning, skilled immigrants may be less likely to naturalize, because many are willing and able to move to other countries that offer a higher return on their skills.

In Regression 2, I include metro area dummies and cluster standard errors by metro area to test my second hypothesis: positive correlations should still persist after controlling for metropolitan level variation. The results offer no evidence for this hypothesis: the coefficients on Chinese PUMA concentration and the enclave interaction term are no longer significant, and the Indian PUMA concentration and interaction term variables remain insignificant. The results for the other explanatory variables remain approximately the same, with some variation occurring in the educational set of variables.

In Regression 3, I test my third hypothesis: positive correlations will persist for high skilled immigrants after running regressions separately for immigrants with at least a bachelor's degree and for those without one. I retain the metropolitan level dummies and clustered errors in these regressions. The results do not support my hypothesis and thus offer no evidence of an amenities effect: Chinese neighborhood variables remain insignificant, and the Indian neighborhood terms for high skilled immigrants are also statistically insignificant. Interestingly, however, the Indian PUMA concentration term for low skilled immigrants is positive and highly significant in terms of magnitude and statistical significance: an incremental increase in co-ethnic concentrations increases the odds of a low skilled Indian immigrant being naturalized nearly five-fold.

The results for the other explanatory variables are approximately the same with a few exceptions. The home ownership variable is no longer significant for low skilled Chinese immigrants, and having at least one child is no longer significant for high skilled Indian immigrants. The gender variable is now statistically significant for both classes of Indian immigrants. However, being a female reduces the odds of naturalization for low skilled Indian immigrants, while it increases the odds for high skilled Indian immigrants.

In Regression 4, I test my hypothesis that the positive coefficients of ethnic concentrations will diminish when interacted with linguistic isolation and poverty and will increase when interacted with proportion naturalized. I retain my metro area controls and continue to separate my regressions by immigrant skill level. The results do not support my hypothesis, though the Indian sample offers some weak evidence. For the Chinese sample, the only statistically significant neighborhood variable is the interacted term between Chinese PUMA concentrations and below the poverty line. Surprisingly, this interaction term increases the odds of naturalization rather than reducing it. Nonetheless, a positive correlation could theoretically exist for a few reasons: first, poor neighborhoods may have a higher level of non-profit organizations targeting their needs and assisting them through the naturalization process. Second, if the poverty status of the neighborhood is indicative of the immigrant's own wealth, poor immigrants stand to gain the most from naturalizing, since they can access a wider suite of welfare benefits.

For the Indian sample, the Indian PUMA concentration term remains statistically significant for low skilled immigrants and jumps in magnitude with the introduced interacted terms. However, with the exception of the poverty interaction term, none of these interacted terms are significant for low skilled Indian immigrants. Like the Chinese sample, the poverty coefficient is positive. In the high skilled Indian sample, the interacted enclave term and poverty term exhibit a weak negative correlation with naturalization odds, and the naturalization interaction term falls in line with expectations and has a positive and highly significant value. The results for the other explanatory variables are consistent with the results from Regression 3.

In Regressions 5 – 8, I adapt the same set of hypotheses and add in an age variable for the survival analysis model. Regression 5 supports the first hypothesis: an incremental increase in Chinese and Indian concentrations is associated with a faster time to naturalization, though the enclave interaction term is statistically insignificant. The results of the other explanatory terms mirror the results of Regression 1. Tertiary education is strong and positively significant for Chinese but insignificant for Indians; females naturalize faster than males; home ownership, English ability, and having a spouse who is a citizen increase the speed at which an immigrant naturalizes for both immigrant groups. For Indians, having a child now is in line with expectations and increases the speed of naturalization.

Regression 6 supports the second hypothesis: after controlling for metropolitan variation, Chinese and Indian PUMA concentrations still remain positive. However, clustering is associated with a slower time to naturalization and is highly significant in terms of magnitude and statistical significance. The other explanatory variables remain approximately the same in comparison to Regression 5. One difference is that females are now less likely to naturalize than Indian immigrants.

Regression 7 offers mixed support for the third hypothesis: after separating immigrants by class, Chinese neighborhoods have no significant effect, but Indian PUMA concentrations for the high skilled segment have a strong positive effect on the speed of naturalization, though clustering strongly reduces this speed. Interestingly, this result is the converse of Regression 3, where Indian PUMA concentrations were significant for low skilled immigrants but not for high skilled. The results of the other explanatory variables mirror the results of Regressions 5 and 6.

Regression 8 does not support the fourth hypothesis: after adding interacted terms of neighborhood social capital with ethnic concentration, Chinese concentrations for low skilled immigrants now reduces the speed to naturalization, although the coefficient stays positive and significant for both high and low skilled Indians. Like Regression 4, the poverty interaction term for low skilled Chinese immigrants is the only significant neighborhoods interaction term for the Chinese sample. In the Indian sample, clustering reduces the effect of the ethnic concentration variable, but none of the other interaction terms has an effect. Once again, the effects of the other explanatory terms are roughly consistent across the survival regressions.

The regression results are displayed in the next few pages. Coefficient terms with a star signify that they are interaction terms, and details of how to interpret the results are included at the bottom of the table.

REGRESSION 1: GENERAL HYPOTHESIS

CITIZEN	(1) Chinese Odds Ratio*	(2) Chinese Odds Ratio*	(3) Indian Odds Ratio*	(4) Indian Odds Ratio*
Years in USA	1.346***	1.346***	1.326***	1.326***
	(0.00778)	(0.00778)	(0.0101)	(0.0101)
Years in USA squared	0.997***	0.997***	0.996***	0.996***
	(9.92e-05)	(9.92e-05)	(0.000156)	(0.000156)
Gender (omitted male)	1.686***	1.686***	1.046	1.046
	(0.0579)	(0.0580)	(0.0380)	(0.0380)
Some High School (omitted no educ.)	0.964	0.964	0.961	0.962
	(0.0670)	(0.0670)	(0.122)	(0.122)
High School/GED	1.186**	1.185**	1.495***	1.495***
	(0.0829)	(0.0829)	(0.187)	(0.187)
Bachelor's degree	1.245***	1.246***	1.175	1.174
	(0.100)	(0.100)	(0.147)	(0.147)
Graduate degree/PhD	0.769***	0.771***	0.869	0.868
	(0.0612)	(0.0614)	(0.109)	(0.109)
Speaks Eng. poorly (omitted no. Eng)	2.271***	2.274***	3.252***	3.248***
	(0.120)	(0.120)	(0.398)	(0.397)
Speaks Eng. well	2.796***	2.804***	4.110***	4.103***
	(0.175)	(0.175)	(0.495)	(0.495)
Speaks Eng. fluently	2.516***	2.526***	3.400***	3.391***
	(0.166)	(0.167)	(0.406)	(0.406)
Family income	1.000	1.000	1.001	1.001
	(0.00106)	(0.00106)	(0.00109)	(0.00109)
American Spouse	4.831***	4.830***	12.21***	12.21***
	(0.170)	(0.170)	(0.452)	(0.452)
Child (at least 1)	0.707***	0.707***	0.810***	0.809***
	(0.0259)	(0.0259)	(0.0344)	(0.0344)
Homeowner	1.121***	1.121***	1.459***	1.459***
	(0.0438)	(0.0438)	(0.0620)	(0.0620)
Chinese PUMA concentration	3.460***	3.492***		
	(0.550)	(0.555)		
*Chinese Cluster		4.517*		
		(3.964)		
Indian PUMA concentration			1.374	1.378
			(0.541)	(0.543)
*Indian Cluster				0.367
				(0.687)
Constant	0.00705***	0.00698***	0.00431***	0.00433***
	(0.000742)	(0.000735)	(0.000698)	(0.000703)
Observations	26,669	26,669	27,752	27,752
Pseudo R2	0.3405	0.3406	0.4205	0.4205

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
*Odds Ratios are interpreted such that OR>1 increases probability of being naturalized by (OR-1)%, OR<1 decreases probability of being naturalized by (1 – OR)%, and OR=1 signifies no influence

REGRESSION 2: SELECTION EFFECT

CITIZEN	(1) Chinese Odds Ratio*	(2) Indian Odds Ratio*
Years in USA	1.351***	1.330***
	(0.0117)	(0.0139)
Years in USA squared	0.996***	0.996***
	(0.000147)	(0.000200)
Gender (omitted male)	1.680***	1.046
	(0.0496)	(0.0487)
Some High School (omitted no educ.)	0.981	0.962
	(0.0575)	(0.157)
High School/GED	1.152	1.514***
	(0.135)	(0.169)
Bachelor's degree	1.194	1.211*
	(0.161)	(0.134)
Graduate degree/PhD	0.773**	0.910
	(0.0994)	(0.0939)
Speaks Eng. poorly (omitted no Eng.)	2.344***	3.298***
	(0.145)	(0.440)
Speaks Eng. well	2.954***	4.257***
	(0.263)	(0.601)
Speaks Eng. fluently	2.704***	3.571***
	(0.199)	(0.511)
Family income	0.999	1.001
	(0.000999)	(0.00146)
American Spouse	4.696***	12.26***
	(0.368)	(1.032)
Child (at least 1)	0.722***	0.818***
	(0.0408)	(0.0390)
Homeowner	1.122**	1.467***
	(0.0535)	(0.0532)
Chinese PUMA concentration	1.238	
	(0.251)	
*Chinese Cluster	6.156	
	(11.65)	
Indian PUMA concentration		1.173
		(0.499)
*Indian Cluster		0.251
		(0.294)
Constant	0.00489***	0.00377***
	(0.000834)	(0.000790)
Observations	26,551	27,606

Robust se in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
*Odds Ratios are interpreted such that OR>1 increases probability of being naturalized by (OR-1)%, OR<1 decreases probability of being naturalized by (1 – OR)%, and OR=1 signifies no influence

REGRESSION 3: AMENITIES EFFECT				
	(1)	(2)	(3)	(4)
	Chinese	Chinese	Indian	Indian
	Low Skill	High Skill	Low Skill	High Skill
CITIZEN	Odds Ratio*	Odds Ratio*	Odds Ratio*	Odds Ratio*
Years in USA	1.292***	1.466***	1.185***	1.414***
	(0.0110)	(0.0222)	(0.0250)	(0.0144)
Years in USA squared	0.997***	0.996***	0.998***	0.995***
	(0.000162)	(0.000212)	(0.000426)	(0.000205)
Gender (omitted male)	1.830***	1.558***	0.875**	1.163***
	(0.0721)	(0.101)	(0.0564)	(0.0668)
Speaks Eng. poorly (omitted no Eng.)	2.393***	1.845***	3.113***	4.003***
	(0.149)	(0.282)	(0.356)	(2.039)
Speaks Eng. well	3.646***	1.731***	4.609***	3.125**
	(0.437)	(0.361)	(0.592)	(1.557)
Speaks Eng. fluently	2.376***	1.684**	4.917***	2.248*
	(0.199)	(0.355)	(0.712)	(1.078)
Family income	0.999	0.994**	1.000	0.999
	(0.00104)	(0.00244)	(0.00270)	(0.00175)
American Spouse	2.979***	8.563***	3.894***	20.88***
	(0.214)	(0.915)	(0.304)	(1.745)
Child (at least 1)	0.832***	0.577***	0.946	0.781***
	(0.0516)	(0.0414)	(0.0713)	(0.0478)
Homeowner	1.026	1.343***	1.503***	1.421***
	(0.0621)	(0.0756)	(0.0972)	(0.0562)
Chinese PUMA concentration	1.096	1.402		
	(0.311)	(0.912)		
*Chinese Cluster	0.948	20.08		
	(1.624)	(57.16)		
Indian PUMA concentration			4.991***	0.757
			(1.916)	(0.407)
*Indian Cluster			0.607	0.108
			(0.365)	(0.195)
Constant	0.00938***	0.00229***	0.0254***	0.00247***
	(0.00146)	(0.000512)	(0.00580)	(0.00123)
Observations	14,699	11,661	7,265	20,163
Pseudo R2	0.2881	0.4629	0.2325	0.5153

Robust se in parentheses

*** p<0.01, ** p<0.05, * p<0.1

*Odds Ratios are interpreted such that OR>1 increases probability of being naturalized by (OR-1)%, OR<1 decreases probability of being naturalized by (1 – OR)%, and OR=1 signifies no influence

REGRESSION 4: SOCIAL CAPITAL EFFECT				
	(1)	(2)	(3)	(4)
	Chinese	Chinese	Indian	Indian
	Low Skill	High Skill	Low Skill	High Skill
CITIZEN	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio
Years in USA	1.291***	1.466***	1.187***	1.412***
	(0.0110)	(0.0220)	(0.0247)	(0.0145)
Years in USA squared	0.997***	0.996***	0.998***	0.995***
	(0.000159)	(0.000210)	(0.000419)	(0.000206)
Gender (omitted male)	1.832***	1.556***	0.870**	1.163***
	(0.0748)	(0.101)	(0.0547)	(0.0666)
Speaks Eng. poorly (omitted no Eng.)	2.406***	1.862***	3.123***	3.999***
	(0.143)	(0.294)	(0.357)	(2.038)
Speaks Eng. well	3.676***	1.761***	4.646***	3.101**
	(0.413)	(0.384)	(0.608)	(1.549)
Speaks Eng. fluently	2.385***	1.716**	4.920***	2.229*
	(0.194)	(0.380)	(0.716)	(1.073)
Family income	0.999	0.994**	1.000	0.999
	(0.00107)	(0.00242)	(0.00268)	(0.00175)
American Spouse	2.974***	8.577***	3.892***	20.85***
	(0.204)	(0.919)	(0.302)	(1.749)
Child (at least 1)	0.834***	0.582***	0.944	0.775***
	(0.0510)	(0.0429)	(0.0718)	(0.0474)
Homeowner	1.034	1.359***	1.505***	1.409***
	(0.0623)	(0.0828)	(0.101)	(0.0551)
Chinese PUMA concentration	0.107	0.939		
	(0.162)	(0.848)		
*Chinese Cluster	1.571	11.83		
	(2.579)	(27.58)		
*Chin. % bad Eng. PUMA	2.553	1.913		
	(3.663)	(1.613)		
*Chin. % naturalized PUMA	2.918	0.651		
	(1.988)	(0.856)		
*Chin. % poor PUMA	1.889***	2.787		
	(0.361)	(2.751)		
Indian PUMA concentration			19.12***	0.596
			(13.00)	(0.430)
*Indian Cluster			3.502	0.0285*
			(8.476)	(0.0543)
*Ind. % bad Eng. PUMA			0.241	0.910
			(0.607)	(2.079)
*Ind. % naturalized PUMA			0.0426	4.059***
			(0.112)	(2.099)
*Ind. % poor PUMA			25.86**	0.0376*
			(41.23)	(0.0746)
Constant	0.00935***	0.00220***	0.0249***	0.00256***
	(0.00135)	(0.000522)	(0.00568)	(0.00128)
Observations	14,699	11,661	7,265	20,163
Pseudo R2	0.2887	0.4631	0.2336	0.5154

Robust se in parentheses

*** p<0.01, ** p<0.05, * p<0.1

*Odds Ratios are interpreted such that OR>1 increases probability of being naturalized by (OR-1)%, OR<1 decreases probability of being naturalized by (1 – OR)%, and OR=1 signifies no influence

REGRESSION 5: GENERAL HYPOTHESIS		
	(1)	(2)
	Chinese	Indian
CITIZEN	Hazard Ratio	Hazard Ratio
Age	1.061***	1.016
	(0.0126)	(0.0124)
Age squared	1.000***	1.000
	(0.000106)	(0.000123)
Gender (omitted male)	1.267***	0.940
	(0.0588)	(0.0357)
Some High School	0.941	0.840
	(0.103)	(0.142)
High School/GED	1.052	0.982
	(0.112)	(0.162)
Bachelor's degree	1.314**	1.236
	(0.150)	(0.204)
Graduate degree/PhD	1.258**	1.119
	(0.146)	(0.186)
Speaks Eng. poorly	1.673***	2.862***
	(0.137)	(0.580)
Speaks Eng. well	2.058***	3.046***
	(0.193)	(0.617)
Speaks Eng. fluently	2.078***	3.247***
	(0.200)	(0.660)
Family income	1.000	1.000**
	(1.95e-07)	(1.53e-07)
Homeowner	1.215***	1.160***
	(0.0671)	(0.0584)
American Spouse	3.497***	10.89***
	(0.163)	(0.523)
Child (at least 1)	0.933	1.384***
	(0.0463)	(0.0705)
Chinese PUMA concentration	2.102***	
	(0.458)	
*Chinese Cluster	1.730	
	(2.154)	
Indian PUMA concentration		6.486***
		(2.530)
*Indian Cluster		0.0539
		(0.118)
Observations	10,883	12,553

Robust se in parentheses
*** p<0.01, ** p<0.05, * p<0.1

*Hazard Ratios (HRs) are interpreted in a similar way to odds ratios. HR>1 increases the speed of/reduces the time to naturalization by (HR – 1)%, HR<1 decreases the speed of/reduces the time to naturalization by (1 – HR)%, and HR=1 signifies no influence

REGRESSION 6: SELECTION HYPOTHESIS		
	(1)	(2)
	Chinese	Indian
CITIZEN	Hazard ratio	Hazard ratio
Age	1.057***	1.014
	(0.00995)	(0.0158)
Age squared	1.000***	1.000
	(8.85e-05)	(0.000152)
Gender (omitted male)	1.281***	0.928**
	(0.0472)	(0.0350)
Some High School	0.957	0.880
	(0.0662)	(0.139)
High School/GED	1.057	1.047
	(0.117)	(0.179)
Bachelor's degree	1.308**	1.299*
	(0.151)	(0.207)
Graduate degree/PhD	1.266*	1.167
	(0.156)	(0.188)
Speaks Eng. poorly	1.628***	2.954***
	(0.115)	(0.557)
Speaks Eng. well	2.002***	3.100***
	(0.220)	(0.646)
Speaks Eng. fluently	2.016***	3.295***
	(0.132)	(0.720)
Family income	1.000	1.000
	(2.63e-07)	(2.60e-07)
Homeowner	1.205**	1.161**
	(0.0887)	(0.0756)
American Spouse	3.347***	10.97***
	(0.374)	(1.122)
Child (at least 1)	0.936	1.394***
	(0.0562)	(0.0892)
Chinese PUMA concentration	1.395**	
	(0.190)	
*Chinese Cluster	1.996	
	(2.362)	
Indian PUMA concentration		6.903***
		(1.332)
*Indian Cluster		0.0192***
		(0.0252)
Observations	10,883	12,553

Robust se in parentheses
*** p<0.01, ** p<0.05, * p<0.1
*Hazard Ratios (HRs) are interpreted in a similar way to odds ratios. HR>1 increases the speed of/reduces the time to naturalization by (HR – 1)%, HR<1 decreases the speed of/reduces the time to naturalization by (1 – HR)%, and HR=1 signifies no influence

REGRESSION 7: AMENITIES HYPOTHESIS				
	(1)	(2)	(3)	(4)
	Chinese	Chinese	Indian	Indian
	Low Skill	High Skill	Low Skill	High Skill
CITIZEN	Hazard Ratio	Hazard Ratio	Hazard Ratio	Hazard Ratio
Age	1.021	1.106***	0.949**	1.075***
	(0.0191)	(0.0259)	(0.0231)	(0.0229)
Age squared	1.000	0.999***	1.000**	0.999***
	(0.000159)	(0.000205)	(0.000210)	(0.000214)
Gender (omitted male)	1.360***	1.209***	0.807**	0.938
	(0.0787)	(0.0776)	(0.0792)	(0.0385)
Speaks Eng. poorly	1.731***	1.342	3.124***	2.405
	(0.147)	(0.292)	(0.679)	(1.483)
Speaks Eng. well	2.190***	1.689**	3.826***	1.501
	(0.369)	(0.361)	(0.834)	(0.915)
Speaks Eng. fluently	1.796***	1.742***	3.846***	1.614
	(0.174)	(0.323)	(0.964)	(0.977)
Family income	1.000*	1.000	1.000	1.000
	(2.72e-07)	(4.81e-07)	(4.95e-07)	(2.15e-07)
Homeowner	1.150*	1.243**	1.159	1.158**
	(0.0910)	(0.123)	(0.154)	(0.0676)
American Spouse	2.091***	4.929***	3.097***	16.56***
	(0.208)	(0.649)	(0.359)	(1.952)
Child (at least 1)	0.969	0.921	1.038	1.514***
	(0.0870)	(0.0581)	(0.0933)	(0.116)
Chinese PUMA concentration	1.115	1.427		
	(0.399)	(0.522)		
*Chinese Cluster	3.528	0.566		
	(8.290)	(0.448)		
Indian PUMA concentration			6.556	5.769***
			(10.02)	(1.611)
*Indian Cluster			0.0335***	0.0102**
			(0.0358)	(0.0213)
Observations	5,565	5,318	2,773	9,780

Robust se in parentheses
*** p<0.01, ** p<0.05, * p<0.1
*Hazard Ratios (HRs) are interpreted in a similar way to odds ratios. HR>1 increases the speed of/reduces the time to naturalization by (HR – 1)%, HR<1 decreases the speed of/reduces the time to naturalization by (1 – HR)%, and HR=1 signifies no influence

REGRESSION 8: SOCIAL CAPITAL HYPOTHESIS				
	(1)	(2)	(3)	(4)
	Chinese	Chinese	Indian	Indian
	Low Skill	High Skill	Low Skill	High Skill
CITIZEN	Hazard Ratio	Hazard Ratio	Hazard Ratio	Hazard Ratio
Age	1.020	1.106***	0.949**	1.076***
	(0.0181)	(0.0259)	(0.0234)	(0.0229)
Age squared	1.000	0.999***	1.000*	0.999***
	(0.000152)	(0.000204)	(0.000213)	(0.000214)
Gender (omitted male)	1.359***	1.208***	0.806**	0.938
	(0.0800)	(0.0782)	(0.0787)	(0.0388)
Speaks Eng. poorly	1.745***	1.353	3.106***	2.411
	(0.144)	(0.310)	(0.672)	(1.491)
Speaks Eng. well	2.239***	1.700**	3.815***	1.509
	(0.372)	(0.386)	(0.828)	(0.923)
Speaks Eng. fluently	1.793***	1.754***	3.811***	1.621
	(0.165)	(0.347)	(0.952)	(0.984)
Family income	1.000*	1.000	1.000	1.000
	(2.57e-07)	(4.75e-07)	(4.96e-07)	(2.16e-07)
Homeowner	1.170**	1.255**	1.158	1.153**
	(0.0887)	(0.133)	(0.158)	(0.0672)
American Spouse	2.084***	4.932***	3.087***	16.57***
	(0.208)	(0.647)	(0.353)	(1.943)
Child (at least 1)	0.969	0.921	1.034	1.513***
	(0.0843)	(0.0600)	(0.0928)	(0.116)
Chinese PUMA concentration	0.0646**	1.804		
	(0.0858)	(2.316)		
*Chinese Cluster	5.612	0.508		
	(12.38)	(0.444)		
*Chin. % bad Eng. PUMA	4.241	0.811		
	(4.579)	(0.951)		
*Chin. % naturalized PUMA	2.594	0.860		
	(2.147)	(1.491)		
*Chin. % poor PUMA	2.604***	1.900		
	(0.276)	(1.564)		
Indian PUMA concentration			10.02*	6.563***
			(12.85)	(2.314)
*Indian Cluster 2			0.0137**	0.0158*
			(0.0257)	(0.0362)
*Ind. % bad Eng. PUMA			0.0650	2.192
			(0.159)	(2.508)
*Ind. % naturalized PUMA			1.503	0.579
			(2.543)	(0.552)
*Ind. % poor PUMA			3.708	0.433
			(7.107)	(0.409)
Observations	5,565	5,318	2,773	9,780

Robust se in parentheses

*** p<0.01, ** p<0.05, * p<0.1

*Hazard Ratios (HRs) are interpreted in a similar way to odds ratios. HR>1 increases speed of/reduces the time to naturalization by (HR – 1)%, HR<1 decreases speed of naturalization by (1 – HR)%, and HR=1 signifies no influence

Chapter 9: Conclusions and Further Research

The results of my analysis are consistent with the consensus in the naturalization literature. There is a positive and significant correlation between naturalization outcomes and years in the United States, English ability, home ownership, and citizenship status of spouse. In particular, the coefficient of English ability and having an American spouse are large in magnitude and statistically significant.

Education is significant in some regressions but not in others: this result may be due to multi-collinearity issues, since education is highly correlated (0.58) with English ability. Interestingly, in my analysis, having a child is significant and negatively correlated with the probability of being naturalized, while most other studies find a significant positive correlation.

In the tests of the four hypotheses (general, selection, amenities, and social capital), the significance of the neighborhood variable varies considerably across regressions. For Chinese immigrants, the coefficients of the PUMA concentrations are positive and significant in Regressions 1, 5, and 6 but are insignificant in the other regressions. For Indian immigrants, PUMA concentrations are positive and highly significant for the low skilled group in logistic Regressions 3 and 4 but positive and highly significant for the high skilled group in survival Regressions 7 and 8. In general, co-ethnic concentrations for Indian immigrants are significant and positively associated with a shorter time to naturalization in all the survival regressions.

Limitations in empirically defining ethnic neighborhoods and missing data could account for the wide confidence intervals of the regression results. Future studies could address these concerns and expand this research in a few ways. First, experimenting with different definitions of segregation and ethnic concentrations could provide a better sense of the degree to which neighborhood ethnic profiles are truly related to naturalization outcomes. Second, using restricted Census microdata (available only to researchers at the Center for Economic Studies) would permit a more granular analysis of neighborhood outcomes, since PUMAs are probably too large of a spatial unit to capture enclave effects. Third, my thesis focuses only on Indian and Chinese immigrants; for a more complete understanding of ethnic neighborhood and naturalization correlations, future studies should replicate these models for other immigrant groups or ethnicities.

Fourth, the interpretation of these results is constrained by the lack of Census data on the legal status of immigrants and their migratory patterns. In a future study, use of the New Immigrant Survey, a panel dataset surveying legal permanent residents who arrived in 2003, could address these concerns. Access to confidential NIS data would tell us the precise geographic locations and sequential moves of immigrants within the United States. Furthermore, the NIS contains behavioral information missing from the empirical models that could play a pertinent role in explaining naturalization choices of immigrants.

Citizenship has important implications for U.S. society: immigrants who are naturalized contribute more to the American economy and can use their vote to shape U.S. policies in new ways. On a local level, a better understanding of the dynamics between neighborhoods and naturalization outcomes can better aid non-profits in targeting their efforts to integrate immigrants. As a result of these implications, it is important to continue and refine research on the determinants of naturalization.

References

Alba, R.D., J.R. Logan, and K. Crowder. "White Ethnic Neighborhoods and Assimilation: The Greater New York Region, 1980-1990." *Social Forces* 75, no. 3 (1997): 883-912.

Andersson, F., S. Burgess, and J. Lane. . "Do as the Neighbors Do: The Impact of Social Networks on Immigrant Employees." Working Paper 4423, IZA Discussion Paper Series, 2009. http://ftp.iza.org/dp4423.pdf (accessed May 1, 2014).

Aranda, C., D.K. Levy, R. Pitingolo, R. Santos, M.A. Turner, and D. Wissoker. "Housing Discrimination Against Racial and Ethnic Minorities 2012." U.S. Department of Housing and Urban Development. http://www.huduser.org/portal/Publications/pdf/HUD-514_HDS2012.pdf (published June 11, 2013).

Auclair, G. and J. Batalova. "Naturalization Trends in the United States." Migration Information Source, Migration Policy Institute. http://www.migrationpolicy.org/article/naturalization-trends-united-states (published October 24, 2013).

Aydemir, Abduhramman. "Ethnic Enclaves and Human Capital Investments: Evidence from a Natural Experiment." Working Paper, IZA Discussion Paper Series, 2009. http://www.iza.org/conference_files/TAM2009/aydemir_a2793.pdf

Barkan, E. R., & N. Khokhlov. "Socioeconomic data as indices of naturalization patterns in the United States: A theory revisited." *Ethnicity* 7 (1980): 159-190.

Bartel, Anne P. "Where Do the New Immigrants Live?" *Journal of Labor Economics* 74, no. 4 (1989): 371-91.

Bauer, T., G. Epstein, and I. Gang. "Enclaves, Language, and the Location Choice of Migrants." *Journal of Population Economics* 18, no. 4 (2005): 649 – 662.

Bayer, P.J., R. McMillan, K. Rueben. "What Drives Racial Segregation? New Evidence Using Census Microdata." Working Paper 859, Economic Growth Center, Yale University, 2003. http://www.econ.yale.edu/growth_pdf/cdp859.pdf (accessed May 1, 2014).

Bayer, P.J. and S.L. Ross. "Identifying Individual And Group Effects in the Presence of Sorting: A Neighborhood Effects Application." Working Paper 200613, Duke University, 2010. http://digitalcommons.uconn.edu/econ_wpapers/200613 (accessed April 22, 2014).

Bloemraad, I. "Becoming a Citizen in the United States and Canada: Structured Mobilization and Immigrant Political Incorporation." *Social Forces* 85, no. 2 (2006): 667-95.

Beckhusen, Julia and Raymond Florax, Thomas de Graaf, Jacques Poot, and Brigitte Waldorf. "Living and Working in Ethnic Enclaves: Language Proficiency of Immigrants in U.S. Metropolitan Areas." *Papers in Regional Science* 92, no. 2 (2013): 305-28.

Bernard, William S. "Cultural Determinants of Naturalization." *American Sociological Review* 1, no. 6 (1936): 943-53.

Bevelander, P., & Don DeVoretz. *The economics of citizenship*. Malmo University, 2008.

Borjas, George. "Ethnic Enclaves and Assimilation." *Swedish Economic Policy Review* 7 (2000): 89-122.

Borjas, George. "To Ghetto or Not to Ghetto: Ethnicity and Residential Segregation." *Journal of Urban Economics* 44, no. 2 (1998): 228-253.

Chiswick, B. R., & P.W. Miller. "Citizenship in the United States: The Roles of Immigrant Characteristics and Country of Origin." Working Paper 3596, IZA Discussion Paper Series, 2008. http://ftp.iza.org/dp3596.pdf (accessed April 22, 2014).

Chiswick, B.R. and P.W. Miller. "Do Enclaves Matter in Immigrant Adjustment?" *City and Community* 4, no.1 (2005): 5-35.

Chiswick, B.R. and P.W. Miller. "Language in the Immigrant Labor Market: The Immigrant Experience in Canada and the United States." Working Paper 784, Queen's University, 1992. http://qed.econ.queensu.ca/working_papers/papers/qed_wp_784.pdf (accessed April 22, 2014).

Chiswick, B.R., Y. Lee, and P. Miller. "The Determinants of the Geographic Concentration among Immigrants: Application to Australia." *Australasion Journal of Regional Science* 7, no. 2 (2001): 125 – 150.

Cutler, David M. and Edward L. Glaeser. "Are Ghettoes Good or Bad?" *The Quarterly Journal of Economics* 112, no. 3 (1997): 827-72.

Department of Homeland Security. *Yearbook of Immigration Statistics 2011*. Washington D.C.: U.S. Department of Homeland Security. https://www.dhs.gov/yearbook-immigration-statistics (accessed March 28, 2014).

DeVoretz, D. J., & S. Pivneko. "Self-selection, Immigrant Public Finance Performance and Canadian Citizenship." Working Paper 1463, IZA Discussion Papers, 2005. http://ftp.iza.org/dp1463.pdf (accessed April 22, 2014).

Duncan, N. T., & B.S. Waldorf. "Becoming a Citizen: The Role of Immigrant Enclaves." *Cityscape* 11, no. 3 (2009): 5-28.

Dunlevy, J. A. "Nineteenth-Century European Immigration to the United States: Intended versus Lifetime Settlement Patterns." *Economic Development and Cultural Change* 28, no. 1 (1980): 77-90.

Edin, P-A., P. Fredriksson and O. Åslund. "Ethnic Enclaves and the Economic Success of Immigrants: Evidence from a Natural Experiment." *The Quarterly Journal of Economics* 118, no. 1 (2003): 329-57.

Evans, M. D. R. "Choosing to be a Citizen: The Time-Path of Citizenship in Australia." *International Migration Review* 22, no. 2 (1988): 243-264.

Family Search. *United States Naturalization Laws.* https://familysearch.org/learn/wiki/en/United_States_Naturalization_Laws#Naturalization_Laws (last updated October 13, 2010).

Gavit, J.P. *Americans by Choice.* New York, NY: Harper Brothers, 1922.

General Social Survey. *Appendix F- Occupational Classification Codes.* http://publicdata.norc.org/GSS/DOCUMENTS/BOOK/GSS_Codebook_AppendixF.pdf (accessed May 1, 2014).

Gonzalez, A. "Mexican Enclaves and the Price of Culture." *Journal of Urban Economics* 43, no. UE972049 (1998): 273 – 291.

GovHK. *Application for Renunciation of Chinese Nationality.* http://www.gov.hk/en/residents/immigration/chinese/renounce.htm (last updated March 2010).

Grebler, Leo. "The Naturalization of Mexican Immigrants in the United States." *International Migration Review* 1, no. 1 (1966): 17-32.

Hall, Matthew, Audrey Singer, Gordon F. De Jong, and Deborah Roempke Graefe. "The Geography of Immigrant Skills: Educational Profiles of Metropolitan Areas." Metropolitan Policy Program, Brookings Institute. (June 2011).

Immihelp. *Renunciation of Indian Citizenship and Obtaining Surrender Certificate.* http://www.immihelp.com/nri/indian-citizenship-renunciation-surrender-certificate/ (accessed May 2, 2014).

Jaeger, D.A. "Local Labor Markets, Admission Categories, and Immigrant Location Choice." Working Paper, College of William and Mary, 2000. http://www.iza.org/conference_files/amm_2004/jaeger_d138.pdf (accessed April 22, 2014).

Jaramillo, J.J. "Why Isn't My Green Card Green?" *Slate Magazine,* July 4, 2012. http://www.slate.com/articles/life/design/2012/07/green_card_history_u_s_immigrants_vital_document_through_the_years_.html (accessed April 29, 2014).

Jasso, G. and M.R. Rosenzweig. “Family Reunification and the Immigration Multiplier: U.S. Immigration Law, Origin-Country Conditions, and the Reproduction of Immigrants.” *Demography* 23*,* no. 3 (1986): 291-311.

Jasso, G., D.S. Massey, M.R. Rosenzweig, and J.P. Smith. “The New Immigrant Survey-Pilot (NIS-P): Overview and New Findings about U.S. Legal Immigrants at Admission.” *Demography* 37, no. 1 (2000): 127-138.

Jenks, R. “United States Citizenship.” Center for Immigration Studies. http://cis.org/articles/1997/jenks43097.htm (published April 1997).

Jones-Correa, M. *Between Two Nations: The Political Predicament of Latinos in New York City.* New York: Cornell University Press, 1998.

Kerr, William R. “The Agglomeration of U.S. Ethnic Inventors.” In *Agglomeration Economics*, edited by Edward L. Glaeser, 237-276. National Bureau of Economic Research, 2010.

Kritz, Mary, M. Gurak, Douglas T. and Lee, Min-Ah. “Why Immigrants Leave New Destinations and Where Do They Go?” Working Paper CES-WP-13-32, US Census Bureau Center for Economic Studies Paper, 2013. ftp://ftp2.census.gov/ces/wp/2013/CES-WP-13-32.pdf (accessed April 22, 2014).

Lapp, K. “Reforming the Good moral character Requirement for U.S. Citizenship.” Working Paper No. 11-23, New York University School of Law, 2012. http://www.repository.law.indiana.edu/ilj/vol87/iss4/5/ (accessed April 29, 2014).

Li, Wei. "Anatomy of a New Ethnic Settlement: The Chinese Ethnoburb in Los Angeles." *Urban Studies* 35, no. 3 (1998): 479-501,

Light, I., G. Sabagh, M. Bozorgmehr, and C. Der-Martirosian. "Beyond the Ethnic Enclave Economy." *Social Problems* 41, no. 1 (1994): 65-80.

Mandorff, M. “Social Networks, Ethnicity, and Occupation.” PhD diss., University of Chicago, 2007. http://gradworks.umi.com/32/72/3272969.html

Mazzolari, F. “Naturalization and its Determinants among Immigrants from Latin America: The Role of Dual Citizenship Rights.” In *Immigrants and Minorities, Politics and Policy,* edited by David L. Leal and Stephen J. Trejo, 133-150. New York and Heidelberg: Springer, 2011.

McMahon, S. “Echoes of Freedom: South Asian Pioneers in California, 1899 – 1965.” *The Library, University of California, Berkeley.* http://www.lib.berkeley.edu/SSEAL/echoes/chapter10/chapter10.html (accessed April 29, 2014).

McManus, W.S. “Labor Market Effects of Language Enclaves: Hispanic Men in the United States.” *Journal of Human Resources* 25, no. 2 (1990): 228-252.

NRI Realty News. *Difference between NRI, PIO, and OCI.* http://www.nrirealtynews.com/nri-info/difference-between-nri-pio-oci.php (accessed May 4, 2014).

Nowrasteh, Alex. "The Rise of Asian Immigration." *Forbes Magazine,* July 9, 2012, http://www.forbes.com/sites/alexnowrasteh/2012/07/09/the-rise-of-asian-immigration/ (accessed May 4, 2014).

On the Issues. *Al Gore on Immigration: 2000 Democratic Nominee for President; Former Vice President.* http://www.ontheissues.org/celeb/Al_Gore_Immigration.htm (accessed May 1, 2014).

Painter, G., and Z. Yu. "Immigrants and Housing Markets in Mid-size Metropolitan Areas." *International Migration Review* 44, no. 2 (2010): 442-476.

Pamuk, Ayse. "Geography of Immigrant Clusters in Global Cities: A Case Study of San Francisco, 2000." *International Journal of Urban and Regional Research* 28, no. 2 (2004): 287-307.

Pastor, M. and J. Scoggins. "Citizen Gain: The Economic Benefits of Naturalization for Immigrants and the Economy." University of Southern California, Center for the Study of Immigrant Integration. http://csii.usc.edu/documents/citizen_gain_web.pdf. (published December 2012).

Pew Research Social & Demographic Trends. *The Rise of Asian Americans.* http://www.pewsocialtrends.org/2012/06/19/the-rise-of-asian-americans/ (published June 19, 2012).

Portes, A. "The Social Origins of the Cuban Enclave Economy of Miami." *Pacific Sociological Association* 30, no. 4 (1987): 340- 372.

Portes, A. and J.W. Curtis. "Changing Flags: Naturalization and Its Determinants among Mexican Immigrants." *International Migration Review* 21*,* no. 2 (1987): 352-371.

Ruggles, S., J.T. Alexander, K. Genadek, R. Goeken, M.B. Schroeder, and M. Sobek. *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

Scott, D.M., P.A. Coomes, and A.I. Izyumov. "The Location Choice of Employment-Based Immigrants among U.S. Metro Areas." *Journal of Regional Science* 1*,* no. 45 (2005): 113-145.

Seitles, Marc. "The Perpetuation of Residential Racial Segregation in America: Historical Discrimination, Modern Forms of Exclusion, and Inclusionary Remedies." *Journal of Land Use & Environmental Law.* 1996. http://dev.law.fsu.edu/journals/landuse/Vol141/seit.htm#FNR*

United States Citizenship and Immigration Services. *Adjustment of Status.* http://www.uscis.gov/green-card/green-card-processes-and-procedures/adjustment-status (last updated March 30, 2011).

United States Citizenship and Immigration Services. *Naturalization for Spouses of U.S. Citizens.* http://www.uscis.gov/us-citizenship/citizenship-through-naturalization/naturalization-spouses-us-citizens (last updated January 22, 2013).

United States Citizenship and Immigration Services. *Origins of the Federal Immigration Service.* http://www.uscis.gov/history-and-genealogy/our-history/agency-history/origins-federal-immigration-service (last updated September 23, 2013).

United States Citizenship and Immigration Services. *Origins of the Naturalization Civics Test.* http://www.uscis.gov/history-and-genealogy/history-and-genealogy-news/origins-naturalization-civics-test (last updated January 15, 2013).

United States Citizenship and Immigration Services. *Why Some C-Files May Lack a Declaration of Intention.* http://www.uscis.gov/history-and-genealogy/genealogy/genealogy-notebook/why-some-c-files-may-lack-declaration-intention (last updated June 27, 2013).

United States Department of State: Bureau of Consular Affairs. *Directory of Visa Categories.* http://travel.state.gov/content/visas/english/general/all-visa-categories.html (accessed April 29, 2014).

United States Department of State: Bureau of Consular Affairs. *Visa Bulletin for January 2013.* http://travel.state.gov/content/visas/english/law-and-policy/bulletin/2013/visa-bulletin-for-january-2013.html (accessed April 29, 2014).

Vigdor, J.L. *From Immigrants to Americans: The Rise and Fall of Fitting In.* New York, NY: Rowman & Littlefield Publishers, 2009.

Venturini, A. and R. Faini. “Home Bias and Migration: Why is Migration Playing a Marginal Role in the Globalization Process?” Working Paper, Center for Household, Income, Labor, and Demographic Economics, 2001. http://www.child-centre.it/papers/child27_2001.pdf

Yang, P. Q. “Explaining Immigrant Naturalization.” *The International Migration Review* 28*,* no. 3 (1994): 449-77.

Zavodny, Madeline. “Determinants of Recent Immigrants’ Locational Choices.” Working Paper 98-3, Federal Reserve Bank of Atlanta, 1998. http://www.frbatlanta.org/filelegacydocs/wp9803.pdf (accessed April 22, 2014).

Zhang and Associates, P.C. *Frequently Asked Questions about Adjustment of Status.* http://www.hooyou.com/i-485/faq.html (last updated October 10, 2012

Appendix A

Based on American Community Survey 2011 data, the Migration Policy Institute estimates the total Foreign-born population in the United States to be 40.4 million (Auclair and Batalova, 2013).

Based on an IPUMS sample of the ACS 2011, Chinese account for 3.941% of the total Foreign Born pool, while Indians account for 4.345%.

The following population estimates of Indian and Chinese foreign born result from this information:

$$Indian\_Sample : 71{,}534 \div 1{,}646{,}338 = 4.345\%$$

$$Chinese\_Sample : 64{,}891 \div 1{,}646{,}338 = 3.941\%$$

$$Indian\_Population : 40{,}400{,}000 \times 0.04345 = 1{,}755{,}380$$

$$Chinese\_Population : 40{,}400{,}000 \times 0.03941 = 1{,}592{,}164$$

Based on 2012 data, The Department of Homeland Security's Office of Immigration Statistics released a 2014 report estimating the population of temporary visa holders (not including short term visitors) by country of origin and visa category:

Temporary Visa Holders

Country	Total	Temporary Workers	Students	Exchange Visitors
India	430,000	320,000	100,000	10,000
China	210,000	30,000	150,000	30,000

$$Indian\_Temp\% = 430{,}000/1{,}755{,}380 = 24.5\%$$

$$Chinese\_Temp\% = 210{,}000/1{,}592{,}164 = 13.1\%$$

Appendix B

Summary Statistics for Full Sample: Yrs. Nat. 2008 – 2011

Variable	n	Mean	Std. Dev.	Median	Min	Max
Citizen	23,589	0.22	0.41	0.00	0.00	1.00
Age	23,589	45.26	13.37	41.00	24.00	94.00
Years in USA	23,589	13.31	6.87	11.00	5.00	75.00
Female	23,589	0.49	0.50	0.00	0.00	1.00
Education Level	23,589	5.34	2.03	6.00	0.00	7.00
English Ability	23,589	2.17	1.05	3.00	0.00	3.00
Family Income	23,589	$120,558	$146,762	$94,525	-$17,800	$1,500,000
American Spouse	23,589	0.60	0.90	0.00	0.00	2.00
Child	23,589	0.70	0.46	1.00	0.00	1.00
Homeowner	23,589	0.67	0.47	1.00	0.00	1.00
Indian PUMA Concentration	23,589	0.03	0.04	0.02	0.00	0.22
Chinese PUMA Concentration	23,589	0.06	0.09	0.02	0.00	0.40
Indian Dissimilarity	23,589	0.51	0.15	0.48	0.00	0.99
Chinese Dissimilarity	23,589	0.46	0.16	0.44	0.00	0.99
% naturalized PUMA	23,589	0.12	0.08	0.11	0.00	0.37
% bad Eng. PUMA	23,589	0.14	0.11	0.11	0.00	0.55
% poor PUMA	23,589	0.11	0.65	0.91	0.19	0.44

Summary Statistics for Chinese Sub-Sample: Yrs. Nat. 2008 - 2011

Variable	n	Mean	Std. Dev.	Median	Min	Max
Citizen	10,970	0.19	0.39	0.00	0.00	1.00
Age	10,970	48.29	13.95	45.00	24.00	94.00
Years in USA	10,970	13.81	7.22	12.00	5.00	75.00
Female	10,970	0.52	0.50	1.00	0.00	1.00
Education Level	10,970	4.75	2.26	5.00	0.00	7.00
English Ability	10,970	1.66	1.11	2.00	0.00	3.00
Family Income	10,970	$98,387	$141,305	$68,300	-$8,741	$1,500,000
American Spouse	10,970	0.56	0.88	0.00	0.00	2.00
Child	10,970	0.66	0.47	1.00	0.00	1.00
Homeowner	10,970	0.65	0.48	1.00	0.00	1.00
Chinese PUMA Concentration	10,970	0.09	0.11	0.04	0.00	0.40
Chinese Dissimilarity	10,970	0.45	0.16	0.42	0.00	0.99
% naturalized PUMA	10,970	0.14	0.09	0.13	0.00	0.37
% bad Eng. PUMA	10,970	0.18	0.13	0.14	0.00	0.55
% poor PUMA	10,970	0.12	0.07	0.11	0.02	0.44

Summary Statistics for Indian Sub-Sample: Yrs. Nat. 2008 - 2011

Variable	n	Mean	Std. Dev.	Median	Min	Max
Citizen	12,619	0.24	0.43	0.00	0.00	1.00
Age	12,619	42.62	12.24	39.00	24.00	94.00
Years in USA	12,619	12.87	6.53	11.00	5.00	65.00
Female	12,619	0.47	0.50	0.00	0.00	1.00
Education Level	12,619	5.84	1.65	6.00	0.00	7.00
English Ability	12,619	2.62	0.75	3.00	0.00	3.00
Family Income	12,619	$139,831	$148,689	$110,612	-$17,800	$1,500,000
American Spouse	12,619	0.63	0.91	0.00	0.00	2.00
Child	12,619	0.74	0.44	1.00	0.00	1.00
Homeowner	12,619	0.68	0.47	1.00	0.00	1.00
Indian PUMA Concentration	12,619	0.04	0.05	0.02	0.00	0.22
Indian Dissimilarity	12,619	0.50	0.15	0.47	0.19	0.99
% naturalized PUMA	12,619	0.11	0.07	0.09	0.00	0.37
% bad Eng. PUMA	12,619	0.11	0.08	0.09	0.00	0.55
% poor PUMA	12,619	0.10	0.06	0.08	0.02	0.43

% Distribution of Naturalized Citizens by Individual Variables: Full Sample (2008 – 2011)		
Variable	**Base**	**% Naturalized**
Citizen	23,589	21.75%
Age		
24 to 35	5,875	14.57%
36 to 45	8,682	25.62%
46 to 55	4,457	24.19%
56 to 65	2,278	23.79%
66 to 75	1,308	22.55%
76 to 85	754	15.25%
86+	235	8.51%
Years in USA		
5 to 15	17,851	20.62%
16 to 25	4,407	26.78%
26 to 35	927	22.87%
36 to 45	268	18.66%
46 to 55	94	7.45%
56 to 65	37	0.00%
66 to 75	5	0.00%
Female	11,604	23.35%
Education Level		
0	1,144	14.69%
1	535	13.08%
2	992	14.62%
3	1,425	18.25%
4	3,512	19.85%
5	783	27.71%
6	5,271	26.31%
7	9,927	22.02%
English Ability		
0	2,613	11.17%
1	3,573	19.59%
2	4,535	22.87%
3	12,868	24.10%
Family Income		
<$20,000	2269	16.75%
$20,000-$49,999	3959	18.44%
$50,000-$99,000	6198	18.83%
$100,000-$199,000	8144	24.18%
$200,000-$300,000	1977	29.99%
>$300,000	1042	27.93%
American Spouse	7,417	48.70%
Child	16,561	23.46%
Homeowner	15,711	25.70%

% Distribution of Naturalized Citizens by Individual Variables: Sub-Samples				
	Chinese		**Indian**	
Variable	**Base**	**% Naturalized**	**Base**	**% Naturalized**
Citizen	10,968	18.95%	12,616	24.19%
Age				
24 to 35	1,711	12.33%	4,164	15.49%
36 to 45	3,905	20.10%	4,777	30.12%
46 to 55	2,650	20.38%	1,804	29.82%
56 to 65	1,258	20.83%	1,020	27.45%
66 to 75	769	24.45%	539	19.85%
76 to 85	515	15.34%	239	15.06%
86+	162	8.02%	73	9.59%
Years in USA				
5 to 15	7,806	17.13%	10,045	23.33%
16 to 25	2,547	24.50%	1,857	29.94%
26 to 35	398	23.12%	529	22.68%
36 to 45	119	17.65%	149	19.46%
46 to 55	64	6.25%	30	10.00%
56 to 65	31	0.00%	6	0.00%
66 to 75	5	0.00%	0	0.00%
Female	5,718	20.67%	5,886	25.96%
Education Level				
0	824	14.81%	320	14.38%
1	391	14.32%	144	9.72%
2	779	15.79%	213	10.33%
3	902	14.08%	523	25.43%
4	2,215	16.98%	1,297	24.75%
5	486	26.13%	297	30.30%
6	1,433	24.70%	3,838	26.92%
7	3,940	20.13%	5,987	23.27%
English Ability				
0	2,192	11.91%	421	7.36%
1	2,729	18.69%	844	22.51%
2	2,687	21.21%	1,848	25.27%
3	3,362	21.92%	9,506	24.87%
Family Income				
<$20,000	1725	17.10%	544	15.63%
$20,000-$49,999	2584	16.02%	1375	22.98%
$50,000-$99,000	2818	17.85%	3380	19.64%
$100,000-$199,000	2871	21.46%	5273	25.66%
$200,000-$300,000	687	26.93%	1290	31.63%
>$300,000	285	22.81%	757	29.85%
American Spouse	3,230	36.75%	4,187	57.92%
Child	7,208	19.06%	9,353	26.86%
Homeowner	7,171	21.68%	8,540	29.07%
LowFB_State	402	15.17%	605	19.83%
LowChin_State	309	15.21%	-	-
LowInd_State	-	-	108	13.89%

% Distribution of Chinese and Indian in Ethnic Cluster by Individual Variables				
	Chinese		**Indian**	
Variable	**Base**	**% in Chin. Cluster**	**Base**	**% in Ind. Cluster**
Citizen	2,077	26.53%	3,052	20.18%
Age				
24 to 35	1,711	28.93%	4,164	20.77%
36 to 45	3,905	24.46%	4,777	20.26%
46 to 55	2,650	27.85%	1,804	26.88%
56 to 65	1,258	30.21%	1,020	28.43%
66 to 75	769	24.58%	539	26.53%
76 to 85	515	25.63%	239	31.38%
86+	162	18.52%	73	21.92%
Years in USA				
5 to 15	7,806	26.36%	10,045	21.62%
16 to 25	2,547	26.46%	1,857	24.66%
26 to 35	398	27.39%	529	28.54%
36 to 45	119	36.13%	149	30.20%
46 to 55	64	31.25%	30	46.67%
56 to 65	31	38.71%	6	33.33%
66 to 75	5	60.00%	0	0.00%
Female	5,718	26.18%	5,886	22.49%
Education Level				
0	824	31.55%	320	28.44%
1	391	32.23%	144	29.17%
2	779	27.86%	213	32.86%
3	902	30.38%	523	32.50%
4	2,215	33.91%	1,297	28.60%
5	486	23.46%	297	25.25%
6	1,433	22.33%	3,838	20.01%
7	3,940	21.75%	5,987	20.96%
English Ability				
0	2,192	27.42%	421	26.60%
1	2,729	30.52%	844	31.16%
2	2,687	25.27%	1,848	23.38%
3	3,362	23.97%	9,506	21.41%
Family Income				
<$20,000	1,725	29.33%	544	37.13%
$20,000-$49,999	2,584	32.51%	1375	33.89%
$50,000-$99,000	2,818	31.01%	3380	26.01%
$100,000-$199,000	2,871	19.44%	5273	17.18%
$200,000-$300,000	687	10.04%	1290	14.73%
>$300,000	285	25.26%	757	26.29%
American Spouse	3,230	5.23%	4,187	22.21%
Child	7,208	25.83%	9,353	21.31%
Homeowner	7,171	27.29%	8,540	22.31%

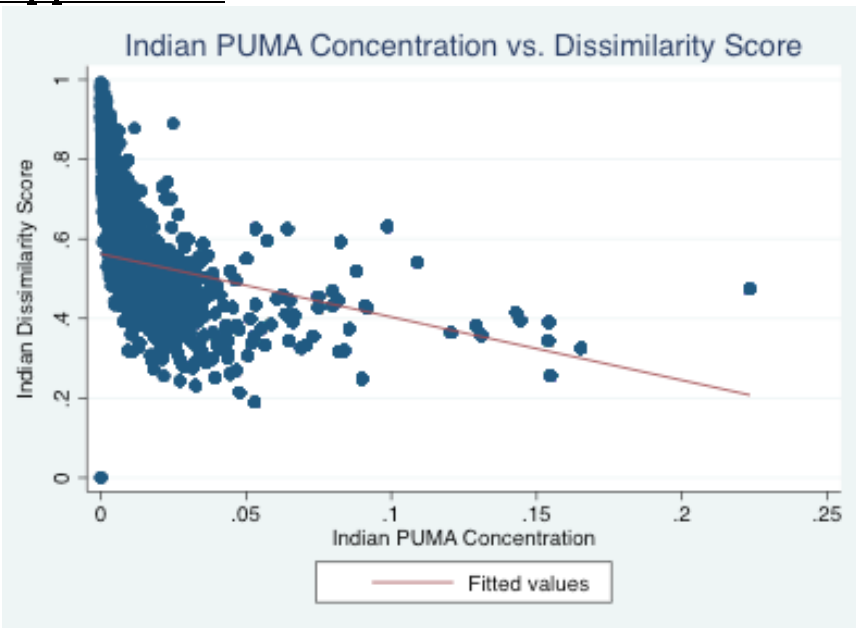
Appendix C

	age	YrsInUSA	englis~y	educ_l~l	family~e	Spouse~t	female
age	1.0000						
YrsInUSA	0.6463	1.0000					
english_ab~y	-0.3815	-0.0187	1.0000				
educ_level	-0.3560	-0.1095	0.6612	1.0000			
family_inc~e	-0.0235	0.0526	0.1946	0.2221	1.0000		
Spouse_Cit	0.1015	0.2516	0.0902	0.0486	0.0411	1.0000	
female	-0.0080	-0.0248	-0.1056	-0.1485	-0.0152	-0.0081	1.0000
child	-0.2765	-0.2735	0.0290	0.0260	0.0309	0.0605	0.0241
homeowner	0.0213	0.1710	0.1659	0.1286	0.0635	0.2658	0.0309
FB_PUMAperc	0.1378	0.0261	-0.3072	-0.2550	-0.0883	-0.0327	0.0221
Ind_PUMAperc	-0.1101	-0.0870	0.1274	0.1432	0.0620	0.0096	-0.0155
Chin_PUMAp~c	0.1500	0.0431	-0.3406	-0.2594	-0.0831	-0.0292	0.0306
Nat_PUMAperc	0.1478	0.0498	-0.2917	-0.2391	-0.0694	-0.0013	0.0272
Ind_Diss	0.0467	0.0249	-0.0977	-0.1386	-0.0689	-0.0226	-0.0008
Chin_Diss	-0.0242	-0.0036	0.0532	-0.0184	-0.0308	-0.0087	-0.0144
FB_Diss	-0.0212	-0.0044	0.0333	0.0170	-0.0006	-0.0221	-0.0125
BadEng_PUM~c	0.1718	0.0374	-0.3920	-0.3478	-0.1322	-0.0449	0.0258
BPL_PUMAperc	0.1125	0.0204	-0.2646	-0.2751	-0.1407	-0.0904	0.0089

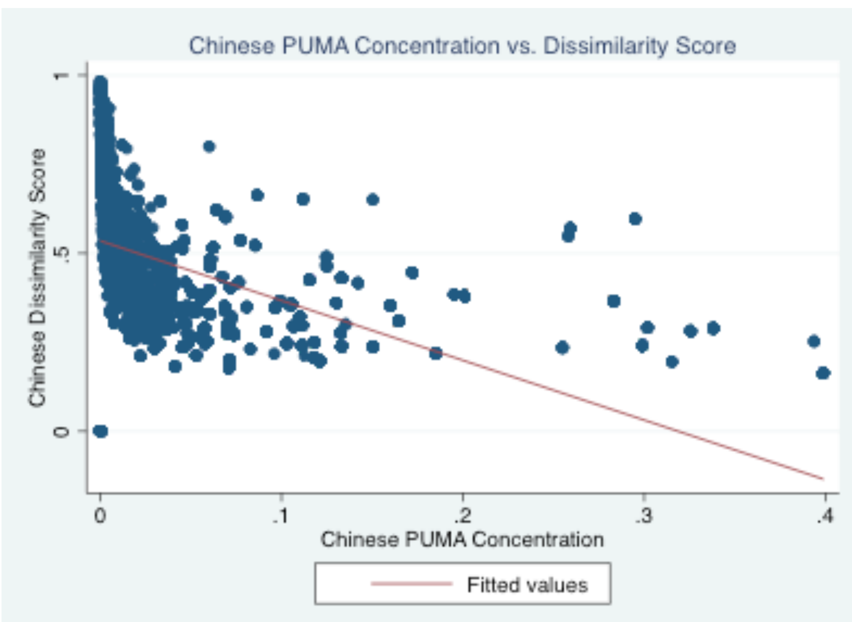
	child	homeow~r	FB~Aperc	In~Aperc	Ch~Aperc	Nat_PU~c	Ind_Diss
child	1.0000						
homeowner	0.1801	1.0000					
FB_PUMAperc	-0.0143	-0.1408	1.0000				
Ind_PUMAperc	0.0917	0.0258	0.3455	1.0000			
Chin_PUMAp~c	-0.0200	-0.1467	0.6782	0.0793	1.0000		
Nat_PUMAperc	0.0056	-0.0856	0.9245	0.3291	0.7527	1.0000	
Ind_Diss	-0.0552	-0.0214	-0.3508	-0.4842	-0.1848	-0.3625	1.0000
Chin_Diss	-0.0344	-0.0008	-0.5240	-0.3359	-0.4689	-0.5835	0.6816
FB_Diss	-0.0335	-0.0499	-0.3982	-0.2024	-0.1570	-0.4033	0.4312
BadEng_PUM~c	-0.0385	-0.1706	0.9136	0.0733	0.6922	0.7993	-0.1086
BPL_PUMAperc	-0.1223	-0.2376	0.1240	-0.3308	0.0726	-0.0477	0.4520

	Chin_D~s	FB_Diss	BadEng~c	BPL_PU~c
Chin_Diss	1.0000			
FB_Diss	0.5173	1.0000		
BadEng_PUM~c	-0.3352	-0.2788	1.0000	
BPL_PUMAperc	0.4174	0.2385	0.3574	1.0000

Appendix D



Indian PUMA Concentration vs. Dissimilarity Score
Indian Dissimilarity Score
0
.2
.4
.6
.8
1
0
.05
.1
.15
.2
.25
Indian PUMA Concentration
Fitted values



Chinese PUMA Concentration vs. Dissimilarity Score
Chinese Dissimilarity Score
0
.5
1
0
.1
.2
.3
.4
Chinese PUMA Concentration
Fitted values