Who Naturalizes?: A Community-Based Analysis of the Factors which Determine Naturalization Rate among Ethnic Groups

Elizabeth Deak
Senior Honors Thesis

Professor John Roos, Advisor

April 10, 2007
Section I: Introduction and Abstract

What makes a person a US citizen? The vast majority of the current US population inherits citizenship at birth, either by being born on US soil or by virtue of having at least one citizen parent. But a growing percentage of the population entered the US for the first time with a passport from another country, and must go through the process of naturalization in order to become a citizen. In the 2000 US Census, the Census Bureau reported that there were over 31 million foreign-born persons residing in the U.S., (11.1 percent of the total population), but that only 12.5 million, or 40.3 percent of the foreign-born population, were naturalized citizens. Latin American immigrants, who make up more than half of the foreign-born population, naturalize at even lower rates—only 31 percent are US citizens. What factors determine who naturalizes and who does not? Past studies have indicated that socioeconomic and demographic factors such as income, education, English fluency, and country of birth may all play a significant role, as well as federal naturalization policy and public warmth towards immigration in general. My senior thesis presents a model of naturalization that is based on the resources available to an immigrant, largely from three sources: individual skills and characteristics, communities, and the government. While I address the impact of individual characteristics such as socioeconomic status and of government policy on naturalization, my focus is on quantitatively studying the effects of community characteristics on community naturalization rate, grouped by country of birth, because it is by far the most underdeveloped area of the existing naturalization literature. For my method of analysis I chose ordinary least squares (OLS) regression on the Current
Population Survey (CPS) March supplement data from 1994 through 2005. I also hope to explain how changes in US policy, and specifically the 1996 Illegal Immigration Reform and Immigrant Responsibility Act, affected naturalization in the US by impeding the access of certain groups to important resource. The emphasis throughout is on Latin American, and specifically Mexican, immigration, since these immigrants make up such a significant portion of the foreign-born population and therefore influence heavily any theories on naturalization in the US.

The thesis itself is divided into three major parts—theory, data, and conclusions. The “Theory” section is intended to provide the reader with some background on the issue of naturalization in the US, including a review of the relevant scholarly literature, a brief history and summary of recent changes in naturalization policy, and a statement of my theory and hypothesis. The “Data” section explains my data and methods, why I chose to conduct my analysis in the way I did, and provides an overview of the limitations of my data and methods before presenting my actual findings. The data and results are split into two parts: one for the global model, the other for the seven regional models I chose to run and the Mexican case. Finally, in “Conclusions” I reevaluate my theory in light of my results and emphasize any original or unusual findings. I also evaluate possible shortcomings or omissions in my analysis for future research. The crux of my research is contained primarily in the last part of the “Theory” section and the “Data” section, but the rest of the paper contains background information and acknowledgements of what I was unable to accomplish that enrich the discussion of what I was able to show, and merit their inclusion. The appendices contain additional resources
on naturalization literature, data tables, an overview of OLS multiple regression and Stata procedures, and a glossary of terms. Comments, questions, or requests for clarification and/or data sources and tables are welcome, either addressed to the author in person or via email at edeak@nd.edu.
Section II: Theory

§1: Literature Review

The only conclusion one can draw with certainty from a thorough review of the existing literature on American naturalization is that there is little widespread agreement on why immigrants choose to naturalize. This is in part due to the fact that the amount of naturalization literature available is still relatively small and underdeveloped, especially in comparison to the immigration literature published in the US over the past fifty years (DeSipio 1987, Pantoja 2006). Most of the naturalization literature can be divided into two broad categories: heavily quantitative work that analyzes a specific data set in order to identify variables which influence naturalization, or work whose primary goal is to establish some comprehensive theory of naturalization, either supported by quantitative analysis and/or qualitative argument. This afforded me an enormous advantage in my research: although no single author has examined the effects of U.S. naturalization policy quantitatively across as comprehensive a data set as I intended to, several authors have tackled pieces of this “big picture,” identifying factors that may be correlated with naturalization using restricted data sets, or offering broad but yet untested theories. Their theories and methods offer examples that I can adapt to my thesis as necessary, giving my work a basis in precedent while allowing plenty of opportunity for original findings and analysis.

1 I include a table at the end of this section summarizing the key characteristics of several representative or noteworthy articles to provide the reader additional clarity in the muddle of quantitative vs. qualitative vs. theory-based studies (Table 1)
It is most instructive to start with the data-based literature, because although the findings are often contradictory they provide the basis for the broader theories. One of the earliest semi-quantitative studies of Mexican naturalization was published in 1966 by Leo Grebler. While conducting a broader study of the socioeconomic status of Mexican-American immigrants in the Southwest, Grebler found that illiteracy and low socioeconomic status (SES), as well as California residency, tended to indicate a lower than normal rate of naturalization. Grebler’s study seems to be indicative of much of the work done on naturalization before the late 1980s. Louis DeSipio mentions Grebler’s study, among many others, in an excellent review of most of the work done in the area of U.S. naturalization policy from about 1925 through 1986, citing the historically negative correlation between blue-collar jobs or low incomes and propensity to naturalize, as well as the historically low percentage of Mexicans who naturalize. DeSipio also notes, however, that the bulk of this research has been done in at best a semi-quantitative manner, and there was a real need for more quantitative research on American naturalization. The lack of widely available complete datasets on the foreign-born population probably contributed to the dearth of such studies.

The release of the 1990 U.S. Census brought about a new wave of truly quantitative studies on naturalization. For the first time, the Census Bureau released the IPUMS records (Individual Level Public Use Microdata, offering detailed individual-level data for 5% and 1% samples of the population) pre-formatted in an easily downloadable, widely accessible data file that saved researchers the work of cutting and pasting hundreds of files or manually entering data into data analysis programs. The Center for Immigration Studies (CIS) released a 1996 article, “Embracing America: A
Look at Which Immigrants Become Citizens,” that used logit analysis to determine which variables influence naturalization rates. The CIS found that in general, higher levels of education, higher household income, English proficiency, non-receipt of public assistance, and more highly skilled occupations correspond to higher rates of naturalization. However, even after controlling for all these factors, national differences do not entirely disappear. In particular, the CIS study concluded that a Mexican college graduate is still less likely to attain U.S. citizenship than a high-school dropout from China solely on the basis of differences in country of origin. The study, however, does not discuss the reasons for this disparity in depth. Further, the CIS used only data from the 1990 census, and addressed the broader question of “who naturalizes?” rather than focusing on differences caused by country of birth (with the exception of Mexico).

Along similar lines is the article Taking the Oath: An Analysis of Naturalization in California and the United States produced by the Public Policy Institute of California. It couples the 1990 individual census data with 1996 and 1997 Current Population Survey (CPS) data to estimate trends in naturalization propensities. The article is mainly concerned with naturalization in California and its policy implications. It does, however, identify key contributing factors to naturalization rates such as level of adaptation to American society, strength of social networks, time spent in the US, education attainment, English fluency, income, and marriage to a US citizen, and supports the hypothesis that Latino immigrants and immigrants from California naturalize at a lower rate than the legal permanent resident (LPR) population at large. A 2003 article by Heather Antecol, Deborah A. Cobb-Clark, and Stephen J. Trejo also used 1990 and 1991 census data from the U.S., comparing it to similar data from Australia and Canada.
Although immigrants to Australia and Canada tended to have higher levels of income, education, English fluency, and naturalization rates, after excluding Latin American immigrants from all three countries the numbers were comparable—that is, the levels of English fluency, income, education, and naturalization rate among all immigrants were approximately the same. This suggests that although income and education affect the average immigrant’s propensity to naturalize, the U.S.’s comparatively low average naturalization rates may be because Mexican immigrants represent such a large portion of its total foreign-born population rather than attributable to any significant policy difference.

Later authors used different datasets, but with similar results. Michael Jones-Correa studies the correlation between naturalization and voting in a 2001 article, with a focus on Latin American immigrants. Jones-Correa tests a limited set of variables (education, length of stay, sex, race, California residency, country allows dual nationality) in a logit regression model over the November 1996 CPS data on Latin American immigrants—a pool of 4,436 adults, 1,076 of whom had naturalized. Jones-Correa provides an excellent model but uses only a very small sample size (in comparison to a total Latin American population of approximately 16 million in 2000) and does not compare his data to the pool of immigrants at large or from other regions. He finds that education and year of entry are positively correlated with naturalization, but sex and race have no effect. Other authors have argued that socioeconomic status has only an indirect influence on an immigrant’s decision to naturalize: an immigrant’s objective status is connected to their satisfaction in general with life in the US, which in turn determines propensity to naturalize (Massey 2006, Portes 1987).
On the theory-based end of the spectrum, most of the work done, like most of the data, focuses on individual characteristics. Bloemraad (2006) instructively breaks down this group of theories into two major categories: those that consider citizenship adoption as the product of individual cost/benefit calculations, and those that assume individual immigrants possess different skills and resources which are necessary to acquire citizenship. The first group is dominated by those who focus on the costs of naturalization, as well as the costs associated with not becoming a citizen, imposed by changing government policy such as the Immigration Reform and Control Act (IRCA) of 1986 and the Illegal Immigration Reform and Immigrant Responsibility Act of 1996 (DeSipio and de la Garza 1997, DeSipio 2004, Bartley 2001, Portes 1987). Rather than discouraging illegal immigrants, such theories contend that the IRCA’s increased border control and employer sanctions simply prevented illegal immigrants from returning to Mexico or finding jobs with any but the lowest-paying employers and contributed to the creation of a Mexican-American underclass largely cut off from the rest of American society (Orrenius 2003, Durand 1999, 2001). Others argue that it is bureaucracy and red tape in an unfamiliar system, the “long gray welcome” of INS paperwork and the costs that come with it that delay or discourage immigrants from becoming citizens (Murguia 2005).

Most of the studies mentioned above, if they adopt any theory, tend to take the second view focusing on naturalization as a function of individual skills and assets: more education, more income, or a better job equip an immigrant with the tools he needs to naturalize. One of the more vocal members of this group is George Borjas. Borjas has written a wide array of scholarly articles and editorials based on the premise that the
disparity in wages and social status between recent immigrants (especially those from
Mexico and Latin America) and the native-born American population is caused primarily
by recent immigrants’ lack of skills which are valuable in the American job market. He
argues that these immigrants naturalize at lower rates because they enter the US lacking
both the individual socio-political resources to achieve naturalization and the
socioeconomic attributes like education or job skills that would allow them the upward
mobility they need to attain such resources in the future. Further, he finds that Mexican
immigrants in particular have a relatively low socioeconomic status compared to
immigrants from other countries, even after correcting for variables like length of
residence in the U.S. However, since Borjas’ primary focus is on the socioeconomic
disparity between the native-born and foreign-born populations, his data is mostly
comparison tables of socioeconomic characteristics like income or occupation between
the native and foreign-born populations, broken down by country of birth.

A relatively new theory in the realm of naturalization literature is what Van Hook,
Brown, and Bean refer to as “social-contextual factors.” The concept is hinted at by
earlier authors who cite “roots in the US” or “satisfaction with life in the US” as positive
predictors of naturalization. Essentially, social-contextual (SC) theory looks at
naturalization as a function of the characteristics of one’s community and social network:
attitudes about the value of naturalization and towards the US in general, political power
or voice, and community access to resources, among other things. Borjas touches upon
the importance of community characteristics with his theory of “ethnic enclaves”; that is,
immigrant groups with large populations in the U.S. tend to remain clustered in certain
areas or regions of the country and in doing so create isolated ethnic environments which
may limit socioeconomic mobility. He considers as “ethnic enclaves” both large, fairly
dense groups such as the Mexican-American population in the Southwest and smaller,
more widely dispersed groups such as certain Mexican-American communities in the
Northeast, as long as they have the key characteristic of being relatively isolated from
American society at large. Such enclaves impede the assimilation process and prevent
immigrants from attaining skills like English fluency that are necessary for most kinds of
employment. Authors like Bloemraad (2006), Pantoja (2006), Basok (2004), and Van
Hook (2006) focus specifically on the social factors which encourage or discourage
immigrants from becoming US citizens. Exclusion from or integration into local
communities (Basok, Bloemraad), the attitude of the state or community towards
immigrants (Van Hook, Basok), the political mobilization of immigrants in the
community (Pantoja 2006, Bloemraad 2006), as well as the resources available to
immigrants within an ethnic community setting, all play a major role in the naturalization
process.

Perhaps the most comprehensive work (for my purposes) on US naturalization
does not address Latino, or even modern, immigrants at all. Irene Bloemraad’s
“Citizenship Lessons from the Past: The Contours of Immigrant Naturalization in the
Early 20th Century” uses logistic regression to analyze Census IPUMS microfile data for
the foreign-born population from 1900 to 1920, focusing on four possible explanations
for varying rates of naturalization: individual resources and skills, regulatory and
bureaucratic barriers to citizenship, relative costs and benefits of citizenship, and the
degree of political mobilization directed to immigrants. She finds that in this period,
place of residence in the U.S. is the most relevant predictor of naturalization. Prior to the
1906 Naturalization Act, naturalization was not a federalized process, and the rules varied from state to state. Bloemraad found that immigrants were more likely to naturalize if they lived in a state with fairly friendly naturalization laws, which usually reflected the state’s general openness to immigration as well as the degree of local political immigrant mobilization. However, Bloemraad also found that the effect of residence on propensity to naturalize decreased over time as the 1906 Act was implemented across the country, suggesting that major policy changes can significantly impact the factors which most strongly influence naturalization.

What I hope to add to the above literature is a similarly comprehensive theory of naturalization which accounts for the disparate rates of naturalization among groups of immigrants in the 1994-2005 period, and addresses the impact of the 1996 changes in U.S. immigration and naturalization policy in a quantitative manner. Most of the existing quantitative studies, especially those analyzing the 1990 Census data, only identify factors which may influence naturalization with little additional explanation: e.g., immigrants with greater levels of English fluency naturalize at higher rates simply because it’s obvious that English fluency makes it easier to naturalize, with no more sophisticated explanation of the social and political factors involved. Other studies work with very small data sets, or exclude populations that I believe to be critical in fully understanding naturalization in the United States. Van Hook’s article, for example, has a sample set of only 969 adults and specifically excludes those born in Mexico, even though in 2000 nearly 30% of the foreign-born population report Mexican birth (US Census Bureau, 2000). To my knowledge, there has been no recent work on
naturalization in the United States that looks at the entire foreign-born population over a
ten-year or greater period, as my research does.
<table>
<thead>
<tr>
<th>Table 1: Review of Selected Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic premise</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Irene Bloemraad, “Becoming a Citizen in the US and Canada: Structured Mobilization and Immigrant Political Incorpor.” (2006)</td>
</tr>
<tr>
<td>Irene Bloemraad, “Citizenship Lessons from the Past: The Contours of Immigrant Nat. in the Early 20th Century” (2006)</td>
</tr>
<tr>
<td>Louis DeSipio, “From Naturalized Citizen to Voter: Context of Natz. and Electoral Participation in Latino Communities.”</td>
</tr>
<tr>
<td>Leo Grebler, “The Naturalization of Mexican Immigrants in the US” (1966)</td>
</tr>
<tr>
<td>Alejandro Portes, John W. Curtis. “Changing Flags: Naturalization and Its Determinants Among Mexican-Americans”</td>
</tr>
<tr>
<td>Public Policy Institute of California, Taking the Oath: An Analysis of Naturalization in California and the United States</td>
</tr>
<tr>
<td>Jennifer VanHook, Susan Brown, Frank Bean, “For Love or Money?: Welfare Reform and Immigrant Naturalization”</td>
</tr>
</tbody>
</table>
§2: Historical Survey

U.S. Immigration and Naturalization Policy: A Brief History

Although the United States has been admitting immigrants and naturalizing new citizens since 1789, the foundations of modern United States naturalization policy were laid in 1952 with the passage of the McCarran-Walter, or Immigration and Nationality Act (INA). For the first time in its history, the US no longer excluded entire ethnic groups or denied citizenship petitions primarily on the basis of race or country of origin, but adopted a set of presumably objective guidelines. Congress formally extended naturalization to all races, stating that “[t]he right of a person to become a naturalized citizen of the United States shall not be denied or abridged because of race or sex or because such person is married” (8 U.S. Code 1422). Members of Communist organizations, former Nazi party members, and those who advocated overthrowing the U.S. government could be barred from naturalizing, however. Congress also tempered the liberalization with a slew of new naturalization regulations and prohibitions. Applicants for naturalization after 1952 had to demonstrate “an ability to read, write, and speak words in ordinary usage in the English language,” as well as demonstrate a basic knowledge of United States government and history (8 U.S. Code 1423). This requirement was not designed to be particularly burdensome—the same code provides that “…ability to read and write shall be met if the applicant can read or write simple words and phrases…and that no extraordinary or unreasonable conditions shall be imposed upon the applicant”—but it nevertheless established a new hurdle, however low, for prospective citizens to overcome. As for immigration policy, the INA maintained the core of the quota-based visa system of the 1920s, setting numerical limits for the number
of visas granted to immigrants from each country per year. These caps were not absolute—the INA encouraged the immigration and naturalization of persons with special skills or education and immediate relatives of U.S. citizens by exempting them from the annual numerical limits. However, the numerical limits for certain countries were so restrictive (e.g., allotting only 100 visas annually to most Asian nations) that President Truman initially vetoed the bill. In his veto message, he wrote that although the bill had its positive aspects, it was based on the discriminatory theory that

“…Americans with English or Irish names were better people and better citizens than Americans with Italian or Greek or Polish names…people of West European origin made better citizens than Rumanians or Yugoslavs or Ukrainians or Hungarians or Baits or Austrians. Such a concept is utterly unworthy of our traditions and our ideals. It violates the great political doctrine of the Declaration of Independence that "all men are created equal." It denies the humanitarian creed inscribed beneath the Statue of Liberty proclaiming to all nations, "Give me your tired, your poor, your huddled masses yearning to breathe free!" (Truman 1952)

The bill nevertheless passed over Truman’s veto, and nationalization policy has remained fundamentally unchanged since (although the quota system for immigration has been liberalized considerably). All applicants for citizenship must demonstrate “good moral character, attachment to the principles of the U.S. Constitution, [and a] favorable disposition to the United States” (USCIS). Immigrants are not explicitly excluded from naturalizing on ethnic, religious, or social grounds, but can be denied for anything ranging from being a self-admitted Communist to a felony conviction. Also barred from naturalizing is any immigrant who has been a public charge in the five years preceding
application for naturalization—presumably all indications that the applicant has less than “good moral character.” General applicants must also fulfill a residency requirement (for most immigrants, five years; for certain refugees and asylees, three years), demonstrate an ability to read, write, and speak English, and have a basic knowledge of U.S. history and government. Certain applicants, like spouses of U.S. citizens and military personnel in the U.S. Armed Forces, can waive some of the general requirements. The process takes about six months, and requires a non-refundable $390 fee, plus the cost of passport photographs (USCIS).

Although it sounds fairly straightforward, anyone who has ever tried to apply for a driver’s license in the United States can imagine the amount of bureaucracy and paperwork involved in applying for US citizenship. The citizenship test has gone fundamentally unrevised since the 1950s, adding yet another layer of inefficiency to the citizenship process. A recent review by researchers from the National Research Council concluded that the test was largely “unsystematic and uneven,” lacking a rational purpose or clear rubrics, and utilizing a civics test that assumed more than a basic command of English (Chudowski 2006). Testing officers have the ability to approve or deny applications with little explanation or formal evaluation, in a well-intentioned but dangerous in practice attempt to give officials the flexibility to consider an immigrant’s background in a citizenship interview. Despite these problems, the US Citizenship and Immigration Service has repeatedly pushed back its target date for revising the test to sometime in 2009.

Instead of revising naturalization policy, Congress has focused on regulating immigration. Obviously, immigration policy has an enormous impact on applications for
citizenship—immigration policy largely determines the pool of immigrants eligible to naturalize. Over the past fifty years, Congress has increased the quotas from most countries and raised the number of immigrants subject to numerical limits (that is, non-immediate relatives of U.S. citizens) to 675,000 annually. About 70 percent of these visas are reserved for other family members, with another 20 percent devoted to employment-based visas. The remaining 10 percent are allocated for “diversity visas,” aimed at nationals of countries that the INS considers “underrepresented” in recent years (104 Statutes At Large 4978). In practice, most are granted to Eastern European or African immigrants.

The last major immigration reform occurred in 1996, when Congress passed the Illegal Immigration Reform and Immigrant Responsibility Act (IIRIRA). Among other things, the IIRIRA requires that all immigrants name a permanent resident or citizen sponsor who must earn at least 125 percent of the poverty level for his or her own family and for the immigrant. The sponsor is legally bound to take financial responsibility for the immigrant for ten years or until naturalization, including but not limited to reimbursing the federal government for any means-tested public benefits received by the immigrant during that period (110 Statutes At Large 3009). The act also limited many federal social benefit plans to United States citizens only and made it much easier to deport aliens convicted of even minor crimes (ibid.). The net effect is that “these provisions add a class dynamic that was not present for family preference immigrants…the likely result is that fewer family preference immigrants from Mexico, the Dominican Republic, and, perhaps, the Philippines will be able to meet the qualifications for award of a visa” (DeSipio and de la Garza, 48). Rather than applying
for visas that they may never be eligible for, immigrants from these countries increasingly turn to illegal immigration.

**Illegal Immigration and the IRCA**

For the first century and a half, there was almost no concern about illegal immigration. The only people with the means and desire to make the arduous journey to America generally fell under the United States’ fairly lax immigration provisions—until 1917, the even the most stringent acts only barred the immigration of prostitutes or criminals, persons who did not freely immigrate, persons infected with contagious diseases, “feebleminded persons,” anarchists, or those “likely to become a public charge” (DeSipio and de la Garza, 26). Then in 1929, shortly after the passage of the restrictive Quota Acts, Congress passed the first of a series of legalization acts granting legal status to any undocumented immigrant who had been living in the U.S. for more than eight years—essentially dismissing the issue.

The issue arose again in the 1950s when immigration from Mexico exploded after World War II. In response to public concern, the INS used a federal short-term agricultural visa program called the Bracero Program to periodically legalize undocumented immigrants from Mexico. In essence, “the INS would make [illegal immigration] literally disappear” (DeSipio and de la Garza, 44). This was not an acceptable long-term solution, however; and in 1986 after long debate Congress passed the Immigration Reform and Control Act (IRCA). It again legalized all undocumented immigrants who had resided in the U.S. for over five years, as well as certain farm workers, and tried to curb illegal immigration by imposing stiff penalties on employers.
who knowingly employed undocumented workers. Over three million persons applied for adjustment to legal status under the IRCA, with nearly 2.7 million approved as legal permanent residents (Rytina 2002). Three-fourths of these were Mexican-born. The IRCA is responsible for most of the swell of naturalization applications in the early to mid 1990s, as immigrants who had gained legal status under the IRCA fulfilled their five-year residency requirement. However, these IRCA legalizees also tended to be less educated and have lower incomes than other immigrants, and naturalized at about 60% of the rate of the immigrant population at large (see Graph 1 below). Despite Congress’ intentions, the IRCA in practice did little to slow the flow of illegal immigrants into the US. There is some evidence that the provisions of the IRCA actually increased the number of illegal immigrants in the US: increased border security deterred undocumented migrants from returning to Mexico when they could not guarantee a safe reentry into the US (Durand 1999, Orrenius 2003).

Since 1986, most Congressional legislation has focused on slowing illegal immigration rather than granting blanket amnesty as in the past. Recent legislation, like the 2006 bill authorizing the construction of a 700 mile long security fence along the U.S.–Mexico border, treats illegal immigration as a border defense issue that can be solved with more guards and tighter border security. In doing so, the United States again fails to address the social and policy-related causes of illegal immigration; only this time, because of widespread public concern about undocumented immigrants, Congress and the INS are unable to sweep the problem under the rug with blanket legalizations. This leaves a huge subset of people—up to 10.5 million as of January 2005, according to the Office of Immigration Statistics (OIS)—that are completely ineligible for citizenship.
Most are poor, working for employers willing to disregard the labor restrictions of the 1986 bill, and concentrated in California and Texas. OIS estimates in the same document that nearly 6 million of these were born in Mexico, with a little under 1.4 million from Honduras, Brazil, Guatemala and El Salvador. Although the illegal immigrant population is by its nature difficult to collect data on, illegal immigrants are a significant and growing subset of the immigrant, and especially the Mexican and Latin American, communities, and thus important to take into account in any comprehensive study of the foreign-born population.
Graph 1. Naturalization Rates through 2001 of IRCA and non-IRCA Immigrants

The USCIS tracked the cohort of immigrants who entered the U.S. from 1979-82. The dark bars represent those who were legalized under the provisions of the IRCA, and the light bars all other immigrants. The data is further divided into three groups: Total (immigrants born in all countries), Mexico (immigrants born in Mexico only), and Other (all immigrants not born in Mexico)

Source: Rytina 2002
§3: Theory

The lack of scholarly agreement on the precise factors which cause an immigrant to naturalize reflects the complexity of the issue. As much as social scientists and researchers might prefer it, immigrants to the US are not discrete data points or easily generalized persons reacting uniformly to government policies. They bring with them certain skill sets and cultural understandings, both heavily shaped by the way of life in their respective countries of origin. They are rooted in communities in the US, whether they choose to associate exclusively with people from their own ethnic groups or to assimilate into middle-class society. And their decisions are shaped not only by government policies, but by the extent to which they understand such policies and the general receptiveness to immigrants that the policy reflects. Although there is no simple way to quantitatively assess every factor that influences an immigrant’s decision to naturalize, I intend to present a model of citizenship which shows naturalization as a function of one’s integration into and familiarity with U.S. culture, norms, and political and bureaucratic processes, influenced not only by one’s individual resources and skills, but also by the relative integration of friends, family, and local leaders and the resources made easily accessible by the government.

Origins and Evolution

When I first began looking at the naturalization data over the last twenty years, my first reaction was to attribute the low rates of naturalization for Mexican-Americans entirely to their relatively low socioeconomic status (SES). It seemed like a fairly straightforward correlation until I began to investigate the reasons behind the
relationship. The naturalization process in the US does not place an objectively higher value on higher income, for example; aside from being able to afford the $390 processing fee, the US government does not impose any official financial criteria for becoming a US citizen. Income is simply one way of quantifying a socioeconomic status which brings with it certain resources or liabilities. A higher income may indicate that an immigrant possesses job skills which are valuable to American employers, such as English fluency, ability to learn new customs and skills quickly, or familiarity with American customs and traditions, and these attributes may make the naturalization process more accessible for various other reasons. The same holds true for all other sociodemographic factors: naturalization is not a direct function of x, y, z sociodemographic variables, but rather a complex equation in which x, y, and z play a role that must be explained.

At the same time as I was developing a theory based on individual factors, I began looking more closely at US naturalization and immigration policy. Part of the reason why I chose to investigate naturalization was because of the recent fervor over illegal immigration and immigrants in general, and I was curious as to whether the supposedly objective criteria of the USCIS citizenship process in practice selected against certain immigrant characteristics – low income or Mexican origin, say. I initially framed my thesis in terms of whether a policy bias existed against Mexican immigrants or not, especially in regards to the 1996 changes in immigration policy and the restriction of federal benefits to US citizens only. I viewed immigration and naturalization policy as a filter, screening out “undesirable” or “un-American” immigrants in much the same way that immigrants with Eastern European or Asian heritage were discriminated against in the first half of the 20th century.
These initial theories lent themselves well to a cost-benefit analysis, with immigrant skills on the positive side and restrictive government policy on the negative side. I began to consider all factors in terms of resources, realizing that an immigrant’s specific skills and attributes affected his propensity to naturalize primarily because they afforded access to certain resources—information about the naturalization process, English classes, the ability to navigate American bureaucracy, or an understanding of the costs of and the value in gaining US citizenship, for example. Likewise, government policy affects naturalization rates because it increases or decreases the amount and number of resources immigrant groups have access to. When I framed the issue this way, I realized that many resources I identified as important were not based solely on individual characteristics, but rather assumed the existence of a community with certain characteristics. Especially after reading Borjas’ theory of ethnic enclaves, I realized that community characteristics had an enormous impact on the resources available to immigrants—a concept that I should have been aware of from the beginning, seeing as I chose to organize my analysis by grouping immigrants according to country of origin. By the end of my research, my theory had evolved from a one-dimensional analysis of individual characteristics as restricted by government policy to a more robust and sophisticated theory of naturalization as access to resources based on individual, community, and government factors, with much more potential to effectively address a question as complicated as why immigrants choose to naturalize.
When and for what reasons do immigrants choose to naturalize?

The simplest answer is that immigrants naturalize when it is relatively convenient and easy to do so, and carries no major costs. In this, I agree with authors such as Louis DeSipio and Rodolfo de la Garza who view naturalization as primarily a cost-benefit problem: an immigrant naturalizes when the benefits outweigh the costs from the individual’s point of view. An individual cost-benefit analysis, however, does not adequately account for all the relevant factors. How easily an immigrant is able to naturalize depends to a large extent on how familiar the immigrant is with the bureaucratic processes and how much assistance and support he or she receives—which are in turn, I believe, functions of an immigrant’s social and cultural integration into American society and the types of institutional barriers in place.

I have so far been relatively vague about the exact nature and relationship to naturalization of the “resources” I expect each source to contribute. To complete the picture, let me discuss the particular relevance of each aspect of my theory—individual characteristics, community characteristics, and government policy—in turn.

- Individual Characteristics and Skills

   It seems most obvious that individual characteristics should affect an immigrant’s propensity to naturalize: naturalization is after all the process by which an individual becomes a US citizen. What is less obvious is exactly what characteristics and how. One key attribute that native-born citizens often take for granted is a certain familiarity with American bureaucracy and customs. The average American citizen might be inconvenienced by long lines or unfriendly government employees at the license bureau,
or frustrated while trying to figure out exactly what the IRS means by that footnote to line 27(a)(i) on a tax form, but has a basic understanding of the way the system works—where to look for information, who to call for help, and even how to complain and be persistent until one succeeds.

Imagine instead being a new immigrant (assume a legal permanent resident, in fact) with a limited command of English. The formal steps to the naturalization process would be to spend five years as a law-abiding, taxpaying resident of the United States, then go to the closest USCIS office to declare your intention to seek citizenship. After another waiting period of approximately six months while USCIS processes your application, you schedule an interview with a naturalization official to take your civics and English tests, and if you pass a few months later take your oath and receive your new US passport.

This formal process, however, assumes that every immigrant not only has a working knowledge of the basic requirements of the naturalization process but has access to the information and paperwork necessary to complete it. The more familiar an immigrant is with American culture, the more likely it is that he or realizes that the first step is to call the local USCIS office or search online for basic information and paperwork to begin the naturalization process. This is where socioeconomic differences factor in: education, English fluency, and income level are all reasonably correlated to a basic understanding of US society, or at least the ability to gain such an understanding. It is logical that a well-educated, white-collar immigrant fluent in English would find it much easier to walk into his local USCIS office and ask questions about the citizenship process, while the hypothetical immigrant discussed above might be too intimidated by
the complex web of paperwork and red tape. The better-educated immigrant would have
greater access to any classes he needed to prepare for the citizenship test, if he needed
any preparation at all, while the non-English fluent immigrant may have to wait months
to enroll an overcrowded government-subsidized English as a Second Language course.

Another related but important issue concerns the value an individual places on
citizenship relative to the perceived costs of attainment. An immigrant living at
subsistence level or just above may find it more difficult to set aside the time and money
necessary to complete the citizenship process. He may be unsatisfied with his experiences
in the US, and even consider returning to his home country in the future. An immigrant
living at a higher socioeconomic level, however, may be more satisfied with life in the
US and consider the amount of time and money involved in the citizenship process to be
nominal. In this way, higher socioeconomic status both lowers the costs and increases the
perceived value of citizenship on an individual level. Individual social and cultural
integration into US society also increase an immigrant’s level of satisfaction with life in
the US and consequently the value placed on citizenship.

- Community characteristics and resources

The relative costs and benefits of citizenship, however, are not determined only
by individual socioeconomic status. The value and relative ease of attainment of
citizenship, as well as degree of integration into US society, are in large part determined
by the community an immigrant belongs to, whether in the US or abroad. The culture of
one’s native country, along with the status of one’s ethnic group in America, becomes
enormously important because new immigrants tend to cluster with other immigrants
from the same country and “follow their lead” to a certain extent. Naturalization is a much more socially motivated process than many past authors have acknowledged. If a new immigrant becomes a member of a community that is in effect an “ethnic enclave,” largely isolated from US society, in which most of the other members are poor, speak little English, and most are not naturalized, the chances of that immigrant naturalizing is very low. The prospects may be even worse for members of small ethnic communities that are relatively isolated even from other communities of the same ethnicity—although Borjas’ main concern was for the large ethnic enclave of Mexican-Americans in the Southwest, a community of that size has at the very least some sort of “strength in numbers”. Large communities, however isolated, simply because of their size are more likely to be able to provide some resources to their members, whereas smaller communities are more limited.

Whatever its size, not only does such a community itself place little value on and offer no support for naturalization, but its members are much less likely to learn English, gain job skills valuable to American employers, or take any other action that might improve their socioeconomic status. Community characteristics are a strong normalizing force on the individual characteristics of its members, pushing all towards a median level of socioeconomic status and integration into US society.

This normalizing force works both ways, however. The high naturalization rate among Cuban-Americans in Miami is an excellent example of how a community with excellent political resources and a high level of integration into US society that places a relatively high value on US citizenship can affect its members’ ability and decision to naturalize. A community originally founded by exiled Cuban elites is now able to offer
its resources to poor refugees as well, allowing new immigrants to work their way up the socioeconomic ladder quickly and offering easy access to naturalization. The Cuban-American community’s positive attitude towards the US is also reflected in its high naturalization rate.

The influence of community on the decision to naturalize extends beyond the borders of the US as well. An immigrant who remains closely tied to his or her home country is less likely to consider settlement in the US permanent and less likely to view US citizenship as something to be valued. This is a significant factor in the Mexican case especially—if one remains connected to a community on the Mexican side of the border and returns frequently, there is little value in having a US passport rather than a Mexican one. An immigrant who is far away from her home country is more likely to have made a permanent move, leaving behind old customs and traditions in order to build a completely new life in the US.

- Government influences

Government policy also has a significant impact on the percentage of immigrants who naturalize, for similar reasons. More restrictive immigration and naturalization policies, especially those specifically aimed at a certain group of immigrants, or immigrants with certain characteristics (a low socioeconomic status, for example), could signal a general public sentiment against these groups of immigrants or against immigrants in general, as well as directly limiting the accessibility of the naturalization process. Therefore these immigrants are less likely to be integrated into the population at
large, and also less likely to receive any support or aid in naturalizing from American citizens outside of their ethnic community.

The government also offers little support for immigrants who want to naturalize. Information on the naturalization process is poorly distributed, and federally funded preparation classes for the citizenship test are overcrowded with waiting lists months long (NYT article). There is little incentive for the government to make the process easier for unempowered immigrants—since they are not citizens, they cannot vote in federal elections, and the groups of immigrants with the greatest need tend to be those with the least political and economic power.

§4: Hypothesis

From this theory, I expect that any factor which reflects an immigrant population’s relative integration into US society and their intentions to stay permanently should be positively correlated with naturalization rate. Integration and intention to stay are themselves closely correlated as well: the longer an immigrant intends to remain in the US, the more incentive he or she has to learn more about US culture and ways of life instead of just passing the days until returning home, and likewise the more an immigrant becomes integrated into US society the more the US becomes home. Length of residence in the US has proved such a strong predictor of naturalization rate because long-term residence increases the chances that an immigrant intends to settle in the US permanently. Also, the longer an immigrant lives in the US, the more opportunities he or she has to form relationships, make connections, understand the workings of US society, and accumulate other resources that could help the naturalization process.
Homeownership, and to a lesser extent, marriage are also useful estimations of permanence. Jones-Correa (2001) specifically suggests homeownership as a useful proxy for an immigrant’s intention to remain in the US and degree of assimilation, and there is evidence that married immigrants tend to remain in the US and become citizens, especially if one’s spouse is US-born (Portes 1987, Massey 2006). Both suggest a long-term commitment to life in the US, whether financial (a mortgage) or familial (children born or raised in the US), and thus should be positively correlated with naturalization rate. Reason for immigration also affects an immigrant’s intention to stay in the US: those coming in search of short-term employment often intend to work in the US for a few years, and then return to their home country. Because the population ages 25-44 represents most of the mobile labor force willing to move from country to country in search of jobs, this group is less likely to intend to settle permanently in the US, and less likely to see US citizenship as something worth acquiring. After having lived in the US for many years, however, some may decide to remain in the US and eventually become citizens.

SES variables such as income and education should also be positively correlated with naturalization, as they suggest a greater ability to function in and a greater level of satisfaction with US society. However, receipt of public assistance is a bit trickier – before 1996, it should be indicative of a low SES and therefore negatively correlated with naturalization (Bouvier 1996, PPIC), but after 1996 the correlation may become positive since only US citizens were legally allowed to receive such benefits3.

---
3 Congress did include a few exceptions for aliens with “strong military connections,” lawful permanent residents with long work histories in the United States, or certain other extremely specific cases (Moore 1999)
Finally, there is some evidence that gender may play a role in determining whether or not an immigrant becomes a citizen. This may be partly because heavily male populations suggest a large contingent of migrant workers in the US for jobs and little else. At least one author also suggests that women tend to be more heavily involved in their communities, and therefore more likely to come into contact with the kind of political and social resources they need to naturalize (Pantoja 2006).

Political and social strength of the community should be an asset in the naturalization process, except in the extreme situations when communities of poor, relatively disadvantaged immigrants become ethnic ghettos which hinder their members’ integration into the US. This means that population density could affect naturalization rate in one of two ways—either a more concentrated immigrant population is a sign of a stronger community with access to more resources, so high population density is positively correlated with naturalization rate, or high population density indicates ethnic ghettos in the more populated regions, in which case low population density would be positively correlated. I expect some variation across world region in this respect. Large communities of Latin American immigrants are more likely to end up as ethnic ghettos than large communities of immigrants from Europe, because the European community is more likely to be integrated into and accepted by the larger community in a particular town or city. Ideally, I would be able to quantify not only the population—the “numeric strength”—of ethnic communities, but also their relative political strength.

Likewise, which ethnic community an immigrant is tied to should have an impact, not only in terms of country of birth but in whether the community is in the US or abroad.
Those with strong connections to their home country are less likely to be part of a strong ethnic community in the US, and thus do not have access to the same kind of community resources as immigrants with strong ties to communities in the US. Factors such as distance from home country should be positively correlated with naturalization rate since a greater distance represents a more complete break with the traditions and cultures of one’s country of birth, as well as a greater chance of permanent settlement in the US.

Government policy should also shift the emphasis on which factors are the strongest predictors of naturalization, although I am not sure in this case what the impact of the 1996 bill will be. SES characteristics may become more important, because all immigrants have a greater incentive to naturalize but only those with the resources to do so (i.e., a high enough SES) will naturalize, or SES characteristics may become less important because the greater ability of immigrants with more resources to naturalize will be statistically counteracted by the greater incentive of poor immigrants to naturalize to avoid losing federal benefits or the threat of deportation. The stipulation that all new immigrants are required to name a financially liable permanent resident or citizen sponsor may also increase the percentage of immigrants who come to the US to settle permanently—if the immigration process becomes more difficult, then those who choose to go through it should be more committed to a long-term stay.
III. Data

§1: Explanation of Data Sources and Methods

For my analysis, I used data from the 1994-2005 Current Population Survey (CPS) March supplements, accessible through the U.S. Census Bureau’s DataFerrett system. The CPS is a monthly survey of 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics that has been conducted for over fifty years. The sample is selected to reflect the civilian non-institutionalized population as accurately as possible. All household members fifteen years old and older are interviewed, and data is collected on employment, earnings, hours of work, occupation, and other labor-related statistics, as well as demographic characteristics like age, sex, race, marital status, citizenship status, and educational attainment. The March supplement contains further information on topics such as health insurance, work experience, and school enrollment.

Individual-level data was not available through the CPS, so I chose to use country in year as my unit of analysis. That is, I analyzed the characteristics of the aggregate population born in a country (e.g., France) in a specific year (say 1998). Excluding countries with incomplete or flawed data, this provided between 90 and 100 observations per year, with the exception of 1994 which only provided 42 countries with complete data because of omitted values in the CPS dataset.

I limited my dataset to the universe of foreign-born persons residing in the U.S. who were not U.S. citizens at birth (that is, excluding those persons born on foreign soil to at least one citizen parent) and are sixteen years or older, although for simplicity I refer to this universe as the “foreign-born population”. From the data available through the
CPS March Supplement, I isolated ten variables with a possible effect on propensity to naturalize:

LEASTPOP: Percentage of foreign-born population living in the two census regions\(^4\) in which the fewest other immigrants from their country of birth were living in that year. Values have a continuous range of .0140501 to .4857235.

EDUC: Percentage of foreign-born population which reports having completed at least “some college.” Values may range continuously from .1065616 to .8987774.

MARRIED: Percentage of foreign-born population that is married. Spouse may be either present in the US or living abroad. Values range continuously from .1148649 to .9649677.

LOWINC: Percentage of foreign-born population classified as living below the federal poverty line. Values range continuously from .0024204 to .7075045.

HOMEOWN: Percentage of foreign-born population owning or buying a home. Values range continuously from .0084748 to .946464.

FEMALE: Percentage of foreign-born population that is female. Values may range continuously from .123235 to .8257405.

WCOCCS: Percentage of foreign-born population reporting a “white-collar” occupation, which I defined as any occupation under the broad census

---

\(^4\) There are four basic regions, defined as follows: Northeast (PA, NY, NJ, CT, RI, ME, MA, VT, NH), South (VA, KY, WV, DC, DE, MD, NC, SC, GA, FL, TN, MS, AL, TX, OK, AR, LA), Midwest (OH, IL, IN, MI, MN, WI, IA, MO, ND, SD, NE, KS), and West (AK, HI, WA, OR, ID, CA, NV, UT, CO, AZ, NM, MT, WY)
categories of “Executive, Administrative, and Managerial,” “Professional Specialty,” “Technicians and Related Support,” “Sales,” or “Administrative Support” occupations. Values range continuously from .0819841 to .9859933.

RESLENGTH: Percentage of foreign-born population living in the US for 10 or more years. Values may range continuously from .0655434 to .9912432

AGE: Percentage of foreign-born population between the ages of 25 and 44. Values range continuously from .101575 to .7908182.

WELFARE: Percentage of foreign-born population receiving food stamps, public assistance, Supplemental Security Income, or unemployment compensation. Values range continuously from .0026256 to .970268.

In addition, I chose to include a variable (DIST) measuring the approximate distance between each country and the US. In most cases, I measured the distance in miles between the country’s capital city and the closest major US airport that flights from that country would most likely fly to—New York for most European cities, Miami for the Caribbean, San Francisco or Los Angeles for Asian countries, for example—in order to create a variable that measured the approximate distance one would have to travel in order to enter the US from a given country. Values range continuously from 0 (for Canada and Mexico) to 7900.

Several of these variables had been identified as possible predictors of naturalization in other studies (namely age, reslength, welfare, educ, married, lowinc, homeowner, female, wcoccs), but I chose to define certain ones in somewhat unique ways. This was partly necessitated by the nature of my data, and partly by the nature of my
theory. While most other studies have focused on the “individual resources and skills” portion, I intended to evaluate the effects of community characteristics, and to a lesser extent the impact of government policy. As far as evaluating community characteristics, I chose to limit age to the population between the ages of 25 and 44 because it tends to be the labor force population that is most mobile, or most likely to migrate in search of work (Alba 2006). This could have one of two effects. Either the 25-44 immigrant population is essentially a migrant population, coming to the US in search of jobs and intending to return to their home country, as is the case with many migrant farm workers (Chavez 2006), or they immigrate with the intention of working and living in the US permanently.

Income (lowinc) is another factor widely positively correlated with naturalization (Antecol 2003, DeSipio 1987, Bouvier 1996, PPIC). I use the percent of immigrants in poverty rather than, for example, median income because it is a closer measure of the proportion of the foreign-born population that has access to the fewest socioeconomic resources. According to at least one study, income can also be seen as a measure of the job skills, including English fluency, workers possess, especially among Hispanic immigrants (Yang 2005).

As for educ, I chose “some college” as my cutoff point because there is evidence that this is the category into which a significant percentage of Mexican-Americans fall, and the focus of my research is ultimately on Mexican immigration (Alba 2006). More generally, using the percentage of immigrants with a high school degree was a bit too broad and simply not as useful as support for my thesis. A high school degree is a good indicator of a basic educational status; having completed at least some college indicates

---

5 According to Alba’s data from the 2000 Census PUMS, about 30 percent of Mexican-Americans fall into this category. Alba criticizes another author, Samuel Huntington, for using “college degree” as his cut-off point and thus “grossly understating” Mexican-American educational attainment.
not only a higher degree of education, but also suggests the possibility of upward socioeconomic mobility. Increasingly, American employers expect their employees to have at least a high-school education in all but the lowest-paying jobs, and it is becoming harder to find stable, well-paying jobs that require only a high-school degree (Alba 2006). Therefore an immigrant with more education than a high school degree, even if he or she never earned a degree, may have an edge in the job market over those with only high-school degrees. Some college may also reflect a greater social mobility, an ability and a desire to go beyond the basic educational structure, even if only for a year or two at a local community institution. \textit{Wcoecs} is a more direct measure of job skills valuable to American employers (such as English fluency and general familiarity with American culture) among the immigrant population, and has also been shown to be positively correlated with naturalization (DiSipio, Bouvier 1996, Rytina 2002).

The effect of immigrant population concentration (\textit{leastpop}) is something that has been less widely studied. There have been several articles written on the negative correlation between California residency and naturalization (PPIC, Grebler 1966, Jones-Correa 2001), with some including New York, Texas, Florida, or Illinois as well. More recent articles suggest that a strong immigrant community may be an asset in the naturalization process (Bloemraad, Van Hook, Basok). Because of the low quality of the data available in certain CPS years on area of residence, I chose the relatively crude method of calculating the percentage of the immigrant population living in the two census regions (see footnote 2, above) in which the fewest immigrants from that country were living. For example, for Mexico I added the number of Mexican immigrants living in the Midwest and Northeast regions and divided it by the total Mexican-born population
living in the US, because the vast majority of Mexican immigrants settle in the South or West regions. In the Mexican case the Midwest and Northeast were consistently the two regions in which the fewest number of immigrants lived, but for other countries (especially those with a more evenly distributed population) there was some variation from year to year, so I reassessed which regions I included for each year. It is admittedly a clumsy variable, but it represents an attempt at quantifying the relative concentration of the immigrant population, which has not been done in any comprehensive way thus far.

Distance from native country (dist) is another variable which has been identified as possibly correlated with naturalization in theory—the farther one is from one’s home country, the more one must cut off ties there and establish “roots” in the US (Borjas 2002, DeSipio 2004, Portes 1987)—but not to my knowledge quantitatively evaluated.

More fundamentally, I have chosen to define my sample set much more broadly. Previous studies have tried to limit their sample sets to only the foreign-born population eligible for naturalization—that is, legally admitted immigrants who are over eighteen years of age and have resided in the US for five years or more—and some have restricted their samples even further, to Latino registered voters or to non-refugee, non-Mexican immigrants. While there is merit in each type of analysis, for my community-based theory I felt it was most logical to analyze the characteristics of the community as a whole. Whether registered voter or illegal alien, whether resident in the US for ten years or two weeks, each immigrant is a member of an ethnic community and as such has an effect on the actions of other community members. Many Mexican households, especially in the Southwest, are made up of both legal and illegal immigrants; according to some estimates, 46 percent of settled Mexican households with legal immigrants also
contain someone without documents (Massey, Malone, and Durand 2002). Members of households containing an undocumented migrant are probably less likely to trust government officials or seek out government resources for fear of having a household member deported. They also are less free to circulate in search of job opportunities, both inside the U.S. or in other countries. Or perhaps a group of new immigrants from Poland join a Polish community, and their negative experiences with Homeland Security upon entering the US significantly weaken the community’s view of the US government. There is some support in the existing literature for using this type of broad sample set: Massey and Bartley (2005) explicitly warn about the dangers of studying legal immigration only, especially as the size of the undocumented immigrant and nonimmigrant populations grow.

In order to assess the impact of changes in government policy, I had to be a bit more creative—there seemed to be no direct way to measure the effects of the 1996 legislation. I decided to use time as a proxy, dividing my data into pre-1996 and post-1996 cohorts and testing each group separately after testing the set as a whole in order to see if there were any changes in which variables were significant.

Neither the CPS nor the decennial censuses exclude illegal immigrants, and although many undocumented immigrants are understandably reluctant to give out any information to the US government, it is fairly well-documented that a large number are included in each survey (Massey and Bartley 2005). Therefore the CPS data I draw on most likely includes a large but underrepresented illegal immigrant population. I made no effort to estimate the effect of the unrepresented illegal immigrant population, simply because of the amount of effort and time involved in (and perhaps impossibility of)
imputing such data at the level of detail I needed\textsuperscript{6}. My dependent variable then becomes percent of total foreign-born population born in a specific country who have become naturalized citizens (\textit{natz}), and reflects low rates of naturalization compared to studies that attempt to exclude illegal immigrants from their samples. Values range from .0175146 to .8698364.

\textit{Method}\textsuperscript{7}

After collecting all the relevant data, I compiled it into one large data file and uploaded it into the Stata 9 data editor. This dataset included slightly over 1,000 observations; that is, 1,000 sets of variables calculated for the subset of the foreign-born population in a specific year that was born in a single foreign country. After running one OLS regression on the entire set as a test, I checked the data set for problem observations—observations with incomplete data, or nonsense values such as a value of 1 for \textit{married}, indicating that 100\% of the population was married. Most of these observations were for countries with several different possible names—Czechoslovakia, the Slovak Republic, and the Czech Republic were all separate entries, as were Russia and the USSR—or for small island countries like Dominica and the Azores. I also dropped all observations from the Oceania world region (Australia, Fiji, and New Zealand) because of data inconsistencies. The end result was a set of 838 observations.

\textsuperscript{6} For a more complete discussion of the factors involves in imputing the illegal immigrant population, see “Estimates of the Unauthorized Immigrant Population Residing in the United States: 1990 to 2000,” issued by the Office of Policy and Planning of the US Immigration and Naturalization Service, available through the USCIS website at www.uscis.org

\textsuperscript{7} My regression analysis method was largely shaped by the web book \textit{Regression With Stata} published by the UCLA Academic technology services (accessible online at www.ats.ucla.edu/stat/stata/webbooks), especially Chapters 1 and 2 on multiple regression and regression diagnostics.
Next, I checked to make sure that my variables were in fact normally distributed—a key assumption of OLS. The welfare, wcoccs, and lowinc variables all indicated normality problems, but by taking their square roots I was able to replace them with the normalized variables swelfare, swcoccs, and slowinc. I ran a second regression analysis with this improved data set, along with a correlation analysis. This indicated a high correlation between swcoccs and educ. After testing the regression with and without educ and swcoccs, I found that I achieved better results (higher R squared, swcoccs became significant) when I dropped educ. I ran the regression analysis one more time to come up with my final global analysis.

I ran the same OLS analysis one more time, but limiting it to only the data from 1994-1995, and similarly for 1996-2005, since the Illegal Immigration Reform and Immigrant Responsibility Act was passed in 1996 and I wanted to judge its effects, if any. I then divided my data by world region\(^8\) and ran a separate regression and correlation analysis for each region to check for any regional variations. I ran one last regression analysis for Mexico alone, excluding the variable dist because it was obviously uniformly zero. Finally, I went back through each data set and ran one more regression including only the variables I found to be significant or close to significant, just to see if the model was affected by eliminating non-significant variables.

After each regression analysis, I also ran several tests on the results to check that neither my data nor my results were fundamentally flawed. Although normal distribution of the residuals\(^9\) is not required in order to obtain unbiased estimates of the regression

---

\(^8\) Africa, Asia, Europe, Central America, Latin America, the Caribbean, or North America

\(^9\) Residuals are a measure of the unexplained variation after fitting a regression model: the difference between the observed value of the variable and the value suggested by the regression model. See the glossary (Appendix B) or Appendix C for further explanation of regression analysis and OLS procedures.
coefficients, it does ensure that the p-values for the t-tests and F-test will be valid. OLS regression merely requires that the residuals be identically and independently distributed. Nevertheless, I checked my data and each subset for normal residual distribution to err on the side of being overly cautious rather than not cautious enough\textsuperscript{10}. I also checked my data for heteroskedasticity\textsuperscript{11} and multicollinearity\textsuperscript{12} as a precaution. I found no major problems except for with the small Mexican data set, which had its own set of peculiarities as I explain later. The final Stata output is attached as Appendix D, including correlation analyses and specific tests run with results.

\textbf{§2: Limitations of Data}

While I made every attempt to be as accurate as possible, flaws exist in every analysis, and the results must be interpreted with a certain degree of caution. First, my results can only be as good as the data I used, which was not always flawless. I tried as best I could to eliminate country-year observations with missing values or those with blatantly inaccurate data (for example, countries for which the Census Bureau reported a 0\% female population, or that over 100\% of immigrants had white-collar occupations); however, it was impossible to vet the accuracy of each value, and I’m certain that some were questionable at best. The Caribbean and Eastern Europe were especially problematic in this regard.

\textsuperscript{10} For clarity I gloss over many of the more technical particulars of my method. An interested reader can consult Appendix C for a more detailed explanation.

\textsuperscript{11} One of the main assumptions for the ordinary least squares regression is the homogeneity of variance of the residuals; that is, if the model is well-fitted, there should be no pattern to the residuals plotted against the fitted values. If the variance of the residuals is non-constant then the residual variance is said to be "heteroskedasticity." (Stata web book, ch. 2)

\textsuperscript{12} When there is a perfect linear relationship among the predictors, the estimates for a regression model cannot be uniquely computed. The term collinearity implies that two variables are near perfect linear combinations of one another, and multicollinearity implies that more than two variables are involved. Ibid.
Second, although I made my best effort, some of my variables are not the most accurate measures of the factors I wanted to investigate. For example, \texttt{leastpop} is designed to quantify the concentration of the immigrant population from a certain country, but it does so very broadly. It probably would be a more accurate measure if I had looked at the percentage of immigrants living in the top five or ten metropolitan statistical areas\textsuperscript{13}, say, but that data was only available for a restricted set of countries and years. I made the decision to keep as many observations as possible at the expense of broadening my variable definition because I felt that it would provide for a better statistical analysis in the end.

Also, although I tried to account for it as much as possible, some of my variables are unavoidably intercorrelated. For example, \texttt{homeown} is affected by \texttt{reslength} because the longer an immigrant is in the US, the more likely he or she is to buy a home, all other factors constant. The same is true for \texttt{slowinc}, \texttt{swcoccs}, and \texttt{swelfare}—an immigrant who has lived in the US longer is more likely to have increased his SES, all other factors constant. I tried to minimize this by limiting \texttt{reslength} to those living in the US for ten years or more, hoping that after ten years most immigrants had made most of the upward socioeconomic progress they would make and bought a home if they ever planned on buying one, but there are always exceptions. Likewise, many of the SES variables are correlated with each other. With these, however, I found that dropping some of the SES variables had little effect on any model aside from decreasing the R-squared value. I also tested for excessive multicollinearity to check that the variables were not too strongly dependent on each other, and found no major problems.

\textsuperscript{13} Census Bureau term referring to a major city and its surrounding metropolitan area.
Most importantly, my data is limited to analysis of community characteristics by country of birth only. That is, I analyze the effect of the characteristics of the aggregate group of immigrants from a country on the naturalization rate of immigrants from that same country. In my opinion, this is an acceptable and useful community to analyze because this is the community that most immigrants rely on, especially immediately after entering the US, and therefore fundamentally shapes their experience in the US. The 1996 change in the sponsorship requirement, which mandated that each immigrant name a citizen sponsor who was legally and financially responsible for the immigrant for up to ten years, only reinforced this connection to the ethnic community: most immigrants likely name a family member or other person from their same country. Still, care must be taken to remember that my results imply nothing about individual characteristics in relation to individual propensity to naturalize, but rather look at community naturalization rate as a function of community characteristics. Anyone interested in the connection between individual characteristics and naturalization can refer to the Survey of Literature section of this thesis, or any of the many other articles and studies that have been published on that topic.

§3: Data Analysis

Before I present my results, it may be helpful for some readers to include a few words on interpreting multiple regression tables and output. The key values in any multiple regression table are the R-squared value, the coefficients and the p-values. The R-squared gives an estimate of the explanatory power of the model; that is, how much of the variation in the dependent variable is explained by the independent values. An R-
squared of 0.60 suggests that about 60 percent of the variation is explained by the model, which in the uncertain world of statistical analysis is considered fairly respectable.

The coefficient measures the change in the dependent variable for every one-unit increase in each independent variable. For example, in my first global regression I found that the coefficient for female was approximately -0.22. This means that for every 10% increase in the percentage of the population that is female, the naturalization rate will drop about 2% \((0.10 \times -0.22 = -0.022)\). A related term is the “beta” value, which gives the standardized coefficient of each variable; that is, it scales each coefficient on the same 0-to-1 scale, allowing direct comparisons between the magnitude of each variable’s effect on natz. If the absolute value of a variable’s beta coefficient is larger than that of another variable, the first variable has an absolutely greater impact on natz.

The p-value indicates whether or not the variable is considered “significant”; that is, it measures the probability that the variable does not significantly affect the dependent variable. Variables with p-values of less than 5%, or 0.05, are generally considered significant in OLS regression. The lower the p-value, the more likely it is that the variable is significant.

Finally, the last row in each regression is labeled “_cons,” and gives the value of the constant term in the regression equation; that is, the part of the naturalization rate which does not depend on any variable. More information about regression analysis can be found in Appendix C.

*Global analysis, 1994-2005*
My first regression analysis was the broadest, using data from all world countries in all years from 1994-2005. Not surprisingly, this regression also had one of the lowest R-squared values (0.6593). I found that distance, homeownership, length of residence, marital status, sex, age, and white-collar occupation were all significant predictors of naturalization rate. Distance, homeownership, more than 10 years US residence, marriage, and white-collar occupation were all positively correlated with naturalization rate, while higher percentages of female immigrants or immigrants between the ages of 25 and 44 were correlated with lower naturalization rates.

The percentage of the population living in the US for over 10 years (reslength) was most strongly correlated with naturalization rate, followed closely by homeown and age. However, since homeown and reslength are reasonably correlated (the correlation coefficient is 0.5010, indicating that a one-unit increase in one variable would cause about a half-unit increase in the other), age has a slightly higher beta coefficient in the final model that homeown. The table displayed in the text (Table 1) shows the result of a regression analysis including the significant variables only; the full table can be found in Appendix D.
Table 1: Regression analysis of global data, 1994-2005, simplified table

Number of observations: 838; R-squared = 0.6593

<table>
<thead>
<tr>
<th>natz</th>
<th>Coef.</th>
<th>Std. Err.</th>
<th>t</th>
<th>P&gt;t</th>
<th>[95% Conf. Interval]</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>dist</td>
<td>0.0000132</td>
<td>1.78E-06</td>
<td>7.4</td>
<td>0</td>
<td>9.70E-06</td>
<td>0.1801773</td>
</tr>
<tr>
<td>leastpop</td>
<td>-0.0718251</td>
<td>0.0368617</td>
<td>-1.95</td>
<td>0.052</td>
<td>-0.1441785</td>
<td>-0.0452732</td>
</tr>
<tr>
<td>swcoccs</td>
<td>0.1205956</td>
<td>0.0302627</td>
<td>3.98</td>
<td>0</td>
<td>0.061195</td>
<td>0.1799961</td>
</tr>
<tr>
<td>married</td>
<td>0.1096981</td>
<td>0.038821</td>
<td>2.83</td>
<td>0.005</td>
<td>0.0334992</td>
<td>0.0722885</td>
</tr>
<tr>
<td>homeown</td>
<td>0.2557355</td>
<td>0.0283083</td>
<td>9.03</td>
<td>0</td>
<td>0.200171</td>
<td>0.2633754</td>
</tr>
<tr>
<td>reslength</td>
<td>0.4221951</td>
<td>0.0251168</td>
<td>16.81</td>
<td>0</td>
<td>0.372895</td>
<td>0.4314252</td>
</tr>
<tr>
<td>female</td>
<td>-0.1504472</td>
<td>0.0430603</td>
<td>-3.49</td>
<td>0.001</td>
<td>-0.2349671</td>
<td>-0.0774555</td>
</tr>
<tr>
<td>age</td>
<td>-0.3276542</td>
<td>0.0323955</td>
<td>-10.11</td>
<td>0</td>
<td>-0.3912411</td>
<td>-0.2372199</td>
</tr>
<tr>
<td>_cons</td>
<td>0.0724767</td>
<td>0.0420281</td>
<td>1.72</td>
<td>0.085</td>
<td>-0.0100173</td>
<td>0.1549707</td>
</tr>
</tbody>
</table>

These results suggest a reasonably strong correlation between factors I identified as indicators of permanence and integration into the US and community naturalization rate. The relatively large positive coefficients of *homeown*, *reslength*, and *dist*\(^{14}\) indicate that, all other things equal, a well-established immigrant community from a faraway country, most of whose members have been living in the US for over ten years and own their own homes, will have a higher naturalization rate than an immigrant community from a country closer to the US whose members largely rent and have only recently arrived in the US. The relatively large negative coefficient of *age* supports this theory as well: if a large percentage of an immigrant community is a young, mobile labor force, then the community will have lower naturalization rates because this group is less likely to naturalize. Measures of permanence and integration are positively correlated; measures of transience and mobility are negatively correlated.

White-collar occupation seems to have an impact as well, albeit not as large. This may be either moderated by or attributable to the high significance of homeownership—it

---

\(^{14}\) Although the actual coefficient of *dist* is very small in comparison to all other coefficients, recall that while all other variables are in terms of percentages (and hence some value between one and zero), *dist* ranges into the thousands of miles and thus must have a much smaller coefficient. A better way of determining the relative impact of each variable on *natz* is by looking at the beta (standardized) coefficient.
makes sense that white-collar workers are more likely to buy homes than other workers, and this is reflected in a correlation of about 0.44 between \textit{wcoccs} and \textit{homeown}.

\textbf{Marriage} likewise has a slight positive impact on naturalization, perhaps because some of the positive effect of marriage on permanence is canceled out by those who have spouses living abroad (and thus more incentive to return home).

However, in this model both \textbf{female} and \textbf{leastpop} are somewhat troubling. Rather than being positively correlated with naturalization rate, as the existing literature suggests, \textbf{female} is actually negatively correlated, although the effect is slight (a coefficient of -0.15 suggests that for every 10\% increase in the percentage female for a population, naturalization rate will decline by 0.015, or 1.5\%). This may be attributable to variation in the social status of women across cultures; maybe in some more patriarchal cultures, women become less likely to naturalize than men because their access to resources is restricted. Hopefully, the regional regressions will clarify this situation.

\textbf{Leastpop} is disappointing not because of the sign of its coefficient, since I had developed hypotheses for either case, but because of its magnitude. Although it indicates that a more highly concentrated ethnic population is an asset in the naturalization process, its effect on community naturalization rate is so slight—about one-tenth the impact of \textit{reslength}—that I’m not quite comfortable yet concluding that it is a good predictor of naturalization rate.

\textit{Global Analysis, 1994-1995}

I then limited my regression to only the data collected in 1994 and 1995; that is, in the years before the passage of the 1996 Illegal Immigrant Responsibility and
Immigration Reform Act. I looked at both including and excluding 1996 in this set, but I found my model was slightly more explanatory (that is, had a higher R-squared value) when I excluded 1996. Even though most of the Act’s provisions did not take effect until 1997, I expect that anticipation of policy changes and an increase in anti-immigrant public opinion could have an effect similar to the policy change itself during 1996.

In this model, \textit{reslength}, \textit{dist}, \textit{married} and \textit{age} remained significant, with \textit{reslength} an even stronger positive predictor of naturalization rate, \textit{dist} and \textit{married} positive predictors, and \textit{age} remaining a negative predictor, all as in the previous model. \textit{Swcoecs} was no longer significant, but in its place \textit{slowinc} became significant and negatively correlated. Because of the high negative correlation between \textit{swcoecs} and \textit{slowinc}, this may be interpreted simply as another form of evidence that low SES is correlated with low naturalization rates. Note also that the R-squared value is significantly higher, indicating a much closer fit between the data and the model in this restricted case.

\textbf{Table 2: Regression analysis of global data, 1994-1995, simplified table.}

<table>
<thead>
<tr>
<th>natz</th>
<th>Coef.</th>
<th>Std. Err.</th>
<th>t</th>
<th>P&gt;t</th>
<th>[95% Conf. Interval]</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>dist</td>
<td>0.0000131</td>
<td>4.55E-06</td>
<td>2.89</td>
<td>0.005</td>
<td>4.09E-06</td>
<td>0.0000222</td>
</tr>
<tr>
<td>leastpop</td>
<td>0.2146127</td>
<td>0.0926663</td>
<td>2.32</td>
<td>0.023</td>
<td>0.0305424</td>
<td>0.398683</td>
</tr>
<tr>
<td>slowinc</td>
<td>-0.1921427</td>
<td>0.0875284</td>
<td>-2.2</td>
<td>0.031</td>
<td>-0.3660071</td>
<td>-0.0182783</td>
</tr>
<tr>
<td>married</td>
<td>0.212268</td>
<td>0.1118939</td>
<td>1.9</td>
<td>0.061</td>
<td>-0.0099955</td>
<td>0.4345315</td>
</tr>
<tr>
<td>age</td>
<td>-0.2400565</td>
<td>0.0872874</td>
<td>-2.75</td>
<td>0.007</td>
<td>-0.4134422</td>
<td>-0.0666709</td>
</tr>
<tr>
<td>reslength</td>
<td>0.5931288</td>
<td>0.0764187</td>
<td>7.76</td>
<td>0</td>
<td>0.4413324</td>
<td>0.7449252</td>
</tr>
<tr>
<td>cons</td>
<td>-0.0231449</td>
<td>0.1104806</td>
<td>-0.21</td>
<td>0.835</td>
<td>-0.242601</td>
<td>0.1963112</td>
</tr>
</tbody>
</table>
The most interesting change was that `leastpop` became more significant and positively signed—that is, the more evenly dispersed the immigrant population is, the higher the naturalization rate. At least for the 1994-1995 period, this indicates that community integration into US society may be more important than numerical strength, since small ethnic communities or even individual immigrant families are more likely to be assimilated into the general population of a place than large ethnic communities.

*Global analysis, 1996-2005*

Next, I repeated the analysis for 1996-2005. I found considerable changes in both which variables are significant and in the signs of certain variables, suggesting that the change in policy did in fact have an impact on the factors which predict naturalization rate. `Married`, `homeown`, `dist`, `swcoccs`, and `reslength` all remained positive as expected, while `age` and `female` were negatively correlated as in previous models, for similar reasons.

**Table 3: Regression analysis of global data, 1996-2005, simplified table.**

Number of obs: 740; R-squared = 0.6517

| natz     | Coef. | Std. Err. | T      | P>|t| | [95% Conf. Interval] | Beta  |
|----------|-------|-----------|--------|-----|----------------------|-------|
| dist     | 0.0000134 | 1.89E-06 | 7.09   | 0   | 9.71E-06 | 0.0000171 | 0.1861511 |
| leastpop | -0.099715  | 0.0391091 | -2.55  | 0.011 | -0.1764946 | -0.0229354 | -0.0634213 |
| slowinc  | 0.0634678  | 0.0345874 | 1.83   | 0.067 | -0.0044349 | 0.1313705 | 0.1055552 |
| married  | 0.1156055  | 0.040471  | 2.86   | 0.004 | 0.0361521 | 0.1950588 | 0.0778577 |
| homeown  | 0.2683896  | 0.0306489 | 8.76   | 0    | 0.2082191 | 0.3285601 | 0.2784606 |
| reslength| 0.4145164  | 0.0256805 | 16.14  | 0    | 0.3641001 | 0.4649328 | 0.4344808 |
| female   | -0.1469083 | 0.0441813 | -3.33  | 0.001 | -0.2336459 | -0.0601707 | -0.0786761 |
| age      | -0.3394182 | 0.0338636 | -10.02 | 0    | -0.4058999 | -0.2729366 | -0.2512061 |
| swcoccs  | 0.1251814  | 0.0323718 | 3.87   | 0    | 0.0616285 | 0.1887343 | 0.1055552 |
| _cons    | 0.0550139  | 0.049044  | 1.12   | 0.262| -0.0412703 | 0.1512981 |
However, leastpop changed from being moderately positively correlated with natz to slightly negatively correlate. This indicates that the 1996 change in policy turned ethnic concentration into an asset instead of a liability. Chavez et al (2006) suggest that this may be because of the anti-immigrant shift in public opinion surrounding the passage of the bill—when the government and surrounding communities have a negative attitude towards immigrants, “ethnic support systems become very important. Social capital depends less on the relative economic or occupational success of immigrants than on the density of ties between them.” It also explains the even smaller negative effect of leastpop on natz in the 1994-2005 model; the positive contribution from the 1994-1995 period canceled out some of the negative contribution from this model, pushing the coefficient closer to zero.

Another feature worth noting is that although slowinc is not quite significant, it is nevertheless positively correlated with natz. Even though one should not place too much importance on this finding, it provides some support for the idea that after 1996 the poor had greater incentive to naturalize, either to continue receiving government benefits or out of fear of more stringent deportation policy.

*Regional analysis, 1994-2005*

Finally, I turned to breaking down the data by world region of origin (Europe, Asia, Africa, Central America, South America, the Caribbean, or North America/Canada). I chose not to divide the regional data into pre- and post-1996 equations because the low number of observations for such a model (in most cases, less than 20
Table 4: Regression analysis of regional data, selected variables

<table>
<thead>
<tr>
<th>Region</th>
<th>Obs</th>
<th>R-squared</th>
<th>Variable name</th>
<th>Coef.</th>
<th>Std. Err.</th>
<th>P&gt;t</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td>217</td>
<td>0.7324</td>
<td>dist</td>
<td>0.0000860</td>
<td>0.0000130</td>
<td>0.0000</td>
<td>0.2713034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>age</td>
<td>-0.3857680</td>
<td>0.0656524</td>
<td>0.0000</td>
<td>0.3333778</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>reslength</td>
<td>0.4084828</td>
<td>0.0412199</td>
<td>0.0000</td>
<td>0.5182110</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>homeown</td>
<td>0.3441117</td>
<td>0.0503066</td>
<td>0.0000</td>
<td>-0.2365364</td>
</tr>
<tr>
<td>Asia</td>
<td>266</td>
<td>0.6122</td>
<td>dist</td>
<td>-0.000015</td>
<td>6.37E-06</td>
<td>0.02</td>
<td>-0.0925749</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>age</td>
<td>-0.2068699</td>
<td>0.0597193</td>
<td>0.001</td>
<td>-0.1413066</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>female</td>
<td>-0.3268452</td>
<td>0.0697513</td>
<td>0</td>
<td>-0.1959777</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>married</td>
<td>0.202603</td>
<td>0.0639139</td>
<td>0.002</td>
<td>0.1353732</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>homeown</td>
<td>0.3685861</td>
<td>0.0467036</td>
<td>0</td>
<td>0.3554188</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>reslength</td>
<td>0.4465374</td>
<td>0.0442062</td>
<td>0</td>
<td>0.4450155</td>
</tr>
<tr>
<td>Africa</td>
<td>59</td>
<td>0.6026</td>
<td>homeown</td>
<td>0.2601606</td>
<td>0.106816</td>
<td>0.018</td>
<td>0.4991525</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>age</td>
<td>-0.5272644</td>
<td>0.0970697</td>
<td>0</td>
<td>0.5053599</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>female</td>
<td>-0.3268452</td>
<td>0.0697513</td>
<td>0</td>
<td>-0.1959777</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>married</td>
<td>0.202603</td>
<td>0.0639139</td>
<td>0.002</td>
<td>0.1353732</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>homeown</td>
<td>0.3685861</td>
<td>0.0467036</td>
<td>0</td>
<td>0.3554188</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>reslength</td>
<td>0.4465374</td>
<td>0.0442062</td>
<td>0</td>
<td>0.4450155</td>
</tr>
<tr>
<td>Central America</td>
<td>101</td>
<td>0.8528</td>
<td>swcoecs</td>
<td>0.4596909</td>
<td>0.0686657</td>
<td>0</td>
<td>0.4078574</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>leastpop</td>
<td>0.5296533</td>
<td>0.0914094</td>
<td>0</td>
<td>0.2359817</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>age</td>
<td>-0.4494851</td>
<td>0.0906241</td>
<td>0</td>
<td>-0.2746565</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>married</td>
<td>0.2053628</td>
<td>0.0819815</td>
<td>0.014</td>
<td>0.1179716</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>slowinc</td>
<td>-0.1133538</td>
<td>0.0536301</td>
<td>0.037</td>
<td>-0.0980309</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>reslength</td>
<td>0.2760411</td>
<td>0.0591765</td>
<td>0</td>
<td>0.2382806</td>
</tr>
<tr>
<td>South America</td>
<td>107</td>
<td>0.5868</td>
<td>age</td>
<td>-0.3748895</td>
<td>0.1476872</td>
<td>0.013</td>
<td>-0.1809789</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>leastpop</td>
<td>-0.2775234</td>
<td>0.1195149</td>
<td>0.022</td>
<td>-0.1671502</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>slowinc</td>
<td>-0.390603</td>
<td>0.0859523</td>
<td>0</td>
<td>-0.3356296</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>reslength</td>
<td>0.4011394</td>
<td>0.0669152</td>
<td>0</td>
<td>0.429827</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>homeown</td>
<td>0.2142716</td>
<td>0.0816039</td>
<td>0.01</td>
<td>0.1978008</td>
</tr>
<tr>
<td>Caribbean</td>
<td>73</td>
<td>0.7848</td>
<td>swcoecs</td>
<td>0.2791149</td>
<td>0.110333</td>
<td>0.014</td>
<td>0.2016971</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>leastpop</td>
<td>0.5327022</td>
<td>0.1697687</td>
<td>0.003</td>
<td>0.2277463</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>age</td>
<td>-0.574223</td>
<td>0.1172342</td>
<td>0</td>
<td>-0.3546008</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>married</td>
<td>-0.4961354</td>
<td>0.1220767</td>
<td>0</td>
<td>-0.2801003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>homeown</td>
<td>0.1855805</td>
<td>0.0743884</td>
<td>0.015</td>
<td>0.2286987</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>reslength</td>
<td>0.3920713</td>
<td>0.1043163</td>
<td>0</td>
<td>0.3372025</td>
</tr>
<tr>
<td>N. Am./ Canada</td>
<td>15</td>
<td>0.9545</td>
<td>slowinc</td>
<td>-0.3788533</td>
<td>0.1126707</td>
<td>0.006</td>
<td>-0.3117061</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>leastpop</td>
<td>-0.3241167</td>
<td>0.0454412</td>
<td>0</td>
<td>-0.6935737</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>female</td>
<td>0.5805859</td>
<td>0.074121</td>
<td>0</td>
<td>0.8077259</td>
</tr>
</tbody>
</table>
observations for the pre-1996 model) caused serious stability problems. I found considerable variation between regions in which variables were significant predictors of naturalization rate as well as in their signs (positive or negative). **Homeown** and **reslength** were the two variables most strongly and consistently correlated with naturalization rate, while **age** was the most consistent negative predictor of naturalization. Table 4 provides a summary of the significant variables for each region.

One feature worth drawing attention to is the apparent tradeoff between the significance of **dist** and **leastpop**. With the exception of Africa, each region shows one or the other, but not both, as being significantly correlated with naturalization rate. Moreover, the direction of correlation with naturalization rate varies between regions: **leastpop** is a positive predictor of naturalization in the Caribbean and Central America, but negative for Canada and other North America and South America. This may have to do with the nature of concentrated ethnic populations from different countries. For Central America and the Caribbean, large ethnic communities may tend to become enclaves isolated from US society, composed mostly of poor, low-educated immigrants who are afraid to move beyond the boundaries of their ethnic community. Illegal immigration may also have an impact, since the majority of illegal immigrants come from these two regions. Besides lowering overall naturalization rate simply because they do not have the ability to naturalize, the presence of illegal immigrants discourages other community members from naturalizing as well—an immigrant may feel uncomfortable seeking out the very immigration and naturalization officers who have the power to deport his unauthorized brother. Large communities of North American or South American immigrants, in contrast, may be politically powerful or well-integrated into US
society. Perhaps in this case an established community could even encourage immigrants to stay in the US rather than returning to their native countries.

It turns out that the unexpected negative correlation between `female` and naturalization rate in the global model can be explained by regional variations as well. `Female` is actually slightly positively correlated (although only significant in North America) in every region except for Africa and Asia. Perhaps as Pantoja (2006) suggests, women are generally more socially active and thus become more integrated into the community and gain access to more resources in Western cultures. However, in more patriarchal societies, women may be restricted from accessing certain resources, their freedom may be limited in certain key ways, or they may not be able to naturalize independently of male relatives. In such ethnic groups, which are more predominant in Africa and Asia than in other regions, women become less likely to naturalize than men because they are faced with social barriers.

Another notable feature is the level of variation among the R-squared values. My model seems to fit the data from Central America, North America, and the Caribbean extremely well, but not so much Asia, Africa, or South America. Since I developed my theory with a focus on Latin American immigrants, it is logical that my theory fits the data from that region best.

I also analyzed the data from Mexico separately. Because I only had 12 observations to work with (one from each year), all from the same country, most of the SES variables ended up being highly correlated. After testing several different combinations, the only consistently significant variables were `homeown`, `slowinc`, and `female`. `Homeown`, however, is strongly positively correlated (0.75 or greater) with
leastpop, swnoccs, and age, as well as being strongly negatively correlated with slowinc and swelfare. This is reflected in the correlations between these variables and natz (see Table 6). Interestingly, for Mexico reslength is strongly correlated with neither homeowner nor natz. Perhaps for Mexican immigrants, length of residence is unimportant because return to Mexico is so easy and illegal immigration is so prevalent. Homeownership is a better predictor of naturalization partly because it indicates an intention to stay in the US rather than returning to Mexico, and partly because it may suggest legal status. Illegal immigrants’ typically low wages coupled with the legal proceedings involved make them unlikely to buy homes.

The strong negative correlation between percent of population below poverty level and naturalization rate supports previous authors’ research. In fact, for Mexican immigrants the beta coefficient suggests that community poverty status is the strongest predictor of naturalization rate. More than other immigrant groups, for Mexican-Americans poverty is often associated with illegal status, migrant workers, or residence in poor ethnic communities, all of which my theory associates with low naturalization rates. And, as Pantoja (2006) predicted, female is positively correlated with naturalization rate, likely because in Western societies women tend to form social networks in the community that provide them with greater access to resources.
Table 5. Regression analysis for Mexico only, 1994-2005, simplified table.

No. of obs = 12, R-squared = 0.9071

<table>
<thead>
<tr>
<th>natz</th>
<th>Coef.</th>
<th>Std. Err.</th>
<th>t</th>
<th>P&gt;t</th>
<th>[95% Conf. Interval]</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>slowinc</td>
<td>-0.4871815</td>
<td>0.1852207</td>
<td>-2.63</td>
<td>0.03</td>
<td>-0.9143012 -0.0600618</td>
<td>-0.6122455</td>
</tr>
<tr>
<td>homeown</td>
<td>0.6952341</td>
<td>0.2938575</td>
<td>2.37</td>
<td>0.046</td>
<td>0.0175975 1.372871</td>
<td>0.5092259</td>
</tr>
<tr>
<td>female</td>
<td>0.6207872</td>
<td>0.2533352</td>
<td>2.45</td>
<td>0.04</td>
<td>0.0365952 1.204979</td>
<td>0.3210961</td>
</tr>
<tr>
<td>_cons</td>
<td>-0.1227937</td>
<td>0.2132859</td>
<td>-0.58</td>
<td>0.581</td>
<td>-0.6146319 0.3690446</td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Correlation table for Mexico only, 1994-2005

<table>
<thead>
<tr>
<th></th>
<th>natz</th>
<th>leastpop</th>
<th>educ</th>
<th>slowinc</th>
<th>homeown</th>
<th>swcoecs</th>
<th>female</th>
<th>reslength</th>
<th>age</th>
<th>swelfare</th>
</tr>
</thead>
<tbody>
<tr>
<td>natz</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>leastpop</td>
<td>0.8495</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>educ</td>
<td>0.4388</td>
<td>0.6352</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>slowinc</td>
<td>-0.8713</td>
<td>-0.7436</td>
<td>-0.6442</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>homeown</td>
<td>0.8939</td>
<td>0.8677</td>
<td>0.5352</td>
<td>-0.8645</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>swcoecs</td>
<td>0.7951</td>
<td>0.8254</td>
<td>0.4178</td>
<td>-0.7871</td>
<td>0.7567</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>female</td>
<td>-0.2538</td>
<td>-0.3529</td>
<td>-0.8786</td>
<td>0.5643</td>
<td>-0.4504</td>
<td>-0.2322</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reslength</td>
<td>-0.1331</td>
<td>0.1086</td>
<td>-0.3911</td>
<td>-0.9071</td>
<td>0.0041</td>
<td>-0.0960</td>
<td>0.3592</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>0.6927</td>
<td>0.7693</td>
<td>0.4105</td>
<td>-0.6417</td>
<td>0.8194</td>
<td>0.8220</td>
<td>-0.4004</td>
<td>-0.1188</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>swelfare</td>
<td>-0.6792</td>
<td>-0.7807</td>
<td>-0.8863</td>
<td>0.8307</td>
<td>-0.719</td>
<td>-0.6868</td>
<td>0.7495</td>
<td>-0.1309</td>
<td>0.6132</td>
<td>1.0</td>
</tr>
</tbody>
</table>
IV. Conclusion

§1: Factors Correlated with Community Naturalization Rate

In interpreting these results, since homeown and, to a lesser extent, married were both designed as measures of an immigrant community’s relative integration into the US, their positive correlations with community rate of naturalization in the majority of the models suggest that naturalization rate does in fact increase as immigrants become more attached to and involved in US society. Homeownership and marriage are also signs of putting down “roots” in the US that indicate an intention to remain in the US permanently, and US citizenship is much more attractive to an immigrant who intends to live in the US for the rest of his life than to one who intends to retire to his home country after a few years. However, higher rates of marriage alone do not categorically indicate a higher community naturalization rate—in the case of the Caribbean, married is actually negatively correlated with rate of naturalization. Perhaps many families send one spouse to the US to work for a period of time, and therefore marriage represents a tie to one’s home country rather than roots in the US.

Reslength is another way of measuring an immigrant’s integration into US society, since the longer an immigrant has been in the US, the more familiar he or she becomes with US institutions and customs. Predictably, it is also positive in all cases.

The relatively strong negative correlation between age and natz in all cases indicates that the most mobile segment of the labor force naturalizes at lower rates than the population at large. This population is more likely to come to the US primarily in search of a job, not a new life in the US, and more likely to intend to return to their home
country eventually. Whether white-collar workers or migrant farm workers, these economic immigrants probably do not seek out resources outside of their ethnic community or try to learn more about US society than they must know to perform their job. If an immigrant community in the US is largely made up of transient workers, the community is also less likely to gain the type of political experience and resources necessary to support its permanent population in the naturalization process.

However, not all variables were consistent across all models. Between the global model for 1994-1995 and 1996-2005, for example, leastpop changed from being positively signed for the 1994-1995 model to negatively signed for the 1996-2005 model. The significance of leastpop, then, may depend on how effective individual characteristics are in determining naturalization rate. After the 1996 legislation was passed restricting federal means-tested benefits to US citizens only and broadening the criteria for deportation, along with signaling general anti-immigrant public opinion, immigrants of all socioeconomic levels had new incentive to naturalize. Poor immigrants had even greater incentive to naturalize in order to retain any federal benefits they were currently receiving, or to assuage their fear of deportation. The greater individual incentives of the poor to naturalize statistically “cancelled out” the greater individual resources of immigrants with higher SES. Therefore, the difference between those immigrants who simply wanted to naturalize and those who were actually able to do so depended on the strength of their communities. An immigrant population that is more concentrated may have some political and social power based solely on the size of the community, no matter what their average SES; for example, the Spanish-speaking population in the US is large enough that many public signs and documents are printed in
both English and Spanish. A larger community based on size alone is also more likely to include citizen members who can assist other members in the naturalization process.

However, for Europe and Asia, it seems that distance from home country is a better predictor of naturalization rate than relative strength of community in the US. For Europe it is positively correlated, as expected, but for Asia distance is actually negatively correlated to naturalization rate. The fact that distance rather than population concentration predicts naturalization rate may be because immigrants from these two continents have been entering the US in large numbers for over two hundred years, and have had time to spread out and form ethnic communities across the US. For example, think of the Chinatowns in San Francisco, New York, and Chicago, or find the Italian Village in almost every major US city. As for the negative correlation between distance and naturalization rate for Asia, I can only speculate that the relatively high rates of naturalization among Chinese immigrants (in 2000, 50% of Chinese immigrants were naturalized, well above the national rate of 40.3%), coupled with low rates of naturalization from places like Malaysia, Indonesia, and Thailand (all around 25% in 2000) was probably the cause. The Asian case is interesting because Asian immigrants have consistently had the highest incomes, levels of education, and naturalization rates among all immigrants to the US. A few outlying countries like Malaysia, Indonesia, and Thailand which are relatively far away but send immigrants with very low SES and naturalization rates might have a significant impact on the dist variable.

There may also be cultural factors which affect naturalization rates. For example, a larger percentage of females in a community is positively correlated with rate of naturalization, in all regions except for Africa and Asia. This may be caused by the more
traditional, patriarchal societies that still exist in parts of Asia and Africa: if women have relatively less power, or are kept more isolated from society at large, then they are less likely to have access to resources than men. Therefore women become a liability in the naturalization process rather than an asset, because they must wait for their male relatives to begin the naturalization process for them.

§2: Shortcomings and Possible Future Improvements

Even after months of work on this project, it was impossible for me to fully research every aspect of my theory as fully as I would have liked. All the research I’ve done focuses almost exclusively on the ethnic community aspect of naturalization, analyzing community naturalization based on community characteristics. While it is a reasonable “first stab” at quantitatively evaluating the impact of ethnic communities and government policy on naturalization, in order to be rigorous it would be helpful to carry out an individual-level analysis. Instead of using individual-level survey data to correlate individual sociodemographic characteristics with naturalization, it could be possible to construct variables which measure both community resources and community size for each individual. This would provide a more specific and directly correlated measure of the impact of community on decision to naturalize. It may even be possible, but tedious, to do this with the IPUMS Census data, by analyzing the characteristics of ethnic communities in different census statistical areas.

Another systematic improvement would be to expand the analysis to all types of communities, not simply ethnic ones. I chose to consider only the community of immigrants with the same country of origin, which still provides useful data because this
is the community that most immigrants rely on in their first years in the US, and many
retain strong social and cultural ties to immigrants of the same ethnic group throughout
their residence in the US. However, the true community of many immigrants extends
beyond ethnic boundaries to include the towns in which they live and their workplaces,
which may be a mix of native-born Americans and immigrants with a range of different
cultures and heritages. I tried to account for this somewhat with my leastpop variable,
assuming that immigrants who were more evenly distributed across the country may be
more likely to interact with non-ethnic communities than those who were more
concentrated. However, a truly systematic study of the impact of community
characteristics on naturalization should include an analysis of all types of communities.

It may also be beneficial to redefine certain variables in my analysis. As I
mentioned earlier, a more rigorous measure of community strength than leastpop would
add weight to my conclusions. Similarly, future research could apply a more direct
measure of attachment to home country than distance. It could be the case that it is
relatively easy to travel from Delhi to New York because of the number of flights daily,
but it might be more difficult to travel between the US and certain South American
countries because of political instability. In this case, the Indian immigrant may have
more contact with and be more attached to a community in his home country than the
South American immigrant, despite the greater distance between India and the US.

It would also be interesting to compare the naturalization rates of those with
spouses residing in the US to those with spouses residing abroad, as well as specifically
analyzing the naturalization rates among women in different ethnic groups. Although
neither of these issues were integral to my thesis, they nevertheless bring up interesting questions that should be further investigated.

Finally, one of my original intentions was to expand my analysis to the 1900-2005 period, rather than just focusing on the last fifteen years, in order to assess the effects of several different government policy changes. Bloemraad (2006) provides an excellent study how the 1906 federal standardization of immigration and naturalization laws changed the, but it would be interesting to see whether further changes, and specifically the 1952 McCarran-Walter Act, had similar effects on the factors which influence naturalization.

§3: Final Summary

I began my research intending to support the theory that naturalization is determined by an immigrant’s access to resources, which in turn is influenced by individual skills and attributes, community factors, and government policy. Although I included some individual attributes in my analysis, I assumed that previous authors had adequately investigated the connection between individual characteristics and propensity to naturalize and instead focused on the community and government aspects of my theory. I considered intention to reside in the US permanently and integration into US society as two of the most important measures of an immigrant’s access to resources, and structured my data set and variables accordingly.

My results did in fact confirm the importance of community characteristics in the naturalization process. Distance from home country, which I interpreted as a measure of attachment to a community abroad and thus a negative indicator of permanence and
integration, proved significant in determining naturalization rate, as did the relative concentration of ethnic groups in the US. Age, which I originally did not consider as a particularly important variable, ended up having a significant impact on community naturalization rate as well. The population between the ages of 25 and 44 tends to be the mobile labor force, more likely to follow employment opportunities across national borders and less likely to intend to stay in the US permanently.

I also determined that government policy has a significant impact on what factors affect naturalization rate and in what direction. High ethnic population concentration had a negative effect on ethnic community naturalization rates before the 1996 legislation, but a positive effect after 1996. This is probably because the 1996 legislation changed the impact of individual characteristics on propensity to naturalize in such a way that community resources became more important. This may also explain why recent literature has focused more on community-based theories of naturalization than those studying the 1990 Census data.

Based on my findings, the low rate of naturalization among Mexican and Latin American immigrants can be attributed to not only their low average socioeconomic status, but also their lack of assimilation and permanence in the US, as indicated by the strong correlation between homeownership and naturalization for the Mexican model. Surprisingly, length of residence does not seem to impact naturalization among Mexican immigrants, perhaps because Mexican immigrants who do not naturalize after living in the US for ten years tend to be illegal immigrants, transient workers, or residents of ethnic enclaves who are unlikely to ever naturalize.
However, there remains much future research to be done on the impact of community characteristics on naturalization. This study is only a starting point; in order to fully understand the effects, more specific research is needed, from expanding the definition of “community” beyond ethnic groups to redefining key variables to be better predictors of the characteristics they are intended to measure.

What my research does suggest is that there is much more that the government could do to encourage immigrants to naturalize, especially on the community level. Immigrants will not choose to naturalize until they have the appropriate resources, including an understanding of why US citizenship is something to be attained and of their place as citizens in US society. While streamlining the bureaucracy involved in the naturalization process and offering assistance with naturalization applications would reduce some of the government barriers in the naturalization process, it is not enough. If integration and permanence are two of the best indicators of naturalization, then the government should create programs to encourage their development. As Louis DeSipio (2004) suggests, immigrants must be incorporated into US society and politics by pointing out their own roles in each. A larger English as a Second Language, or an expanded community outreach program to make government agencies more accessible to all ethnic groups could also aid in the assimilation process.

More fundamentally, the US should be inviting temporary workers to stay, rather than granting more nonimmigrant and temporary visas, or trying to restrict immigration from Mexico. Measures intended to limit immigration from Mexico have only resulted in a flood of illegal immigrants, which the government has consistently been unable to stem in any way. Large transient populations of immigrants fairly isolated from US society at
large negatively affect the permanence, integration, and naturalization rates of their entire ethnic community. Perhaps blanket legalization of illegal immigrants is not the solution, but neither is further restricting passage between the US and Mexico and constructing a wall between the two countries. If the government fears the “Mexicanization” of the Southwest, the best solution may be to focus on helping Mexican immigrants to participate as US citizens economically, culturally, and politically, encouraging their naturalization and integration into the US, rather than relegating them to the status of outsiders from Mexico.