I Z A

IZA DP No. 3596

Citizenship in the United States: The Roles of Immigrant Characteristics and Country of Origin

Barry R. Chiswick
Paul W. Miller

July 2008

Citizenship in the United States:
The Roles of Immigrant Characteristics and Country of Origin

Barry R. Chiswick
*University of Illinois at Chicago
and IZA*

Paul W. Miller
*University of Western Australia
and IZA*

ABSTRACT

Citizenship in the United States:
The Roles of Immigrant Characteristics and Country of Origin[*]

This study develops and estimates a model of the naturalization process in the US. The model is based on both the characteristics of immigrants and features of their countries of origin. The empirical analysis is based on the 2000 US Census. Both the characteristics of immigrants and the origin-country variables are shown to be important determinants of citizenship status. The individual characteristics that have the most influence are educational attainment, age at migration, years since migration, veteran of the US armed forces, living with family, and spouses' educational attainment. The country of origin variables of most importance are their degree of civil liberties and political rights, GDP per capita, whether the origin country recognizes dual citizenship, and the geographic distance of the origin country from the US.

Corresponding author:

Barry R. Chiswick
Department of Economics
University of Illinois at Chicago
601 South Morgan Street
Chicago, IL 60607-7121
USA
E-mail: brchis@uic.edu

CITIZENSHIP IN THE UNITED STATES: THE ROLES OF IMMIGRANT CHARACTERISTICS AND COUNTRY OF ORIGIN

I. INTRODUCTION

Compared to the burgeoning literature on immigrant labor market adjustment in the US, the citizenship literature is relatively sparse. Indeed, DeSipio (1987, p.402) surmised "...the social science literature on the naturalization process is weak and few statistically valid generalizations can be made about the effect of specific cultural, economic, political or familial variables on naturalization".

Yet naturalization rates among immigrants in the US have the potential to impact a wide range of economic, social and political outcomes, and this is reflected in the few academic studies that analyze them. Thus, Jasso and Rosenzweig (1986) investigated the determinants of naturalization rates because of their links to the family reunification provisions that dominate US immigration policy. Portes and Curtis (1987) see naturalization as a key factor in immigrants' political influence. Other authors (see DeSipio, 1987) view naturalization as a measure of adjustment or "Americanization".

Naturalization rates have also been investigated in order to understand the influence that public policy can have on immigrants' decisions to become a citizen. This has involved examination of policy change within a country, and also comparisons across countries. The Green Card Replacement Program, introduced in 1992, was expected to be associated with a greater propensity towards citizenship, as long-term permanent residents turned to naturalization rather than replace their cards (Jones-Correa, 2001). Similarly, the 1995 Citizenship USA program was aimed at encouraging naturalization. The 1996 Personal Responsibility and Work Opportunity Reconciliation Act (the Welfare Reform Act), which limited non-citizens' access to social services, was also expected to have a major impact on citizenship applications (Bloemraad, 2002; Jones-Correa, 2001). Comparisons between the US and Canada have been used to illustrate the apparently major role that institutional factors have on naturalization patterns (Bloemraad, 2002). Comparisons of the decisions of various arrival cohorts have been used to infer the

impact of changes in citizenship legislation in Germany (Constant, Gatavlina and Zimmermann, 2007).

The current paper seeks to provide a theoretical framework and quantitative overview of the incidence of naturalization (*i.e.*, the process by which immigrants become citizens) in the US. It does this through analysis of cross-tabulations from the 2000 US Census, a multivariate analysis of the same data, and by linking information about the origin countries of the immigrants to these data.[1] The plan of this paper is as follows. Section II provides a brief overview of the ways the citizenship decision has been modelled in the literature. Section III reviews the data on citizenship status among the foreign born, and introduces a series of cross-tabulations that highlight the considerable variations in rates of citizenship in the US. Section IV outlines the specification of the estimating equation, and discusses several key issues in estimation. The focus of this section is the incorporation of information on the countries of origin of immigrants in the US. This includes per capita incomes, measures of political, civil and economic freedom, linguistic distance, dual citizenship recognition, the status of English as an official language in the country of origin, and selectivity in migration and return migration, as measured by the geographic distance between the country of origin and the major port of entry in the US and by an index of sojourner behavior. Section V presents the statistical results from the analysis of the determinants of immigrants being naturalized. Section VI provides a summary and conclusion.

II. MODELING THE CITIZENSHIP DECISION

A. Conceptual Framework

Theoretically, when seeking to understand the citizenship decision one should focus on structural equations for the costs and benefits associated with citizenship. For a review of the costs and benefits, see Yang (1994). However, measures of these are not available in the data sets otherwise most amenable to detailed statistical analysis.

---

[1] It thus offers an update of Yang's (1994) model of citizenship for the US, a major extension to the set of origin-country influences incorporated into that model, and a refinement of the way that several of the origin-country variables have been measured in the naturalization literature.

Consequently, reduced form rather than structural equations are estimated.[2] The conceptual frameworks used to motivate these reduced form multivariate studies of citizenship have typically recognized the roles of four types of variables: personal characteristics of the immigrant; visa category; features of the country of origin (see Jasso and Rosenzweig, 1986; Yang, 1994) and ethnicity of the neighborhood in the destination country (Portes and Curtis, 1987; Yang, 1994). Evans (1988) further categorizes the personal characteristics as reflecting either commitment to the destination country or the immigrant's position in society (termed structural variables by Evans).[3] In addition, as differences in naturalization rates could be due to "characteristics common to immigrants from a certain country (*e.g.*, shared cultural attitudes or an inability to return home)" (Bloemraad, 2002, p.194), the set of personal characteristics needs to incorporate birthplace variables, or, as is a common practice, separate analyses need to be undertaken for major birthplace groups. Thus, the likelihood of an immigrant being a citizen can be expressed as:

$$
\begin{aligned}
\Pr(\textit{Citizen}) = f\,(&\textit{Commitment Variables},\ \textit{Structural Variables},\\
&\textit{Birthplace},\ \textit{Visa Category},\ \textit{Origin Country Characteristics},\\
&\textit{Ethnicity of Neighborhood})
\end{aligned}
\tag{1}
$$

Evans' (1988) commitment variables were held to incorporate direct and indirect measures of the extent of one's commitment to the destination country, and comprised Home Ownership, Speaks Only English, No Religion, Marital Status, and Years Since Migration. The Structural (*i.e.*, position in society) variables were Education Level, Family Income and Gender. The number or presence of children can be added to the former list, and age at migration can be added to the latter list: Portes and Curtin (1987) and Yang (1994) suggest that immigrants with children, especially those born in the

---

[2] DeVoretz and Pivnenko (2005), however, estimate a structural model that includes an expected wage differential. A mover-stayer type framework was applied in this study.

[3] Portes and Curtis (1987) also consider attitudes and orientations towards aspects of the destination society, though direct measures of these are seldom available. In the absence of direct measures, background characteristics, such as educational attainment, could be used to proxy attitudinal factors.

destination, have stronger roots to the destination country, while Jasso and Rosenzweig (1986) and Yang (1994) argue that immigrants who arrive when young are more likely to naturalize, as their attachment to the origin is weaker and they will gain the benefits associated with citizenship over a longer time period. Bloemraad (2002) expanded the list of structural variables to include employment status, while Yang (1994) included a variable for immigrants who served in the US Armed Forces as a measure of commitment and to capture the effect of the relaxation of the duration of residence requirements for citizenship for this group. Income and employment status variables may be endogenous in models of the citizenship decision, and, along with variables like occupation and self-employment status used by Yang (1994), are not considered in the current analysis, which can thus be viewed as a reduced form model.

Visa category has been included in studies of immigrant outcomes when it has been available (see, for example, Chiswick and Miller, 2006). Jasso and Rosenzweig (1986) argued that immigrants who enter a country on employment-related visas will have a higher probability of becoming citizens as a wider range of employment opportunities are open to citizens than to non-citizens.[4] Portes and Curtis (1987) also include the cost of naturalization (measured by being the spouse of a US citizen) in their analysis of Mexican immigrants in the US. There is a shorter required period of residence in the US (three years instead of five years) for the spouses of US citizens, which provides a "greater facility for citizenship change" (Portes and Curtin, 1987, p.365), although other factors, such as the reluctance of a US born spouse to leave the US, may be more important.

The extent of information flows, origin-country attractiveness, and direct costs of returning to the country of origin were the country of origin variables incorporated into the analysis by Jasso and Rosenzweig (1986). Specifically, they argued (p.301) that characteristics that make the country of origin "more attractive…in terms of social, economic, and political conditions, the less likely one is to naturalize". Moreover, immigrants with more information prior to immigrating were argued to be more likely to naturalize, as were those whose origin countries were more distant from the destination

---

[4] Jasso and Rosenzweig (1986) use Immigration and Naturalization Service data, and hence could explore the impact of visa category on the naturalization decision.

country (see also Pachon, 1987). Specifically, the country of origin variables included in the study by Jasso and Rosenzweig (1986) were: (i) GNP per capita; (ii) literacy rates; (iii) physical distance from the origin country to the nearest major port of entry in the US; (iv) centrally planned economy; (v) Voice of America Broadcasts; (vi) English as an official language; and (vii) US military presence.[5] Yang (1994) included origin-country variables for GNP per capita, socialist country, physical distance, English as an official language, refugee-sending country and dual citizenship recognition. The first four of these overlap variables employed by Jasso and Rosenzweig (1986), and the latter two are innovations of the Yang (1994) study. Jones-Correa's (2000) analyses also suggests that dual nationality practices of immigrants' origin countries will be an important consideration.

Portes and Curtis (1987) include a number of variables in the category of "Residential Patterns and Social Relations", namely size and location of the area of residence, the ethnicity of the neighborhood, opportunities to interact with Anglos, ethnicity of employers and co-workers, and the number of relatives and friends. These measures were collected in the data set they used. The ethnicity of the neighborhood in Portes and Curtis' (1987) study of citizenship among Mexican immigrants was coded 1 if the neighborhood was "Anglo" and zero otherwise. Some counterparts can be constructed with the census data used in the current and other studies. Thus, Yang (1994), for example, includes a variable for the number of immigrants from the same ethnic origin who immigrated before 1975 in his analysis of 1980 US Census data. Yang argued that an ethnic neighborhood could either shelter the immigrant from the negative consequences of not being a citizen, or assist aspects of assimilation, including the acquisition of citizenship. There was no regional dimension to this variable, and a preferable approach appears to be to use methods analogous to Chiswick and Miller

---

[5] As well, Jasso and Rosenzweig (1986) distinguish Western Hemisphere immigrants from other source regions, owing to their exclusion from family reunification visas in the US prior to 1978. This variable should not be relevant to analysis of the 2000 US Census. The only personal characteristic (other than visa class) included in their analysis was age at migration, although the authors conducted separate analyses for males and females, and examined the probability of naturalization at a specific length of residence in the US (10 years).

(2005a), where a neighborhood variable is constructed with reference to a specific area of current residence (State or PUMA).[6]

Thus, a reasonably encompassing model of citizenship for immigrants in English-speaking countries can be described as:

Pr(*Citizen) = f* (Education, Marital Status, Age, Years Since Migration, Home Owner, Speaks Only English, Gender, Presence of Children, Veteran, Birthplace, Visa Category, GNP/Capita*, Literacy Rates*, Physical Distance*, Centrally Planned Economy*, English Speaking*, US Military Presence*, Refugee-sending*, Sojourner Propensity*, Dual Citizenship*, Ethnicity of Neighborhood*)   (2)

where variables with an asterisk are the origin-country characteristics discussed above. With minor modifications this model can be applied to the determinants of citizenship for any destination.

B. Empirical Evidence

The multivariate studies of the factors influencing naturalization have shown that both the individual-level and country-level variables included in equation (2) are important, though the studies are generally incomplete in terms of the range of variables included in any particular estimating equation. DeSipio's (1987) early review concluded (p.396) that "…the findings of these studies are often contradictory and the levels of confidence of the majority of their findings are low".

(i) *Individual Characteristics*

Evans' (1988) analyses, based on individual-level variables only, revealed a positive relationship between duration of residence in Australia and the probability of being an Australian citizen. Australian citizenship was less likely among immigrants from English-speaking countries than among immigrants from non-English-speaking countries. However, among the latter group, immigrants who spoke only English at

---

[6] Yang (1994) also included a variable for the percentage of the immigrant group living in urban (*i.e.*, central city) areas. Again, this variable is not described as having a regional dimension.

home had a greater probability of becoming a citizen. The other control variables for commitment included in the analysis (marital status, homeowner, no religion) did not appear to affect the likelihood of becoming a citizen. Moreover, none of the structural variables had a systematic impact on citizenship acquisition across the immigrant groups examined, suggesting there is no association between social stratum and citizenship[7].

Portes and Curtis' (1987) study of a small sample of Mexican immigrants in the US was also based on individual-level variables, and the survey used contained a particularly rich set of these. Only a few of the variables, however, were significant predictors of naturalization: visa category (spouse of US citizen), knowledge of English, home ownership, and number of children. The latter two were given a commitment interpretation in the study.

Bloemraad (2002) reports that among adult Portuguese immigrants in Canada and the US, citizenship increased with duration of residence, age (to 49 years), and educational attainment, and was more likely among those who spoke English and those who were homeowners. Gender, marital status, employment status and income were not significant determinants of whether immigrants were citizens.

An array of findings has also emerged from the literature examining the determinants of citizenship in Europe. Constant, Gatavlina and Zimmermann (2007) report that females, the more highly educated and immigrants having close German friends were all more likely to become citizens in Germany. Country of origin was also an important determinant of naturalization. Age at arrival was negatively related to citizenship acquisition, as was duration of residence. This latter result was attributed to immigrants' responses and adaptations to citizenship legislation: German citizenship law was revised in 1999 to reduce the residence requirements from 15 years to eight years.

Fougère and Safi's (2008) study for France also shows that country of origin is a major determinant of citizenship. Immigrants with higher levels of skill (indexed by educational attainment and occupation) and females were more likely to become citizens. However, the impact of age at migration and marital status varied considerably between males and females. Similar to the finding in Constant *et al*. (2007), immigrants arriving

---

[7] Evans (1988) conducted separate analyses for five immigrant groups: (i) Anglophones; (ii) Mediterranean; (iii) Northwest Europe; (iv) Third World; (v) East Europe.

in France after 1982 were shown to have higher probabilities of citizenship than those who arrived before 1982. A more conventional period of immigration effect is reported in Bevelander and Veeman's (2006) study for the Netherlands. This study also shows that country of origin is a major determinant of citizenship. However, while the incidence of citizenship was found to increase with educational attainment, they found that gender was not an important determinant of citizenship.

(ii) *Institutional Characteristics*

A feature of Bloemraad's (2002) study is that the US and Canadian census data for Portuguese immigrants (living in Massachusetts or Ontario) were pooled, and a location variable used to isolate "institutional" influences on the naturalization decision.[8] The results revealed that living in Ontario had a positive impact on the probability of being a naturalized citizen, which indicated the importance of institutional explanations for differences in the incidence of citizenship in the US and Canada. Bloemaard (2002) argued that the Canadian government has an interventionist relation with Canada's immigrants which promotes close ties between the state and the migrant organizations. She noted that the Canadian government provides symbolic and material support for ethnic associations and leaders, whereas the American government was argued to have autonomous and neutral relations with immigrants that render a disconnect between government and ethnic organizations. This latter characteristic of a state does not promote naturalization as much as the interventionist government, and this appears to be what the statistical analysis of the pooled data for Canada and the US reveals.

(iii) *Visa Category*

The visa category and origin-country variables discussed above in relation to equation (2) are a focus of the study for the US by Jasso and Rosenzweig (1986). They indicated that visa category was an important determinant of the propensity to naturalize, though the ranking of naturalization rates by visa category was shown to differ for male and female immigrants. Bevelander and Veenman (2006) report that refugees are much

---

[8] Limiting the analysis to one group of immigrants (Portuguese) controlled for "group-level" factors often discussed in the literature, while the other explanatory variables included in the analysis controlled for individual-level factors. The location variable was thus interpreted as an institutional factor.

more likely to obtain Dutch citizenship than other immigrants. Unfortunately, information on visa type is seldom available in the large data sets that are otherwise most useful for research into immigrants' economic and social outcomes, including the Census data used in this study.

(iv) *Origin Country Variables*

Jasso and Rosenzweig (1986) reported that GNP per capita, literacy rates, physical distance between the origin country and the major port of entry in the US, centrally planned economy and Voice of America Broadcasts were consistently significant in the statistical analysis, though the seemingly universal coverage of Voice of America broadcasts (in the origin-country language) limits the relevance of this variable in a contemporary study. The variables for English as an official language and US military presence were not consistently significant in their study.

In comparison, the variable English as an official language was highly significant (and negative) in Yang's (1994) study. The other five origin-country variables included in Yang's study were also statistically significant. They showed that per capita GNP of the origin country was associated with lower rates of citizenship, socialist countries and refugee-sending countries were associated with higher rates of citizenship, while there was a positive association between the naturalization rate and the geographic distance between the origin country and the closest major port of entry in the US. In contrast to Jones-Correa (2001) and Mazzolari (2007), Yang (1994) reports that immigrants from countries that allow dual citizenship have lower naturalization rates in the US.[9] It is also noted that the individual characteristics in Yang's (1994) model were generally highly significant, with impacts in the expected direction. Comment is provided in the discussion of the multivariate findings below.

Finally, living in an Anglo neighborhood was associated with greater rates of citizenship in the study by Portes and Curtis (1987). It was argued (p.365) that this came about through exposure leading to "greater knowledge of American society and a greater appreciation of the benefits of citizenship". While the neighborhood ethnicity variable was statistically significant in some of the initial models discussed in Portes and Curtis

---

[9] This is argued by Yang (1994, p.474) to reflect immigrants from countries that allow dual citizenship perceiving this as entailing responsibilities rather than benefits.

(1987), it was not included in their final "parsimonious" model. In contrast, Yang (1994) and Fougère and Safi (2008) report that the larger an ethnic group size, the higher the naturalization rate.[10]

Thus, this review of the major studies of naturalization patterns reinforces the conclusion of DeSipio's (1987) survey: that few individual-level variables are systematically related to citizenship status. It shows an apparently stronger role for origin-country influences, though as with the individual-level variables, there are differences in results across studies, and several of the findings (*e.g.*, that in relation to dual citizenship in Yang's (1994) study) appear anomalous. The analyses below offer a more encompassing study, which provides the basis for a re-assessment of the roles of a number of the determinants of citizenship.

III. DATA AND CROSS-TABULATIONS

A. Requirements for Naturalization

Table 1 provides information on the basic requirements for immigrants to become citizens of the US.[11]

---

[10] The finding for Fougère and Safi (2008) refers to the relative size of the ethnic community variable as opposed to their variable for the number of foreigners in the area, which was intended to measure the deterrent effect of the length of the potential queue.

[11] Persons born in the US are automatically citizens, regardless of their parents' nationality, nativity, visa status or legal status.

Table 1
Eligibility Requirements for Citizenship in the US

Criteria	Comments
Minimum Age	18 years.
Status	Must be a legal permanent resident.
Residency Requirement	Must have lived in the US as a permanent resident alien for at least five years, with absences totalling no more than one year, and have residence in one state for at least three months. There are a number of exceptions under which one may apply for citizenship before 5 years of residence are completed. The most common exceptions are the spouses of US citizens and veterans of the US armed forces who can apply after 3 years of residence.
Language Requirement	Must show ability to read, write, speak and understand "ordinary" English (exceptions: those over 55 years of age living in the US 15 years or more, those over 50 living in the US for 20 years of more, those with impairments).
Knowledge Requirement	Must demonstrate knowledge and understanding of fundamentals of US history and government ("special consideration" is given to those with impairments or over 65 years of age with at least 20 years of residence).
Grounds for Refusal	Certain criminal offences, and/or a failure by the candidate to show that they are of "good moral character".
Oath of Allegiance	Required.
Cost	$595 plus a biometrics fee of $80, so the total fee is $675.
Dual Citizenship	The oath of allegiance includes a phrase renouncing "foreign allegiances", but there is no enforcement of this provision.

Source: Updated from Bloemraad (2002), Table 2.

To become a citizen a person must be a permanent resident alien (generally for at least five years), be at least 18 years of age, demonstrate a basic knowledge of English and of US history and government, and be of "good moral character." The oath of allegiance requires the renouncement of foreign allegiances, but there is no apparent enforcement of this provision. However, there is a potentially important role for the permitting of dual citizenship in the country of origin. As Bloemraad (2002, p.205) argues, "…the critical issue is not whether the host country recognizes dual citizenship, but whether the home country does", as some immigrants may lose, or perceive that they will lose, the benefits of citizenship of their country of origin when they become US citizens.[12]

---

[12] Pachon (1987, p.305) states "Many new immigrants also believe that acquiring U.S. citizenship means giving up rights and benefits in their native countries. Mexican immigrants, for example,

B. Descriptive Statistics

Using simple cross-tabulations, this section examines variations in rates of citizenship across birthplace groups and according to a number of other key personal characteristics of immigrants in the US. The data for the analyses are from the 2000 US Census Public Use Microdata Sample (PUMS), 1 percent file. This data set contains information on whether immigrants are naturalized, together with information on key factors that are of interest when discussing the take-up of citizenship, namely birthplace, duration of residence in the country, proficiency in the English language, age, gender and educational attainment.[13] There are no data on the year in which the immigrant became a US citizen.

The cross-tabulations that follow are limited to the foreign born in the US Census aged 25 to 64 years who arrived in the US as adults (*i.e.*, aged 18 or more years). This age bracket covers the groups with most "choice" with respect to the citizenship decision.[14] Moreover, given the most general residence requirements for citizenship (see Table 1), the information in these cross-tabulations is restricted to immigrants with more than five years of residence in the US.

The PUMS from the 2000 US Census has a very extensive birthplace categorization, containing over 200 country codes. While use is made of this in the multivariate analysis (see Section V), the birthplace data are aggregated into 22 broad birthplace regions for the descriptive analyses of this section. These birthplace regions are based on Chiswick and Miller (2008). Information on rates of US citizenship for 25-64 year olds (who arrived in the US aged 18 or more years) living in the US for at least

---

are often under the mistaken impression that they will forfeit all of their rights to property ownership if they relinquish Mexican citizenship".

[13] Jasso and Rosenzweig (1986, p.301) argue that census data have limited appeal for the study of the determinants of citizenship as "…the Census does not identify those foreign-born persons eligible to naturalize; moreover, in a given Census year, only those survivors of past cohorts of immigrants who did not emigrate from the United States are represented". Other authors (Bloemraad, 2002; Evans, 1988; Yang, 1994), however, use the Census data largely without reservation.

[14] Jasso and Rosenzweig (1986) focus on immigrants who were 21-65 years of age at admission, Evans (1988) restricted her sample to 20-64 year olds, Portes and Curtis (1987) examined Mexican immigrants aged 18-60 years.

five years from each broad birthplace region is presented in Table 2, in total and separately for males and females.

Table 2
Rates of US Citizenship, 25-64 Year Old Immigrants, by Major Birthplace Region, 2000 US Census

Birthplace Region	Rate of Citizenship			Relative Frequency[a]
	Males	Females	Total	
United Kingdom	33.64	44.50	39.98	2.08
Ireland	42.97	54.86	49.23	0.59
Western Europe	41.05	48.44	45.85	2.98
Southern Europe	62.25	61.20	61.75	3.28
Eastern Europe	53.54	56.09	54.84	2.55
Former USSR	62.04	60.61	61.28	2.30
Indo China	69.20	61.78	65.39	5.24
Philippines	70.83	71.28	71.11	5.90
China	57.78	63.77	61.00	5.57
South Asia	54.99	54.97	54.98	5.27
Other Southern Asia	47.93	49.80	48.99	1.29
Korea	51.37	58.87	56.04	3.07
Japan	24.61	32.37	29.80	0.90
Middle East	70.26	69.80	70.07	3.82
Sub-Saharan Africa	51.41	46.60	49.45	2.48
Canada	33.13	36.92	35.28	1.87
Mexico	25.66	26.57	26.08	26.08
Cuba	48.82	63.09	55.80	2.19
Caribbean	45.90	53.01	49.79	8.12
Central and South America-SP[b]	34.06	40.40	37.37	12.95
Central and South America-Non-SP	63.89	62.93	63.36	0.94
Australia, New Zealand, Oceania	36.71	41.99	39.56	0.53
Total	**44.23**	**47.94**	**46.13**	**100.00**

Note: (a) = Refers to proportion of the foreign born aged 25 to 64 who immigrated at age 18
              or more years and who have resided in the US for five or more years;
         (b)   SP = Spanish speaking.
Source: 2000 United States Census, 1% PUMS.

The mean rate of citizenship among the foreign born in the US is 46.1 percent. For females the rate of citizenship is 47.9 percent, and for males it is four percentage points lower, at 44.2 percent. Immigrants from Mexico have the lowest rates of citizenship: 25.7 percent for males and 26.6 percent for females. This low rate is likely due to proximity of the origin country, which increases the probability of return migration and repeated to and fro migration, and the large number who are ineligible for naturalization because they are in an illegal status (see Pachon, 1987; Jasso and Rosenzweig, 1986). However, they do not by themselves account for the overall low rate of citizenship in the US.   In the absence of immigrants from Mexico, the mean

citizenship rate in the US would only be 53.2 percent (compared with 46.13 percent across all immigrants).

The other birthplace-gender groups with low rates of citizenship (below 40 percent) are male and female immigrants from Japan, male and female immigrants from Canada, and male immigrants from Spanish-speaking countries in Central and South America and the UK, Australia and New Zealand. The birthplace-gender groups with high rates of citizenship (or around 70 percent) are male and female immigrants from the Philippines and the Middle East and male immigrants from Indo China. In the case of Japan, the low rate especially for men (only 25 percent) may be associated with the prohibition of dual citizenship among Japanese nationals who live abroad and the relatively high proportion of individuals on temporary visas for professionals and managers.[15] The rate of citizenship is generally higher among female immigrants, although Indochina and Sub-Saharan Africa are noticeable as exceptions to this pattern.

Of course, as noted in Section II, other factors besides country of origin may have a role to play in the naturalization decision. The roles of duration of residence, proficiency in English, and several other possible determinants of citizenship among immigrants in the US are explored in Tables 3 to 6.

Table 3 presents the US citizenship rates of immigrants by their English language proficiency (see requirements in Table 1). It is interesting to note that immigrants who speak another language at home but speak English very well are the most likely to take up US citizenship (57 percent). This rate is eight percentage points higher than that of monolingual English speakers (49 percent). The degree of English language proficiency has a positive association with the rate of citizenship in the US, with the naturalization

---

[15] Information on the conditions under which countries allow dual citizenship is difficult to compile. See Renshon (2001) for relevant information. Jones-Correa (2001) provides a detailed account of the impact of dual citizenship provisions on the propensity of immigrants to acquire US citizenship. Using Immigration and Naturalization Service data from 1965 to 1997, which span the period when nine countries acknowledged dual nationality (and so the data permitted before and after comparisons of the impact on naturalization rates), Jones-Correa (2001) shows that the recognition of dual nationality by sending countries has a positive association with naturalization rates in the US. Moreover, the effect is shown to be greater where the dual nationality policy was the result of pressure from immigrant communities in the US rather than being an initiative of the government of the foreign country. Mazzolari (2007) arrives at a similar finding to Jones-Correa (2001), though Yang's (1994) results contrast with these positive impacts of dual citizenship rights.

rate decreasing from 57 percent for immigrants who speak another language at home but speak English very well, to 54 percent, 34 percent and 15 percent, respectively, for immigrants who speak another language at home but speak English well, not well, and not at all. Similar patterns of citizenship rates by proficiency in English are apparent for males and females separately. These could be associated with the language requirements for citizenship and with the shorter duration of residence in the US of those with limited English skills.

Table 3
Rates of US Citizenship, 25-64 Year Old Immigrants by English Language Proficiency, 2000 US Census

English Proficiency	Rate of Citizenship			Relative Frequency[a]
	Males	Females	Total	
English Only	45.64	51.26	48.74	14.63
Very Well	55.85	59.05	57.43	28.75
Well	50.30	58.09	53.99	25.28
Not Well	30.31	36.78	33.66	22.29
Not at All	14.19	15.24	14.82	9.05
Total	**44.23**	**47.94**	**46.13**	**100.00**

Note: (a) = Refers to proportion of the foreign born aged 25 to 64 who immigrated at age 18
               or more years and who have resided in the US for five or more years.
Source: 2000 United States Census, 1% PUMS.

Table 4 reports the rates of US citizenship of immigrants in 2000 by year of arrival in the US. Note that, for completeness, the table also contains information on immigrants who arrived in 1995-2000. This year of arrival category will be dominated by immigrants who are not eligible for citizenship, as only the spouses of US citizens, immigrants with service in the US Armed Forces, and some refugees with fewer than five years residence in the US can be naturalized. The discussion will therefore focus on the other arrival cohorts. These data show that there is a positive relationship between the rate of citizenship and the immigrants' duration of residence in the US. The rate of citizenship increases from 21 percent for those who arrived in the US between 1990 and 1994 to 77 percent for those arrived in 1969 or earlier. This strong relationship holds true for both male and female immigrants. As duration of residence initially increases beyond five years there are very rapid increases in the incidence of citizenship, but this rate of increase slows as duration of residence increases. This pattern has been remarked upon previously (see Jasso and Rosenzweig, 1986). Apparently this is due to those with the

greatest demand for US citizenship applying relatively quickly after they become eligible rather than delaying their application.

Table 4
Rates of US Citizenship, 18-60 Year Old Immigrants by Year of Arrival in the US,
2000 US Census

Year of Arrival	Rate of Citizenship			Relative Frequency[a]
	Males	Females	Total	
1995-2000	5.28	5.70	5.48	24.53
1990-1994	20.46	22.20	21.37	21.22
1985-1989	37.23	39.87	38.54	17.98
1980-1984	53.97	57.50	55.69	13.69
1975-1979	61.89	67.28	64.58	9.15
1970-1974	65.43	72.03	68.87	6.31
1969 or earlier	74.72	78.40	76.85	7.13
Total[b]	**34.29**	**37.99**	**36.16**	**100.00**

Notes: (a) Refers to proportion of the foreign born aged 25 to 64 who immigrated at age 18 or more years;
   (b) = Includes immigrants who arrived between 1995 and 2000.
Source: 2000 United States Census, 1% PUMS.

Table 5 examines the rate of citizenship of immigrants in the US by their marital status at the time of the census. It is observed that immigrants who are married (spouse present) have a greater rate of citizenship than their unmarried counterparts: 49 percent for the former group and 39 percent for the latter group. The same percentage (51 percent) of married male immigrants as of married female immigrants have taken out US citizenship, whereas among the non-married, there is a pronounced gender differential, with 46 percent of non-married female immigrants being citizens compared to only 33 percent of non-married male immigrants.

Table 5
Rates of US Citizenship, 18-60 Year Old Immigrants by Marital Status,
2000 US Census

Marital Status	Rate of Citizenship			Relative Frequency[a]
	Males	Females	Total	
Married	49.42	49.12	49.27	68.21
Non-married	32.75	45.50	39.41	31.79
Total	**44.23**	**47.94**	**46.13**	**100.00**

Note: (a) = Refers to proportion of the foreign born aged 25 to 64 who immigrated at age 18 or more years and who have resided in the US for five or more years.
Source: 2000 United States Census, 1% PUMS.

18

Table 6 presents the rates of citizenship for immigrants by their spouses' birthplace and citizenship status. Rates of citizenship are highest for married immigrants whose spouses are in the naturalized foreign-born group. This is the case overall (incidence of citizenship of 75 percent) and for both male and female immigrants (rates of citizenship of 78 percent and 73 percent, respectively). These rates of citizenship are considerably higher than those for immigrants whose spouses are native born, that is, citizens by birth (48 percent for males and 57 percent for females). Those who are married to foreign-born individuals who are not citizens of the US are least likely to take up US citizenship, with rates of citizenship of 26 percent for male immigrants and 19 percent for female immigrants. These patterns suggest that the citizenship decision is a joint-decision for a married couple, and it is possible that both would prefer to maintain similar citizenship status, and that they tend to apply together.

Table 6
Rates of US Citizenship, 18-60 Year Old Married Immigrants by Birthplace and Citizenship Status of Spouse, 2000 US Census

		Birthplace/Citizenship of Spouse			
	US		Overseas		Relative
Individual	Mainland	Citizen	Non-Citizen	Territories	Frequency[a]
Males	48.25	78.08	26.12	50.97	49.29
Females	57.09	72.85	19.16	50.87	50.71
Total	**53.54**	**75.34**	**22.95**	**50.91**	**100.00**

Note: (a) = Refers to proportion of the foreign born aged 25 to 64 who immigrated at age 18
             or more years and who have resided in the US for five or more years.
Source: 2000 United States Census, 1% PUMS.

Hence, this brief review of the variation in the incidence of citizenship in the US reveals pronounced differences by birthplace group, by year of arrival, proficiency in English, marital status and birthplace/citizenship of the spouse (among the married). The multivariate analyses that follow will build upon this information by including a wider set of individual characteristics and by also considering the roles that the characteristics of the countries of origin of the immigrants have on the citizenship decision.

IV.   SPECIFICATION OF ESTIMATING EQUATION

Two broad factors are considered in the multivariate analysis: the individual characteristics of immigrants, similar to those reviewed in Section III, and the characteristics of the countries of origin of the immigrants in the US. Appendix A provides definitions of all variables used in the analysis.

A. Immigrant Characteristics

The individual characteristics considered for inclusion in the model of citizenship are gender, educational attainment, age, years since migration, service in the US Armed Forces, family structure, speaks a language other than English at home, region of residence (Southern states, metropolitan area), whether the immigrant lived abroad 5 years ago, presence of children, married with spouse present, and, where relevant, spouse's educational attainment and whether the spouse is foreign born. The family structure variables distinguish immigrants who live in a family household (either a couple or lone-parent family), individuals living alone, or individuals living in other types of households.   Within family households, couple and lone-parent households are distinguished through the "married with spouse present variable", and in these instances information on the spouse is obtained. In addition, information on the presence of children of the wife is assigned to the husband. Immigrants who live in a family household, particularly those with children, can expect to have a greater commitment to the US and thus be more likely to naturalize. The variable for whether the immigrant lived abroad five years ago is an innovation in the current study. It provides a measure of "to and fro migration" which may reflect less commitment to the destination country. Moreover, absence from the US can lengthen the waiting period and delay the approval of the application for naturalization (Table 1).

Gender effects are incorporated into the statistical analyses presented below, both through the estimation of separate equations for males and females, and through estimation of equations on data pooled across males and females with a full set of gender interaction terms. Comment is also provided on the findings from an estimating equation based on data pooled across males and females that contains a dichotomous variable for females.

20

Those with higher levels of schooling would be more likely to naturalize for several reasons, including the higher level of ability, greater English language proficiency and greater ease in passing the history/civics portion of the naturalization test. Moreover, a wider range of job opportunities for the more highly educated citizens may provide an economic incentive for naturalization to increase with educational attainment. Based on the discussion presented in Section II, the incidence of citizenship is expected to be positively associated with service in the US Armed Forces, presence of children, married with spouse present, and living in a family household. This expectation is based on the lower costs of acquiring citizenship (*e.g.*, the lower residency requirements for immigrants married to a US citizen or with service in the US Armed Forces), and the lower probability of return migration for individuals with a greater commitment to the US, as proxied, for example, by military service, being married and having children and other family members in the household.

The incidence of citizenship is expected to be negatively associated with speaking a language other than English at home, as this may be indicative of a lower degree of assimilation and commitment to the US than in cases where English is spoken at home. A lower degree of commitment to the US is also the basis for the anticipated negative relationship between the likelihood of being a citizen and an immigrant living abroad five years ago. Similarly, there is a theoretical expectation that the incidence of citizenship will be negatively related to age at arrival (age when duration of residence is also included in the estimating equation). Immigrants who arrive at an older age have made greater investments specific to their country of origin, including social ties, and have a shorter time period over which any benefit associated with citizenship may be received. Jasso and Rosenzweig's (1986) findings support this expectation. However, Yang (1994) reported that the incidence of citizenship rises with age at migration up to an age of around 40 years, and then declines. A greater likelihood of citizenship among those who immigrate at an older age is possible where they feel more vulnerable given the recent welfare reforms, and turn to citizenship as a form of security. As these reforms are quite recent (1996), however, any influence attributed to them would be expected to be slight. A quadratic specification was used for age in preliminary estimations in this paper.

21

However, the results of this experimentation revealed that the estimating equation could be based on a linear age variable.

The incidence of citizenship is expected to be positively and strongly related to duration of residence, with the strongest effects occurring shortly after immigrants become eligible to acquire citizenship. Accordingly, the years since migration information is included in quadratic form: a non-linear pattern has also been reported in prior research (*e.g.*, Jasso and Rosenzweig, 1986; Yang, 1994).

There are several individual characteristics which have been used in previous studies but which have not been included in the estimating equation below. The main variables in this category are for home ownership, income, employment status, occupational status, and self-employment status. These variables are endogenously determined with citizenship status, and in the absence of identifying instruments that would enable this endogeneity to be accommodated, the preferred strategy is to work with a reduced form equation. Another omitted variable is visa status at entry. Unfortunately, the US Census does not ask the visa status at entry or at the time of the Census.

B. Country of Origin Variables

As discussed in Section II, previous studies of the determinants of citizenship have included a range of variables describing the country of origin of immigrants. This study uses a number of the variables that have been a feature of prior work, and introduces a series of new variables that can be used in place of several of the more crudely measured variables in previous studies.

New to this study are summary measures of Political Rights (PR) and Civil Liberties (CL), obtained from the publication *Freedom in the World* (see Appendix A). This publication covers the majority of the places of birth of immigrants in the US. It assigns a numerical rating—on a scale of 1 to 7—for political rights and an analogous rating for civil liberties, with a rating of 1 indicating the highest degree of freedom (*e.g.*, Switzerland, Australia in 2000) and a rating of 7 the least amount of freedom (*e.g.*, Afghanistan, North Korea in 2000). These data are available from 1973 for most countries.

The same publication combines these two ratings (PR and CL) to generate a third, encompassing, measure of whether a country can be is classified as Free, Partly Free, or Not Free. As the original PR and CL measures have greater variation they are used in the current study. The PR and CL measures (scaled from 1 to 7) replace the need to use dichotomous variables for whether an immigrant's country of origin was centrally planned (or Communist), and for whether the immigrant is from a refugee-sending country. The PR and CL measures are positively correlated, with a correlation coefficient of around 0.8 in most samples used below. While both are included in the estimating equations presented, results from when one of the variables is omitted are also discussed.

Another country-level variable that is of interest to this study is the measure of Economic Freedom (EF) obtained from Gwartney and Lawson's (2007) *The Economic Freedom of the World*. This uses a variety of data (see Appendix A) to rate the degree of economic freedom in a country on a 0 to10 scale, with 10 being most free and 0 the least (*e.g.*, 8.4 for Switzerland in the most recent data available, 5.3 for Nigeria and 4.4 for the Congo). This measure is available for only a subset (around 130) of the countries of origin of immigrants to the US, in 5-yearly intervals, from 1970. It is argued that economic freedom leads to economic prosperity, and consistent with this it was found that the measure of economic freedom was so highly correlated with the measure of per capita Gross Domestic Product, that only one of the variables could be included in the equation. As the measure of GDP per capita was available for a greater number of countries, and for a more extensive time period, it is the preferred measure.

Data on real Gross Domestic Product per capita were obtained from the Penn World Table. This source covers 188 countries and is available on an annual basis from 1950. The Penn World Table presents information in a common set of prices in a common currency so that real quantity comparisons can be made, both across countries and over time.

There are three further country of origin variables that are novel to the current study of citizenship. The first of these is a minority language concentration variable. This is a measure of group identity, and was formed by assigning each respondent a value equal to the percentage of the population aged 18 to 64 in the state in which he/she lives who reports the same non-English language as the respondent. In the construction of this

variable, only the twenty-five largest non-American Indian language groups nationwide are considered. These constitute 92 percent of all responses where a language other than English is used at home. Representation in the other language groups is so small numerically that the proportions are approximately zero, and this value is assigned. Those who reported speaking only English at home are assigned the mean value of the minority language concentration measure for other-language speakers of their birthplace group. However, all immigrants from predominately English-speaking countries are assigned zero for this variable. The variable is thus to be interpreted as an interaction between the English-speaking origin dummy variable (see below) and the minority language concentration variable. This measure has been used in the study of immigrants' language skills and earnings by Chiswick and Miller (2005a).

A measure of the propensity of the immigrant's birthplace group to engage in to and fro migration is also incorporated into the estimating equation. This is the proportion of the birthplace group that had first arrived in the US more than five years ago but which lived abroad in 1995, five years prior to the Census. This index of sojourner behavior will capture aspects of repeat migration, as well as temporary movements. It complements the individual measure of whether a specific immigrant had lived abroad five years ago. Like the individual-level measure, there is expected to be a negative association between the sojourner index and the naturalization rate.

The final variable that is introduced to the citizenship literature in the current study is a measure of linguistic distance (LD). This is included in the model of citizenship on the grounds that individuals with mother tongues that are more distant from English will have greater difficulty learning English and may have attitudes, orientations and patterns of assimilation that differ from those of immigrants whose mother tongues are closer to English, and these differences may influence the citizenship decision.[16]

The measurement of LD is based on the difficulty of learning a foreign language for English-speaking Americans, and is derived from a set of language scores (LS) measuring achievements in acquiring speaking proficiency in foreign languages by English-speaking Americans at the US Department of State, School of Language Studies,

---

[16] Portes and Curtin (1987) argue that attitudes and orientations towards the destination country will impact on the citizenship decision.

reported by Hart-Gonzalez and Lindermann (1993). It is described in detail in Chiswick and Miller (2005b), and the main feature is that it is measured on a scale from 1 to 3 (in increments of 0.25), with a score of 1 for languages that are most distant from English (*e.g.*, Korean, Japanese), and a score of 3 for languages that are closest to English (*e.g.*, Norwegian, Swedish). Foreign-born persons from non-English speaking countries who speak only English at home are assigned the mean value of the linguistic score measure for individuals reporting a foreign language from their birthplace group.

The reciprocal of the LS scores from Hart-Gonzalez and Lindermann (1993) is used as the measure of linguistic distance (*i.e.* LD = 1/LS), so that the more distant languages (*e.g.* Korean) are given a value of 1 and the languages closest to English (*e.g.*, Swedish) are given a value of 1/3. Consistent with this, all immigrants from predominately English-speaking countries are assigned zero for the linguistic distance variable.

The remaining three origin-country variables have been used in earlier studies. They are for the status of English as an official language in the origin country (Yang, 1994), the geographic distance of the origin country from the closest major port of entry into the US (Jasso and Rosenzweig, 1986; Yang, 1994)), and whether the origin country recognizes dual citizenship (Yang, 1994; Jones-Correa, 2001; Mazzolari, 2007).

Data for the English-speaking countries were obtained from the official languages for different countries listed in Banks' (1988) *Political Handbook of the World: 1988*, which is the same source as used by Yang (1994). Note that in instances where a country happens to have two or more official languages and English is only one of them, it is still considered an English-speaking country.

The miles between the major city in the immigrant's country of origin and the nearest large port of entry into the United States (New York, Miami, Los Angeles) were constructed from data in Fitzpatrick and Modlin's (1986) *Direct Line Distances, United States Edition*, the standard source for such data (see, for example, Yang, 1994; Chiswick and Miller, 2008).

Finally, the information on whether the origin country recognizes dual citizenship was compiled with reference to Renshon (2001). This is a more recent classification than

the United Nations and other data employed by Yang (1994), and is now widely cited in the literature (see Millbank, 2000; Bloemraad, 2004).

Given the above comments, the estimating equation proposed for use in this study is, with expected signs of the partial effects—positive, negative or ambiguous— provided following the variable, is:

Pr(*Citizen*) = *f* (Educational Attainment +, Age -, Years Since Migration +, Years Since Migration Squared -, Veteran +, Speaks Language Other Than English -, South ?, Metropolitan ?, Presence of Children +, Lives Alone -, Lives in Family +, Married-spouse present +, Spouse's Education +, Spouse Native Born +, Lived Abroad 5 Years Ago -, Political Freedom* +, Civil Freedom* +, GDP/Capita* -, Geographic Distance* +, Linguistic Distance* -, Minority Language Concentration* -, Dual Citizenship* +, English Speaking Origin* +, Sojourner Index* -)   (3)

where an asterisk denotes a variable capturing characteristics of the origin country of the particular immigrant. All other variables are for characteristics of the individual immigrant from the Census PUMS 1 percent file.

C. Sample Restrictions

There are several considerations that need to be kept in mind when determining the sample to use in a model of the incidence of citizenship. First, the analysis needs to be restricted to individuals who arrived as adults, as immigrants who arrived as children may obtain citizenship status as part of their parents' naturalization (see Mazzolari, 2007). Second, the duration of residence requirements set out in Table 1 need to be incorporated into the sample selection. With these criteria in mind, the analyses which follow are based on immigrants aged 25 to 64 years at the time of the 2000 US Census, who were at least 18 years of age at the time they arrived in the US, and who have resided in the US for at least 5 years (or 3 years if married to a native-born person or if they have served in the US Armed Forces).

D. Timing of the Variables

Previous studies that have incorporated summary measures of the countries of origin of immigrants have compiled these measures with reference to a single year. Thus Yang (1994), in a study of naturalization decisions based on 1980 US Census data, uses the origin countries' GNP per capita in 1976. Jasso and Rosenzweig (1986) analyze data from the Immigration and Naturalization Service covering the naturalization decisions by 1981 of persons who became permanent resident aliens in Fiscal Year 1971. The origin country GNP per capita variable was for 1970.

There are no data in the Census on the year in which those who became citizens were naturalized. Yet variables like GDP per capita, and the political, civil and economic freedom of the immigrants' countries of origin will vary over time, and this suggests that a superior approach to that of previous studies is, where possible, to link the date of measurement of the variable to the date when the decision to naturalize might have been contemplated. Four approaches are considered below. Under the first two approaches, the variables for the political, civil and economic freedom of the immigrants' countries of origin, together with that for GDP per capita, were measured at specific points in time, paralleling the treatment offered in prior research. The two points in time chosen were 1985 and 1995. For the third approach, the variables were measured five years after the immigrant arrived in the US, a time when most would have first become eligible for naturalization. In the fourth approach, the value of each particular measure (largest value of PR and CL, smallest value of GDP per capita relative to US GDP per capita) that might have prompted a decision to naturalize in the period from five years after arrival in the US to 2000 was used. In cases where information on the particular year that arises from these time-varying calculations is not available, the closest year for which information is available is used. As noted above, the data on GDP per capita are available from 1950, and the data for PR and CL are available from 1973.

The values that are obtained for these alternative measures are highly correlated, as shown in Table 7.

Table 7
Correlation Coefficients between Key Origin-Country Variables
by Type of Measurement

1. Males

	PR (1985)	PR (1995)	PR (Arrival+5)
PR (1995)	0.668	1.000	0.723
PR (Arrival+5)	0.724	0.723	1.000
PR (Most Extreme)	0.687	0.726	0.934
	CL (1985)	CL (1995)	CL (Arrival+5)
CL (1995)	0.732	1.000	0.799
CL (Arrival+5)	0.777	0.799	1.000
CL (Most Extreme)	0.754	0.813	0.932
	GDP (1985)	GDP (1995)	GDP (Arrival+5)
GDP (1995)	0.947	1.000	0.955
GDP (Arrival+5)	0.956	0.955	1.000
GDP (Most Extreme)	0.961	0.976	0.977

2. Females

	PR (1985)	PR (1995)	PR (Arrival+5)
PR (1995)	0.657	1.000	0.733
PR (Arrival+5)	0.731	0.733	1.000
PR (Most Extreme)	0.697	0.744	0.937
	CL (1985)	CL (1995)	CL (Arrival+5)
CL (1995)	0.722	1.000	0.794
CL (Arrival+5)	0.779	0.794	1.000
CL (Most Extreme)	0.757	0.805	0.936
	GDP (1985)	GDP (1995)	GDP (Arrival+5)
GDP (1995)	0.948	1.000	0.956
GDP (Arrival+5)	0.960	0.956	1.000
GDP (Most Extreme)	0.963	0.974	0.983

Note: "Arrival+5" refers to five years after the year of arrival in the US, and "Most Extreme" refers to the value of the variable between "Arrival+5" and 2000 that would be expected to impact on the naturalization decision the most.
Source: Derived data described in Appendix A.

Among males, the correlation coefficient for the Political Rights (PR) variable ranges from 0.67 to 0.93. The correlation coefficient for the Civil Liberties (CL) variable ranges from 0.73 to 0.93, whereas those for the GDP per capita variable are higher, being 0.95 or more. While similar results are therefore expected under the four alternatives, it is still of interest to establish the extent of any differences in findings across the measures, as this can guide future research. Naturally, variables such as the geographic distance between

the country of origin and the US will not differ across time, and variables for whether the origin country can be classified as an English-speaking country will differ by little across time. The timing issue does not arise for these origin-country variables. Data on the acceptance of dual citizenship across time was not compiled, owing to the difficulties in obtaining information.[17] Similarly, there are no comparable data over time in the country of origin sojourner index.

V.   STATISTICAL ANALYSIS

The model of equation (3) is estimated separately for males and females. The significance of differences in the estimated impacts associated with particular variables for males and females is assessed through estimating the model on data pooled across males and females with gender interaction terms. The equation is estimated using both OLS and a probit model. OLS estimates are presented in the text and results from the probit models are presented in Appendix B.[18] The two approaches to estimation yielded similar findings.

In general, the estimates for the individual characteristics were not sensitive to the way that variables for the characteristics of the countries of origin were measured. Hence the discussion of results will proceed in two parts. First, the results for the individual characteristics will be discussed, using the findings obtained when the time-varying

---

[17] Jones-Correa (2001) provides some information in this regard, albeit for a very limited number of countries and a restricted time period.

[18] Results from the OLS and Probit models may be compared using approximate partial effects for the Probit estimates. Thus, let X_{ik} be the kth element of the data vector X_i and β_k be the kth element of the parameter vector β. Then the derivative for the probit model with respect to X_k is given by $\phi(\beta X_i)\beta_k$, where ϕ is the standard normal density function. A useful way of evaluating this is to do so at the value of the probit index βX that solves the equation $\beta X = \Phi^{-1}(\overline{Citizenship})$, where $\overline{Citizenship}$ is the mean citizenship rate for the sample (of 0.46) and Φ^{-1} is the inverse normal cumulative distribution function. Applied to the current study, this gives a value of 0.4. That is, multiplying the probit coefficients by 0.4 will provide an approximate partial effect that can be compared with the OLS coefficient.

variables were constructed using 1985 data.[19] Second, the results for the origin-country variables will be discussed en-masse.

A. Immigrant Characteristics

Table 8 presents the estimates for the individual characteristics. The first column of this table is for males and the second is for females.

The individual characteristics are highly significant as a group: the F-test of their joint contribution yielded a test statistic of 803.65 for males and of 1072.34 for females. Omitting the individual characteristics reduces the adjusted R^2 from 0.250 to 0.080 for males, and from 0.268 to 0.060 for females, indicating that the individual characteristics are more important than the country of origin variables.

The Table 8 results largely conform to the expectations outlined above. Citizenship is more likely among the better educated, and the partial effect is statistically stronger for females than for males. Among males, for example, there would be a 13 percentage point difference in the predicted rate of citizenship among individuals with the highest educational qualifications in the sample (the equivalent of 20 years of schooling) and those who left school following completion of grade 10. For females, the difference in the rates of citizenship for these groups would be three percentage points greater, at 16 percentage points.

There is a positive association between age at migration and the rate of citizenship, and this association does not differ significantly between males and females. As noted earlier, there was no substantial evidence in support of a quadratic age-citizenship relationship. This evidence differs from that reported in prior studies, with Jasso and Rosenzweig (1986) reporting a negative association between citizenship and age at migration, and Yang (1994) reporting that the incidence of citizenship initially rises with age at migration, peaks around age 40, and then declines.

---

[19] The full set of results for all four measurements of the origin-country variables is presented in Appendix B.

Table 8
OLS Estimates of Individual Characteristics in Model
of Citizenship, by Gender, 2000 US Census

Variable	Males	Females[a]
	OLS	OLS
Constant	-0.456	-0.444
	(14.46)	(14.12)
Educational	0.013	0.016*
Attainment	(26.46)	(33.97)
Age at Migration	0.003	0.002
	(9.57)	(8.19)
Years Since	0.043	0.047*
Migration (YSM)	(46.08)	(54.59)
YSM Squared/100	-0.058	-0.059
	(26.56)	(30.16)
Lived Abroad 1995	-0.078	-0.106
	(7.61)	(9.52)
Veteran of US	0.126	0.084
Armed Forces	(13.29)	(3.57)
Speaks non-Eng.	-0.012	0.001
Language at Home	(1.68)	(0.14)
South	0.000	0.001
	(0.03)	(0.16)
Metropolitan	0.023	0.006
	(1.19)	(0.34)
Lives Alone	0.004	0.004
	(0.42)	(0.45)
Non-Family	-0.037	-0.047
Household	(4.08)	(4.10)
Married, Spouse	0.012	0.004
Present	(1.03)	(0.43)
Spouse's Education	0.004	0.003
	(5.83)	(5.75)
Spouse Foreign Born	0.008	-0.038*
	(1.18)	(6.33)
Children aged 0-17	-0.001	0.007
Present	(0.13)	(1.56)
Adjusted R^2	0.250	0.268
Sample Size	53,095	56,608

Note: (a) * = coefficients that are significantly different (at 5% level) for females from that of males.
Source: 2000 US Census, 1 percent PUMS and derived data described in Appendix A.

As mentioned earlier, the age at migration-citizenship relationship has traditionally been discussed from the perspective of the impact that age at migration has on the period over which any benefits associated with citizenship will be received. Yang

(1994, p.472) rationalizes his non-linear relationship as follows: "This evidence suggests that as age at immigration increases, the perceived benefits of naturalization first increase and then decline beginning in the early 40s, and therefore the likelihood of naturalization follows the same trajectory". The Table 8 results, however, point to the positive naturalization-age profile to around age 40 identified by Yang (1994) in the 1980 Census now extending into the older age group. The difference in the pattern of effects in Table 8 and Yang (1994) may be linked to the policy changes which have limited non-citizens' access to social services (see Section I). If this is the case, it points to citizenship offering greater access to economic benefits and social services to those who immigrated at an older age.[20]

The relationship between the probability of being a US citizen and duration of residence in the US is non-linear, rising with duration of residence at a decreasing rate up to around 37 (male) to 40 (female) years in the US, and then declining with duration of residence. As the sample is restricted to immigrants aged 25 to 64 years, who arrived as adults (age 18 and older), the age-naturalization relationship is positive for almost all the immigrants in the sample.[21,22] The citizenship-duration of residence relationships for males and females are presented in Figure 1.

There are three main features of Figure 1. First, rates of citizenship are somewhat greater for females than for males over most durations of residence. In estimations based on data pooled across males and females with a dummy variable for females, the *ceteris paribus* difference in naturalization rates was 4 percentage points (coefficient of 0.037, 't' = 13.89) in favor of females. Second, there is a slightly steeper citizenship-duration of residence relationship for females than for males, but the difference is not economically meaningful. Third, the negative, statistically significant squared duration term means that

---

[20] Comparison of citizenship rates in the cells of a birthplace $\times$ age at migration $\times$ years since migration cross-tabulation across the 1980, 1990 and 2000 censuses may yield insights into this, and is a topic for future research.

[21] Fully 97 percent of the male immigrants in the sample have a duration of residence in the US of 35 or fewer years and 99 percent of the female immigrants in the sample have a duration of residence in the US of 40 or fewer years.

[22] The use of a Gompertz function (specifically exp(-0.1* YSM)) for the years since migration variable yields similar findings to the quadratic specification.

the impact of duration of residence is greatest in the years immediately after immigrants become eligible to naturalize. For example, there is a large difference, of 43 percentage points, in the probability of being a citizen for male immigrants with 5 years residence in the US and those with 20 years residence. In comparison, there is a smaller difference of only 16 percentage points in the probability of being a citizen for male immigrants with 20 years residence in the US and those who have been in the US for 35 years. This is the same pattern of effects as was report by Yang (1994).

Figure 1

Relationship Between Citizenship and Duration of Residence by Gender
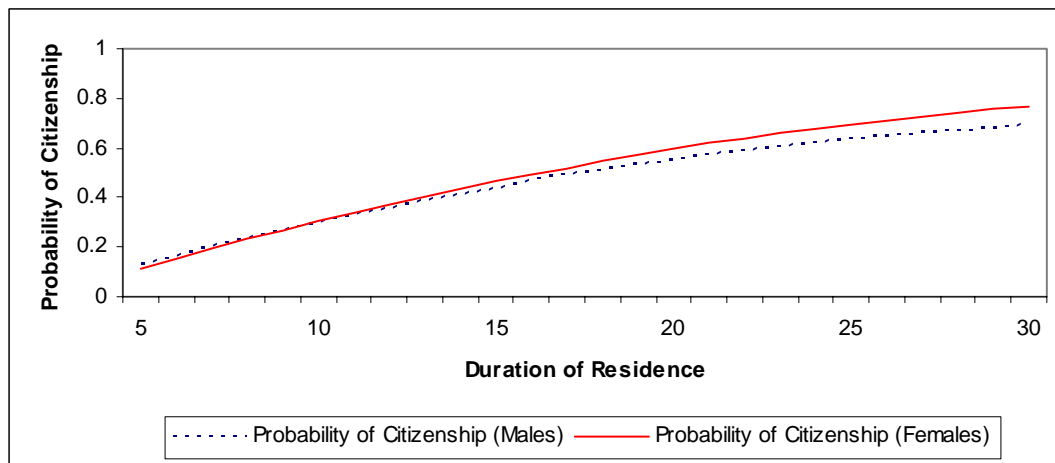
Source: Authors' calculations from Table 8.

Being a veteran of the US Armed Forces is associated with a major change in the probability of US citizenship: 13 percentage points among males and 8 percentage points among females, although these effects do not differ significantly. This is the largest partial effect of the dichotomous variables included in the model. Yang (1994) also reports that a veteran variable has the largest partial effect in his model of citizenship based on 1980 US Census data.

The effect on the citizenship rate of living abroad five years prior to the Census is slightly more pronounced among female immigrants, an 11 percentage points lower rate compared to 8 percentage points for males, although this 3 percentage points excess over the impact for males is not significantly different. This sizeable impact, however, is

relevant only to a small part of the sample (around 2.5 percent of the immigrants lived abroad in 1995) and so it would have little bearing on overall rates of citizenship.

The family structure variables have modest effects on the likelihood of being a US citizen. The omitted category is individuals in family households. Compared to this benchmark group, individuals who live with others in a non-family household (*i.e.*, they live with non-relatives) are four-to-five percentage points less likely to be citizens. This effect is the same for males and females. The greater likelihood of being a citizen associated with membership of a family does not depend on whether it is a couple or sole-parent family (*i.e.*, the "Spouse Present" variable is statistically insignificant).

However, there is evidence that the spouse's characteristics matter to the likelihood of becoming a citizen. Specifically, the citizenship rates increase significantly with the educational attainment of the spouse. The partial effect of the spouses' educational attainment is one-quarter to one-fifth of that of the individual's own educational attainment. Among females, having a foreign-born spouse reduces the probability of citizenship (or equivalently, there is a positive effect on the incidence of citizenship associated with having a native-born spouse). Among males, however, the incidence of citizenship among the married does not vary with the nativity of the spouse. The gender difference in the impact of this variable suggests complex interactions within the household between the spouses.

The remaining individual characteristics variables included in the model (speaks a language other than English at home, location, and the presence of children) are not statistically significant at the 5 percent level. It is noted that Yang (1994) reported that children had a positive impact on the citizenship outcome. While the specification of Table 8 differs from that in Yang (1994) by including variables for the spouse (educational attainment and nativity) and for family structure, tests show that this does not account for the insignificance of the children variable in the current analysis. Note, however, that those with no children aged 0 to 17 at home includes those who never had children and those whose children are older than 17.

B. Country of Origin Variables

Table 9 presents the results for the origin-country variables. As a group, these are highly significant, and almost all variables are individually statistically significant. The F-test of their joint contribution yielded a test statistic of 303.72 for males, and of 348.44 for females. When the country of origin variables were omitted from the model the adjusted R^2 fell from 0.250 to 0.211 for males and from 0.268 to 0.227 for females. Hence, while the origin-country variables enhance the explanatory power of the estimating equation, they are of lesser importance as a group than the individual characteristics presented in Table 8. The goodness of fit of the estimating equation does not differ appreciably across the four measurements for the PR, CL and GDP per capita variables.[23]

Immigrants from countries that recognise dual citizenship are more likely than immigrants from other countries to be US citizens. This is consistent with the evidence in Jones-Correa (2001) and Mazzolari (2007), but contrasts with Yang's (1994) finding that immigrants from countries that allow dual citizenship have lower naturalization rates in the US. The size of the estimated impact, at around two-to-three percentage points, is consistent with Jones-Correa's (2001, p.1023) conclusion that "Dual nationality has relatively small, but positive, effects on immigrants' naturalisation as US citizens".[24]

---

[23] There is one apparently anomalous result, for the PR variable when this is defined with reference to 1995. This suggests that basing the measurement of variables like the political rights in a country on the most recently available information may not be appropriate.

[24] The reason for Yang's (1994) finding is unclear, though it is most likely attributable to Jones-Correa (2001), Mazzolari (2007) and the current study using more recent information on the dual citizenship practices of countries. Yang (1994) examined 1980 census data and used older sources for his dual citizenship variable. Jones-Correa (2001) provides an account of a number of the important changes in citizenship policies for Latin American countries, which shows there have been many changes over the past two decades.

Table 9
OLS Estimates of Origin Country Characteristics in Model
of Citizenship, by Gender, 2000 US Census

Variable	(i)[a]	(ii)	(iii)	(iv)
1. Males				
Origin Country	0.023	0.010	0.022	0.027
Dual Citizenship	(3.78)	(1.62)	(3.54)	(4.48)
Origin English-	0.008	0.005	0.000	0.021
speaking Country	(0.59)	(0.38)	(0.03)	(1.53)
Civil Liberties	0.024	0.017	0.015	0.018
(CL)	(6.35)	(4.85)	(5.43)	(5.34)
Political Rights	0.004	-0.010	0.003	0.008
(PR)	(1.64)	(3.23)	(1.32)	(2.68)
GDP	-0.006	-0.008	-2.075	-1.616
	(9.64)	(14.69)	(14.35)	(9.90)
Geographic	0.022	0.024	0.023	0.022
Distance	(19.57)	(21.24)	(20.65)	(20.11)
Linguistic	-0.044	0.034	-0.012	0.007
Distance	(2.04)	(1.55)	(0.56)	(0.33)
Sojourner Index	-3.387	-3.917	-3.677	-3.666
	(12.43)	(14.05)	(13.28)	(13.22)
Min. Lang	-0.001	-0.002	-0.002	-0.001
Concentration	(3.99)	(4.77)	(4.66)	(4.08)
Adjusted R^2	0.250	0.247	0.250	0.251
2. Females[b]				
Origin Country	0.022	0.010	0.020	0.026
Dual Citizenship	(3.65)	(1.64)	(3.38)	(4.28)
Origin English-	0.094*	0.100*	0.088*	0.107*
speaking Country	(7.92)	(8.45)	(7.62)	(9.08)
Civil Liberties	0.021	0.011	0.012	0.010
(CL)	(5.70)	(3.22)	(4.53)	(3.11)
Political Rights	0.002	-0.010	-0.000	0.006
(PR)	(0.96)	(3.46)	(0.20)	(2.14)
GDP	-0.011*	-0.011*	-3.122*	-2.964*
	(18.32)	(23.25)	(23.19)	(19.32)
Geographic	0.012*	0.016*	0.014*	0.014*
Distance	(12.02)	(15.02)	(14.12)	(13.96)
Linguistic	0.071*	0.168*	0.109*	0.133*
Distance	(3.94)	(8.88)	(6.05)	(7.29)
Sojourner Index	-3.206	-3.795	-3.618	-3.610
	(11.97)	(14.09)	(13.36)	(13.34)
Min. Lang	-0.001	-0.002	-0.002	-0.002
Concentration	(3.69)	(4.46)	(4.32)	(4.21)
Adjusted R^2	0.268	0.263	0.267	0.267

Notes: (a) Model (i) defines the country-level variables with respect to 1985; Model (ii) defines the country-level variables with respect to 1995; Model (iii) defines the country-level variables with respect to five years after the year of arrival; Model (iv) defines the country-level variables with respect to the most extreme values from five years after the year of arrival to 2000.

(b) * = coefficients that are significantly different (at 5% level) for females from that of male.

(c) = When the Civil Liberties (CL) variable was omitted from the model, the coefficient for the Political Rights (PR) variable increased from 0.004 to 0.018 for males and from 0.002 to 0.014 for females in the column (i) model. Similar changes are recorded for models (iii) and (iv). The coefficient on the PR variable in model (ii) became positive for males, and was significant at the 10 percent level. For females the coefficient remained negative but was insignificant.

The variable for immigrants from English-speaking countries is insignificant in the analyses for males. Among females, however, immigrants from English-speaking countries are associated with higher rates of US citizenship, with the partial effect being up to 10 percentage points. This finding contrasts with the result reported by Yang (1994), and suggests that immigrants from countries that are "further" from the US (at least in terms of language) see fewer advantages to naturalization. Note, however, that the English-speaking country variable is used in the construction of both the linguistic distance and minority language concentration variables. Hence, a better comparison of the effects of being from an English-speaking country might involve a contrast with immigrants from non-English-speaking countries having the mean values of these two variables. This alternative comparison shows that both females and males from English-speaking countries have higher predicted rates of citizenship than immigrants from non-English-speaking countries, by 7 percentage points among females and by close to 3 percentage points among males.

The civil liberties (CL) variable is highly significant. Moreover, the estimated coefficient, being between 0.010 and 0.024, indicates that it is quantitatively important. This variable is measured on a scale from 1 to 7. Hence, in Table 9 equation (i), where the variable is defined with reference to 1985, the estimated impact of 0.024 indicates a difference of 14 percentage points in the incidence of citizenship between countries with the least (*e.g.*, Afghanistan, North Korea) and the greatest civil liberty (*e.g.*, Switzerland, Australia).

The political rights (PR) variable has a positive coefficient in the equations other than for that based on the 1995 data, but varies in statistical significance. The minor importance of this variable appears to be associated with the high correlation between the political rights and civil liberties variables that was commented on earlier. When the civil liberties variable is omitted from the model, the political rights variable is significant and positive in all models other than those estimated using the 1995 data (where it is

positive and statistically significant at the 10 percent level for males and insignificant for females), with a coefficient only slightly smaller than that of the civil liberties variables in the models presented in Table 9.

The real GDP per capita variable has a negative coefficient in each set of results, and is highly significant in each instance. The form of the variable differs between the first two sets of results (where the variable is defined with reference to 1985 and 1995) and the final two sets of results (where the variable is defined for varying years, depending on the year the immigrant entered the US). For the first two, the variable is simply the GDP of the origin country, whereas for the second two, it is the GDP of the origin country as a percentage of that in the US of the same year. Hence, comparable coefficients may be obtained by multiplying the GDP coefficient in the first model by 244 (US GDP in hundreds of dollars in 1985) and the GDP coefficient in the second model by 293 (US GDP in hundreds of dollars in 1995). On this basis, it appears that the negative impact of the real GDP per capita of the origin country has been reasonably consistent over time. It is also more pronounced for females than for males. That is, poor economic conditions in the country of origin are more likely to lead to US citizenship among females than they are among males. As with the argument advanced in relation to the gender differences in the positive effect of age at migration on citizenship, citizenship may be seen as offering economic or other benefits in times of distress, and from this perspective the variable should have the greatest impact among groups that are the most vulnerable or disadvantaged. Women are usually viewed as an economic minority, and immigrant women, particularly those from non-English-speaking countries, are particularly disadvantaged (see Le and Miller, 2008).

The physical or geographic distance of the origin country from the US is argued to increase the cost of return migration and of visiting family/friends in the origin country. Consistent with this, the further the country of origin from the US, the more likely it is that an immigrant will become a US citizen. Moreover, distance appears to matter less for females than for males. It is not clear why this difference emerges.

The effect of the linguistic distance variable differs between men and women. Among men it has the expected significant negative effect on the propensity for naturalization when the other country of origin variables are defined for 1985, but it is not

statistically significant when other time periods are used (compare column (i) with columns (ii) through (iv) in the top panel of table 9). Among women, on the other hand, it constantly has a significant positive effect on naturalization, other variables being the same. This is an unexpected result. Tests were performed, including deleting from the sample immigrants from Mexico and from the English-speaking countries, to ascertain why this effect emerges. Yet, the unexpected significant positive effect of linguistic distance on citizenship for females persists.

The sojourner index is a highly significant determinant of the likelihood that immigrants will naturalize. That is, immigrants from countries where a relatively high proportion lived abroad in 1995 are less likely to become US citizens, even if they themselves lived in the US in 1995. The predicted rate of citizenship will vary, for example, by over 10 percentage points when comparing groups where only one percent lived abroad in 1995 (for example, immigrants from Italy), to the high rate of 4 percent lived abroad in 1995 (for example, immigrants from Japan). It would be a 7 percentage points lower naturalization rate for immigrants from Mexico (3 percent lived abroad in 1995) compared to immigrants from Italy.

Finally, the minority language concentration variable is negative and statistically significant in all equations. Immigrants who live in regions where there are proportionately many others with whom they share a mother tongue are less likely to take out US citizenship. Living among others with whom one shares a non-mainstream identify, in this case language, apparently reduces the incentives to naturalize. This is in contrast to Yang (1994) who reported that the larger an ethnic group size nation-wide, the higher the naturalization rate. As noted earlier, however, Yang (1994) did not include a regional dimension to his ethnic group size, and this appears to matter. The current results are consistent with Portes and Curtis (1987), who reported that living in an Anglo neighborhood was associated with a higher incidence of citizenship.

VI.   SUMMARY AND CONCLUSION

This paper has extended research on the determinants of naturalization among immigrants to the United States through an analysis of the 2000 Census of Population. Innovative features include expanding the analysis of country of origin characteristics,

including better measures than have been used previously for political, civil and economic rights and freedom, and GDP per capita in the origin, among other variables. Separate analyses are performed by gender using OLS and probit analyses.

Immigrants are found to be more likely to become citizens if they perceive that the benefits from being naturalized are greater and the costs of satisfying the requirements are lower. Thus, naturalization rates increase with the immigrant's level of schooling, duration in the US, proficiency in English, and service in the US Armed Forces, but are lower among sojourners.

Naturalization rates are also higher the less attractive is the prospect of returning to the country of origin. They are higher the lower are the political rights, civil liberties, economic freedom and GDP per capita in the origin, the further the origin is from the US and the less to and fro migration between the origin and the US. Naturalization is also higher if the origin permits dual citizenship.

Overall, the individual's own characteristics are more important than country of origin characteristics for explaining differences in naturalization rates.

The findings in this paper have implications for how the United States could increase the rate of naturalization among its immigrants without relaxing naturalization requirements. Given that the US allows dual nationality, it is in a position to encourage other countries where this is prohibited (or where certain rights in the origin are lost if an emigrant becomes a citizen of the destination) to do the same. Changes in US immigration policy that encourage the migration of the more highly educated, those more proficient in English, those who bring their spouse and minor children with them and those who are expected to stay a larger period of time (as distinct from sojourners) would enhance the overall naturalization rate. Moreover, citizenship rates could be increased by encouraging immigrants to attend "Americanization" classes where they would learn English and US history, government, civics and culture. Encouraging immigrant enlistment in the US Armed Forces, in part through the deployment of recruiting officers, would further enhance naturalization rates.

Limiting features of the analyses, however, are that there are no data in the Census on the respondent's visa history (i.e., type of visa obtained and when he or she became a permanent resident alien after entering the US) or among those naturalized, the

year in which this occurred. The analysis of the naturalization decision would be greatly enhanced if in the future data sets on immigrants, or the population as a whole, included information on visa status at entry, visa history in the US, and year of naturalization.

RFERENCES

Banks, Arthur S., (1988). *Political Handbook of the World 1988*, CSA Publications, USA.

Bevelander, Pieter and Veenman, Justus, (2006). "Naturalization and Socioeconomic Integration: The Case of the Netherlands", Institute for the Study of Labour Discussion Paper No. 2153, Bonn.

Bloemraad, Irene, (2002). "The North American Naturalization Gap: An Institutional Approach to Citizenship Acquisition in the United States and Canada", *International Migration Review*, Vol. 36, No. 1, pp.193-228.

Bloemraad, Irene, (2004). "Who Claims Dual Citizenship? The Limits of Postnationalism, the Possibilities of Transnationalism, and the Persistence of Traditional Citizenship", *International Migration Review*, Vol. 38, No. 2, pp.389-426.

Chiswick, Barry R. and Miller, Paul W., (2005a). "Do Enclaves Matter in Immigrant Adjustment?", *City and Community*, Vol. 4, No. 1, pp.5-35.

Chiswick, Barry R. and Miller, Paul W., (2005b). "Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages", *Journal of Multilingual and Multicultural Development*, Vol. 26, No. 1, pp.1-11.

Chiswick, Barry R., and Miller, Paul W., (2006). "Language Skills and Immigrant Adjustment: The Role of Immigration Policy", in D. Cobb-Clark and S. Khoo (eds.) *Public Policy and Immigrant Settlement*, Edward Elgar Publishing, Cheltenham, United Kingdom.

Chiswick, Barry R. and Miller, Paul W., (2008) "Modelling Immigrants' Language Skills", *Research in Labor Economics*, Vol. 27, pp. 75-128.

Constant, Amelie F., Gataulina, Liliya and Zimmermann, Klaus F., (2007). "Naturalization Proclivities, Ethnicity and Integration", Institute for the Study of Labour Discussion Paper No. 3260, Bonn.

DeSipio, Louis, (1987). "Social Science Literature and the Naturalization Process", *International Migration Review*, Vol. 21, No. 2, pp.390-405.

DeVoretz, Don J. and Pivnenko, Sergiy, (2005). "Self-Selection, Immigrant Public Finance Performance and Canadian Citizenship", Institute for the Study of Labour Discussion Paper No. 1463, Bonn.

Evans, Mariah, (1988). "Choosing to be a Citizen: The Time-Path of Citizenship in Australia", *International Migration Review*, Vol. 22, No. 2, pp. 243-264.

Fitzpatrick, Gary L. and Modlin, Marilyn J., (1986). *Direct-line Distances: United States Edition*, Metuchen, NJ: The Scarecrow Press Inc.

Fougère, Denis and Safi, Mirna, (2008). "The Effects of Naturalization on Immigrants' Employment Probability (France, 1968-1999)", Institute for the Study of Labour Discussion Paper No. 3372, Bonn.

Gwartney, James and Lawson, Robert, (2007). *Economic Freedom of the World: 2007 Annual Report*, co-published by The Fraser Institute in Canada and seventy-one other research institutes in various countries.

Hart-Gonzalez, Lucinda and Lindermann, Stephanie, (1993). "Expected Achievement in Speaking Proficiency, 1993" School of Language Studies, Foreign Service Institute, U.S. Department of State, Mimeographed.

Heston, Alan, Summers, Robert and Aten, Bettina, (2006). *Penn World Table Version 6.2*, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Jasso, Guillermina and Rosenzweig, Mark R., (1986). "Family Reunification and the Immigration Multiplier: U.S. Immigration Law, Origin-Country Conditions, and the Reproduction of Immigrants", *Demography*, Vol. 23, No. 3, pp.291-311.

Jones-Correa, Michael, (2001). "Under Two Flags: Dual Nationality in Latin America and its Consequences for Naturalization in the United States", *International Migration Review*, Vol. 35, No. 4, pp. 997-1029.

Le, Anh T. and Miller, Paul W., (2008). "Glass Ceiling and Double Disadvantage Effects: Women in the US Labor Market", *Applied Economics*, (forthcoming).

Mazzolari, Francesca, (2007). "Dual Citizenship Rights: Do They Make More and Better Citizens", IZA Discussion Paper No. 3008, Bonn, Germany.

Millbank, Adrienne, (2000). "Dual Citizenship in Australia", Current Issue Brief 5, 2000-01, Parliamentary Library, Parliament of Australia, Canberra.

Pachon, Harry P., (1987). "An Overview of Citizenship in the Hispanic Community", *International Migration Review,* Vol. 21, No. 2, pp.299-310.

Portes, Alejandro and Curtis, John W., (1987). "Changing Flags" Naturalization and its Determinants Among Mexican Immigrants", *International Migration Review,* Vol. 21, No. 2, pp.352-371.

Renshon, Stanley A., (2001). *Dual Citizenship and American National Identity*, Center for Immigration Studies, Washington, USA.

US Census Bureau, (2005). *Public Use Microdata Sample: 2000 Census of Population and Housing Technical Documentation*, US Census Bureau, Washington.

Yang, Philip Q., (1994). "Explaining Immigrant Naturalization", *International Migration Review*, Vol. 28, No. 2, pp.449-477.

1. Individual Characteristics

Data Source: 2000 Census of Population of the United States, Public Use Microdata Sample (PUMS), 1 percent sample of the foreign born, except where noted otherwise (see US Census Bureau, 2005).

Definition of Population: Foreign-born men and women aged twenty-five to sixty-four, who migrated to the US aged 18 or more years, and who have met the residency requirements for US citizenship set out in Table 1. Only residents of the 50 States and the District of Columbia are considered.

Citizenship: This is a dichotomous variable, set to one for the foreign born who are citizenships of the US, and set to zero for those who are not US citizens at the time of the 2000 US Census.

Years of Education (EDUC): This variable records the total years of full-time equivalent education. It has been constructed from the Census data on educational attainment by assigning the following values to the Census categories: completed less than fifth grade (2 years); completed fifth or sixth grade (5.5); completed seventh or eighth grade (7.5); completed ninth grade (9); completed tenth grade (10); completed 11th grade (11); completed 12th grade or high school (12); attended college for less than one year (12.5); attended college for more than one year or completed college (14); Bachelor's degree (16); Master's degree (17.5); Professional degree (18.5); Doctorate (20).

Age: This is the individual's age in years. As a years since migration variable is included in all estimating equations, the age variable can be interpreted as age at migration.

Years Since Migration (YSM). This is computed from the year the foreign-born person came to the United States to stay.

Abroad: This is a binary variable, set to equal one for individuals who resided outside the United States 5 years ago, but who first came to the US to stay more than five years ago.

Location: The two location variables record residence in a non-metropolitan area (NON-MET) or in the Southern States (SOUTH). The states included in the latter are: Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia.

Family Structure: Three types of households are distinguished in the analyses: (i) Family households (involving either couple or lone-parent families); (ii) Sole-person

households; and (iii) other non-family multiple person households (*e.g.*, individuals living with non-relatives).

Marital Status (MARRIED): This is a binary variable that distinguishes individuals who are married, spouse present (equal to 1) from all other marital states. It thus also distinguishes the couple and lone-parent types of family households.

Spouse's Details: Individuals who are married, spouse present, have details on their spouse linked to their individual characteristics using the hierarchical nature of the PUMS. The details that are linked to partners are: (i) spouse's educational attainment; (ii) spouse's nativity (native born or foreign born); and, for males, (iii) the information on the presence of children.

Mother Tongue: This is a binary variable, set equal to one for individuals who speak a language other than English at home (assumed to be their mother tongue). It is set equal to zero for individuals who speak only English at home.

Veteran: This is a binary variable set equal to one for someone who had served in the US Armed Forces, and set equal to zero otherwise.

2. Country of Origin Characteristics

Political Rights and Civil Liberties: Data for Political Rights (PR) and Civil Liberties (CL) were obtained from the publication *Freedom in the World* on the Freedom House website. This publication is the standard-setting comparative assessment of global political rights and civil liberties. Published annually since 1972, the survey includes both analytical reports and numerical ratings for 193 countries and 15 select territories. Each country and territory is assigned a numerical rating—on a scale of 1 to 7—for political rights and an analogous rating for civil liberties; a rating of 1 indicates the highest degree of freedom and 7 the least amount of freedom. These ratings are combined by the compilers to generate a third measure, of whether a country can be classified as Free, Partly Free, or Not Free. Countries whose combined average ratings for Political Rights and for Civil Liberties fell between 1.0 and 2.5 were designated "Free"; between 3.0 and 5.5 "Partly Free", and between 5.5 and 7.0 "Not Free".

Source: http://www.freedomhouse.org

Economic Freedom: Information on Economic Freedom (EF) is from *The Economic Freedom of the World: 2007 Annual Report*, by James Gwartney and Robert Lawson and co-published by The Fraser Institute and the seventy-one other research institutes of the Economic Freedom Network. *The Economic Freedom of the World* ratings are based on 42 government polices that impact on economic freedom. These are organized around five broad areas: (i) Size of Government; (ii) Legal Structure and Security of Property Rights; (iii) Access to Sound Money; (iv) Freedom to Trade Internationally; (v)

Regulation of Credit, Labor and Business. The ratings are based on a 0 to10 scale, with 10 being most free and 0 the least.

Source: http://www.freetheworld.com

Gross Domestic Product (GDP): Data on real GDP per capita at constant prices were obtained from the Penn World Table. The Penn World Table contains national accounts economic time series information covering 188 countries. Its expenditure entries are denominated in a common set of prices in a common currency so that real quantity comparisons can be made, both between countries and over time. It is also noted that there are several countries where information is not available from the Penn World Table. In these cases, World Bank estimates of GDP per capital, current international dollars (PPPs) were used where possible, and these were obtained from the United Nations Statistics Division.

Version 6.2 of the Penn World Tables was used. See Alan Heston, Robert Summers and Bettina Aten, Penn World Table, Version 6.2, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, September 2006.

Source: http://pwt.econ.upenn.edu/php_site/pwt_index.php
        http://unstats.un.org/unsd/cdb/cdb_help/cdb_quick_start.asp

English-Speaking Countries: Data for the English-speaking countries are obtained from the official languages for different countries listed in the *Political Handbook of the World: 1988*. This information is coded into a dichotomous variable. If a country happens to have a few official languages and English is only one of the many, it is still considered as an English-speaking country.

Source: Arthur S. Banks (1988), *Political Handbook of the World 1988*, CSA Publications, USA.

Dual Citizenship: There are a number of lists of countries that recognize dual citizenship. The primary source of the information in the current study is Renshon (2001). This was cross-checked against other lists (*e.g.*, at CUNY) and discrepancies reconciled as far as possible through access to web sites of the official agencies in the relevant countries.

Source: Stanley A. Renshon., (2001), *Dual Citizenship and American National Identity*, Center for Immigration Studies, Washington, USA.

Sojourner Index: This variable records the means of the variable *Abroad* by birthplace groups. It is created from the 2000 US Census 5 percent PUMS and is based on individuals aged 20 to 64 who had lived in the US for more than five years.

Linguistic Distance: This is a measure of the difficulty of learning a foreign language for English-speaking Americans. It is based on a set of Language Scores (LS) measuring achievements in speaking proficiency in foreign languages by English-speaking Americans at the US Department of State, School of Language Studies, reported by Hart-Gonzalez and Lindermann (1993). It is described in detail in Chiswick and Miller (2005b). The Language Scores range from a low of 1.0 (*e.g.*, Korean) to a high of 3.0 (*e.g.* Swedish). The Linguistic Distance measure is LD = 1/LS, that is, LD equals 0.33 for Swedish and 1.0 for Korean.

In the construction of this variable, for the 3 percent of the sample who report a language for which there is no linguistic score, the sample mean value is assigned as the language score. Foreign-born persons from predominately English-speaking countries are assigned a value of zero.

Source: Lucinda Hart-Gonzalez and Stephanie Lindermann (1993). "Expected Achievement in Speaking Proficiency, 1993" School of Language Studies, Foreign Service Institute, U.S. Department of State, Mimeographed.

Direct-Line Distance: The miles between the major city in the immigrant's country of origin and the nearest large port of entry in the United States (New York, Miami, Los Angeles) were compiled from data in Fitzpatrick and Modlin's (1986) *Direct Line Distances, United States Edition*.

Source: Gary L. Fitzpatrick. and Marilyn J. Modlin (1986), *Direct-line Distances: United States Edition*, Metuchen, NJ: The Scarecrow Press Inc.

Minority Language Concentration: Each respondent is assigned a measure equal to the percentage of the population aged eighteen to sixty-four in the region in which he/she lives, who reports the same non-English language as the respondent. This measure is based on the state (50 States and the District of Columbia) of residence—a focus that corresponds to that used by Chiswick and Miller (2005a). In the construction of this variable, only the twenty-five largest non-American Indian language groups nationwide are considered. These constitute 92 percent of all responses (native and foreign born) where a language other than English is used at home. Representation in the other language groups is so small numerically that the proportions are approximately zero, and this value is assigned. Foreign-born persons from predominately English-speaking countries are assigned a value of zero. Hence the variable used in the estimating equation is to be interpreted as the interaction between the English-speaking country variable and the Minority Language Concentration variable.

Source: Barry R. Chiswick.and Paul W. Miller, (2008), "Modelling Immigrants' Language Skills", *Research in Labor Economics*, Vol. 27, pp. 75-128.

Means and standard deviations for these variables are reported in Table A.1, with the first column being for men and the second column for women.

Means and Standard Deviations of Variables by Gender, 2000 US Census

Variable	Males	*Females*
Incidence of Citizenship	0.441	0.476
	(0.50)	(0.50)
Educational Attainment	11.531	11.302
	(5.05)	(4.78)
Age at Migration	43.780	44.560
	(9.94)	(10.16)
Years Since Migration	16.744	17.156
(YSM)	(8.41)	(9.07)
Lived Abroad 1995	0.028	0.023
	(0.16)	(0.15)
Veteran of US Armed Forces	0.038	0.005
	(0.19)	(0.07)
Speaks non-Eng. Language	0.866	0.842
at Home	(0.34)	(0.36)
South	0.266	0.264
	(0.44)	(0.44)
Metropolitan	0.991	0.992
	(0.10)	(0.09)
Lives Alone	0.063	0.056
	(0.24)	(0.23)
Non-Family	0.049	0.029
Household	(0.22)	(0.17)
Married, Spouse Present	0.677	0.662
	(0.47)	(0.47)
Spouse's Education	7.867	8.026
	(6.66)	(6.99)
Spouse Foreign Born	0.576	0.521
	(0.49)	(0.50)
Children aged 0-17 Present	0.454	0.492
	(0.50)	(0.50)
Origin Country Dual	0.773	0.756
Citizenship	(0.42)	(0.43)
Origin English-speaking	0.209	0.237
Country	(0.41)	(0.43)
Civil Liberties	4.065	3.926
(CL)	(1.59)	(1.65)
Political Rights	3.859	3.717
(PR)	(1.86)	(1.91)
GDP	6.245	6.604
	(4.59)	(4.99)
Geographic Distance	3.571	3.680
	(2.69)	(2.70)
Linguistic Distance	0.400	0.397
	(0.24)	(0.26)
Sojourner Index	0.026	0.025
	(0.01)	(0.01)

Min. Lang Concentration	2.368	2.116
	(6.09)	(5.78)
Sample Size	53,095	56,608

Source: 2000 US Census, 1 percent PUMS.

DETAILED REGRESSION RESULTS

Table B.1

OLS Estimates of Model of Citizenship, Males, United States, 2000

Variable	(i)[a]	(ii)	(iii)	(iv)
Constant	-0.456	-0.374	-0.402	-0.433
	(14.46)	(12.48)	(13.54)	(14.54)
Educational	0.013	0.013	0.013	0.013
Attainment	(26.46)	(26.71)	(26.75)	(26.05)
Age at Migration	0.003	0.003	0.003	0.003
	(9.57)	(10.02)	(10.30)	(9.98)
Years Since	0.043	0.043	0.043	0.040
Migration (YSM)	(46.08)	(46.12)	(46.01)	(41.97)
YSM Squared	-0.058	-0.059	-0.060	-0.055
	(26.56)	(27.13)	(27.42)	(25.08)
Lived Abroad 1995	-0.078	-0.080	-0.080	-0.080
	(7.61)	(7.78)	(7.79)	(7.78)
Veteran of US	0.126	0.123	0.122	0.124
Armed Forces	(13.29)	(13.00)	(12.95)	(13.13)
Speaks non-Eng.	-0.012	-0.020	-0.022	-0.020
Language at Home	(1.68)	(2.87)	(3.23)	(2.96)
South	0.000	0.002	0.000	-0.003
	(0.03)	(0.48)	(0.09)	(0.59)
Metropolitan	0.023	0.023	0.021	0.022
	(1.19)	(1.16)	(1.10)	(1.13)
Lives Alone	0.004	0.003	0.007	0.005
	(0.42)	(0.35)	(0.76)	(0.57)
Non-Family	-0.037	-0.038	-0.035	-0.035
Household	(4.08)	(4.15)	(3.85)	(3.87)
Married, Spouse	0.012	0.006	0.012	0.011
Present	(1.03)	(0.54)	(1.03)	(0.98)
Spouse's Education	0.004	0.004	0.004	0.003
	(5.83)	(5.73)	(5.75)	(5.65)
Spouse Foreign	0.008	0.015	0.009	0.011
Born	(1.18)	(2.10)	(1.27)	(1.58)
Children Aged 0-	-0.001	0.001	0.001	0.001
17 Present	(0.13)	(0.13)	(0.21)	(0.13)
Origin Country	0.023	0.010	0.022	0.027
Dual Citizenship	(3.78)	(1.62)	(3.54)	(4.48)
Origin English-	0.008	0.005	0.000	0.021
speaking Country	(0.59)	(0.38)	(0.03)	(1.53)
Civil Liberties	0.024	0.017	0.015	0.018
	(6.35)	(4.85)	(5.43)	(5.34)
Political Rights	0.004	-0.010	0.003	0.008
	(1.64)	(3.23)	(1.32)	(2.68)
GDP	-0.006	-0.008	-2.075	-1.616

	(9.64)	(14.69)	(14.35)	(9.90)
Geographic	0.022	0.024	0.023	0.022
Distance	(19.57)	(21.24)	(20.65)	(20.11)
Linguistic Distance	-0.044	0.034	-0.012	0.007
	(2.04)	(1.55)	(0.56)	(0.33)
Sojourner Index	-3.387	-3.917	-3.677	-3.666
	(12.43)	(14.05)	(13.28)	(13.22)
Min. Lang	-0.001	-0.002	-0.002	-0.001
Concentration	(3.99)	(4.77)	(4.66)	(4.08)
Adjusted R^2	0.250	0.247	0.250	0.251
Sample Size	53,095	53,095	53,095	53,095

Note: (a) Model (i) defines the country-level variables with respect to 1985; Model (ii) defines the country-level variables with respect to 1995; Model (iii) defines the country-level variables with respect to five years after the year of arrival; Model (iv) defines the country-level variables with respect to the most extreme values from five years after the year of arrival to 2000.

Source: See Appendix A.

Table B.2

OLS Estimates of Model of Citizenship, Females, United States, 2000[a]

Variable	(i)[b]	(ii)	(iii)	(iv)
Constant	-0.444	-0.377	-0.397	-0.413
	(14.12)	(12.63)	(13.66)	(14.11)
Educational	0.016*	0.016*	0.016*	0.016*
Attainment	(33.97)	(33.42)	(34.16)	(33.88)
Age at Migration	0.002	0.002	0.002	0.002
	(8.19)	(8.35)	(8.65)	(8.46)
Years Since	0.047*	0.047*	0.047*	0.044*
Migration (YSM)	(54.59)	(54.94)	(54.80)	(50.16)
YSM Squared	-0.059	-0.061	-0.061	-0.056
	(30.16)	(31.00)	(31.27)	(28.51)
Lived Abroad	-0.106	-0.107	-0.107	-0.107
1995	(9.52)	(9.54)	(9.57)	(9.61)
Veteran of US	0.084	0.081	0.082	0.081
Armed Forces	(3.57)	(3.40)	(3.43)	(3.39)
Speaks non-Eng.	0.001	-0.006	-0.009	-0.009
Language at Home	(0.14)	(0.96)	(1.38)	(1.40)
South	0.001	0.003	0.001	-0.000
	(0.16)	(0.56)	(0.12)	(0.09)
Metropolitan	0.006	0.009	0.008	0.008
	(0.34)	(0.48)	(0.40)	(0.40)
Lives Alone	0.004	0.001	0.005	0.005
	(0.45)	(0.17)	(0.56)	(0.59)
Non-Family	-0.047	-0.052	-0.047	-0.047
Household	(4.10)	(4.53)	(4.18)	(4.12)
Married, Spouse	0.004	-0.011	-0.002	0.001
Present	(0.43)	(1.16)	(0.20)	(0.06)
Spouse's	0.003	0.003	0.003	0.003
Education	(5.75)	(6.24)	(5.96)	(5.84)
Spouse Foreign	-0.038*	-0.026*	-0.033*	-0.035*
Born	(6.33)	(4.30)	(5.54)	(5.81)
Children aged 0-17	0.007	0.008	0.008	0.008
Present	(1.56)	(1.85)	(1.77)	(1.74)
Origin Country	0.022	0.010	0.020	0.026
Dual Citizenship	(3.65)	(1.64)	(3.38)	(4.28)
Origin English-	0.094*	0.100*	0.088*	0.107*
speaking Country	(7.92)	(8.45)	(7.62)	(9.08)
Civil Liberties	0.021	0.011	0.012	0.010
	(5.70)	(3.22)	(4.53)	(3.11)
Political Rights	0.002	-0.010	-0.000	0.006
	(0.96)	(3.46)	(0.20)	(2.14)
GDP	-0.011*	-0.011*	-3.122*	-2.964*
	(18.32)	(23.25)	(23.19)	(19.32)
Geographic	0.012*	0.016*	0.014*	0.014*
Distance	(12.02)	(15.02)	(14.12)	(13.96)
Linguistic	0.071*	0.168*	0.109*	0.133*

Distance	(3.94)	(8.88)	(6.05)	(7.29)
Sojourner Index	-3.206	-3.795	-3.618	-3.610
	(11.97)	(14.09)	(13.36)	(13.34)
Min. Lang	-0.001	-0.002	-0.002	-0.002
Concentration	(3.69)	(4.46)	(4.32)	(4.21)
Adjusted R^2	0.268	0.263	0.267	0.267
Sample Size	56,608	56,608	56,608	56,608

Notes: (a) * = coefficients that are significantly different (at 5% level) for females from that of male.

(b) Model (i) defines the country-level variables with respect to 1985; Model (ii) defines the country-level variables with respect to 1995; Model (iii) defines the country-level variables with respect to five years after the year of arrival; Model (iv) defines the country-level variables with respect to the most extreme values from five years after the year of arrival to 2000.

Source: See Appendix A.

Table B.3

Probit Estimates of Model of Citizenship, Males, United States, 2000

Variable	(i)[a]	(ii)	(iii)	(iv)
Constant	-3.019	-2.737	-2.842	-2.895
	(28.94)	(27.79)	(28.97)	(29.46)
Educational	0.042	0.042	0.042	0.041
Attainment	(26.10)	(26.28)	(26.28)	(25.68)
Age at Migration	0.008	0.009	0.009	0.009
	(9.66)	(10.08)	(10.38)	(10.07)
Years Since	0.137	0.136	0.136	0.127
Migration (YSM)	(42.08)	(42.09)	(41.64)	(38.69)
YSM Squared	-0.193	-0.196	-0.194	-0.182
	(25.18)	(25.59)	(25.18)	(23.65)
Lived Abroad	-0.323	-0.328	-0.327	-0.326
1995	(7.89)	(8.03)	(8.00)	(7.97)
Veteran of US	0.423	0.414	0.416	0.419
Armed Forces	(12.78)	(12.50)	(12.54)	(12.65)
Speaks non-Eng.	-0.029	-0.054	-0.060	-0.056
Language at Home	(1.33)	(2.49)	(2.76)	(2.58)
South	-0.003	0.003	-0.003	-0.011
	(0.20)	(0.22)	(0.18)	(0.70)
Metropolitan	0.082	0.081	0.076	0.078
	(1.26)	(1.25)	(1.17)	(1.19)
Lives Alone	0.018	0.017	0.025	0.021
	(0.64)	(0.60)	(0.90)	(0.75)
Non-Family	-0.121	-0.122	-0.115	-0.116
Household	(3.83)	(3.87)	(3.65)	(3.66)
Married, Spouse	0.060	0.043	0.055	0.055
Present	(1.65)	(1.18)	(1.51)	(1.51)
Spouse's	0.010	0.009	0.010	0.009
Education	(4.84)	(4.74)	(4.84)	(4.76)
Spouse Foreign	0.023	0.043	0.029	0.034
Born	(1.05)	(1.98)	(1.32)	(1.55)
Children Aged 0-	0.002	0.006	0.007	0.005
17 Present	(0.10)	(0.36)	(0.39)	(0.29)
Origin Country	0.073	0.029	0.067	0.086
Dual Citizenship	(3.75)	(1.50)	(3.46)	(4.42)
Origin English-	0.028	0.021	0.007	0.068
speaking Country	(0.69)	(0.51)	(0.17)	(1.65)
Civil Liberties	0.076	0.055	0.049	0.037
	(6.35)	(4.86)	(5.48)	(3.49)
Political Rights	0.012	-0.032	0.008	0.035
	(1.40)	(3.46)	(1.07)	(3.82)
GDP	-0.019	-0.023	-6.290	-5.262
	(9.60)	(15.37)	(14.62)	(10.42)
Geographic	0.065	0.071	0.067	0.066
Distance	(19.28)	(20.86)	(20.14)	(19.80)
Linguistic	-0.140	0.106	-0.037	0.033

Distance	(2.12)	(1.58)	(0.56)	(0.50)
Sojourner Index	-10.735	-12.486	-11.632	-11.895
	(14.31)	(16.94)	(15.61)	(15.82)
Min. Lang	-0.004	-0.005	-0.005	-0.004
Concentration	(3.52)	(4.28)	(4.10)	(3.62)
Chi-Squared	14899.43	14691.70	14876.06	14922.96
Pseudo R^2	0.205	0.202	0.205	0.026
Sample Size	53,095	53,095	53,095	53,095

Note: (a) Model (i) defines the country-level variables with respect to 1985; Model (ii) defines the country-level variables with respect to 1995; Model (iii) defines the country-level variables with respect to five years after the year of arrival; Model (iv) defines the country-level variables with respect to the most extreme values from five years after the year of arrival to 2000.

Source: See Appendix A.

Table B.4

Probit Estimates of Model of Citizenship, Females, United States, 2000[a]

Variable	(i)[b]	(ii)	(iii)	(iv)
Constant	-2.943	-2.728	-2.822	-2.825
	(28.07)	(27.41)	(28.70)	(28.67)
Educational	0.053*	0.052*	0.053*	0.053*
Attainment	(33.32)	(32.75)	(33.42)	(33.23)
Age at Migration	0.007	0.007	0.008	0.008
	(8.57)	(8.66)	(8.95)	(8.81)
Years Since	0.144	0.145	0.144*	0.136*
Migration (YSM)	(48.57)	(48.98)	(48.52)	(45.05)
YSM Squared	-0.186	-0.190	-0.188	-0.176
	(27.26)	(28.07)	(27.57)	(25.65)
Lived Abroad	-0.407	-0.405	-0.405	-0.408
1995	(9.43)	(9.42)	(9.39)	(9.46)
Veteran of US	0.300	0.287	0.292	0.287
Armed Forces	(3.31)	(3.18)	(3.23)	(3.17)
Speaks non-Eng.	0.011	-0.008	-0.014	-0.017
Language at Home	(0.54)	(0.39)	(0.71)	(0.87)
South	0.003	0.008	0.001	0.000
	(0.21)	(0.55)	(0.09)	(0.02)
Metropolitan	0.032	0.041	0.035	0.035
	(0.48)	(0.61)	(0.53)	(0.52)
Lives Alone	0.011	0.003	0.012	0.014
	(0.39)	(0.10)	(0.41)	(0.51)
Non-Family	-0.149	-0.163	-0.152	-0.149
Household	(4.04)	(4.42)	(4.13)	(4.04)
Married, Spouse	0.007	-0.042	-0.014	-0.005
Present	(0.22)	(1.34)	(0.46)	(0.16)
Spouse's	0.010	0.011	0.011	0.010
Education	(5.48)	(5.97)	(5.74)	(5.62)
Spouse Foreign	-0.117*	-0.078*	-0.100*	0.106*
Born	(6.11)	(4.11)	(5.26)	(5.57)
Children aged 0-	0.025	0.029	0.027	0.027
17 Present	(1.71)	(1.98)	(1.85)	(1.84)
Origin Country	0.069	0.034	0.069	0.087
Dual Citizenship	(3.66)	(1.81)	(3.64)	(4.57)
Origin English-	0.291*	0.320*	0.292*	0.346*
speaking Country	(7.95)	(8.87)	(8.12)	(9.46)
Civil Liberties	0.063	0.034	0.043	0.012
	(5.57)	(3.09)	(4.93)	(1.18)
Political Rights	0.006	-0.032	-0.005	0.031
	(0.82)	(3.48)	(0.72)	(3.48)
GDP	-0.036*	-0.035*	-9.854*	-9.784*
	(19.56)	(24.90)	(23.99)	(20.68)
Geographic	0.037*	0.049*	0.042*	0.044*
Distance	(11.70)	(14.80)	(13.48)	(13.75)
Linguistic	0.221*	0.537*	0.367*	0.448*

	(i)	(ii)	(iii)	(iv)
Distance	(3.87)	(9.17)	(6.37)	(7.73)
Sojourner Index	-10.843	-12.889	-12.161	-12.504
	(13.83)	(17.04)	(15.77)	(16.13)
Min. Lang	-0.004	-0.005	-0.005	-0.005
Concentration	(3.45)	(4.12)	(3.85)	(3.88)
Chi-Squared	17149.19	16845.82	17099.46	17127.04
Pseudo R^2	0.220	0.216	0.220	0.220
Sample Size	56,608	56,608	56,608	56,608

Notes: (a) * = coefficients that are significantly different (at 5% level) for females from that of male.

(b) Model (i) defines the country-level variables with respect to 1985; Model (ii) defines the country-level variables with respect to 1995; Model (iii) defines the country-level variables with respect to five years after the year of arrival; Model (iv) defines the country-level variables with respect to the most extreme values from five years after the year of arrival to 2000.

Source: See Appendix A.

APPENDIX C

TREATMENT OF MISSING VALUES IN CONSTRUCTION OF ORIGIN COUNTRY VARIABLES

The following notes outline the way that missing values have been assigned when constructing the origin country variables.

<u>Political Rights</u>:

1) Data for Czechoslovakia from 1993-2000 are assigned the corresponding values for Czech Republic.
2) Data for Azores Island are assigned the corresponding values for Portugal.
3) Data for England, Scotland, Wales, Northern Ireland, Guernsey, Jersey and Isle of Man are assigned the corresponding values for United Kingdom.
4) Data for Czech Republic from 1973-1992 are assigned the corresponding values for Czechoslovakia.
5) Data for Slovakia (1973-1992), Bosnia & Herzegovina (1973-1991), Croatia (1973-1990), Macedonia (1973-1991), Slovenia (1973-1990), Serbia (1973-2000) and Kosovo (1973-2000) are assigned the corresponding values for Yugoslavia.
6) Data for Estonia, Latvia, Lithuania, Armenia, Azerbaijan, Belarus, Georgia, Moldova, Russia, Ukraine, Kazakhstan, Tajikistan, Turkmenistan and Uzbekistan from 1973-1990 are assigned the corresponding values for USSR. Data for Kyrgyzstan are assigned the corresponding values for USSR from 1973-1991 and for Russia from 1992-2000.
7) Data for USSR from 1992-2000 are assigned the corresponding values for Russia.
8) Data for Hong Kong are assigned the value of 1 based on its economic freedom.
9) Data for Korea are assigned the corresponding values for South Korea.
10) Data for Macau are assigned a middle ranking of 3 throughout the data period. It was a colony of Portugal until 1999 when it became a Special Administrative Region of China.
11) Data for Bermuda are assigned the top value, of 1, throughout the data period, as it was a British colony.
12) Data for Belize from 1973-1981 are assigned the top value, of 1 as it was a British colony.
13) Data for Antigua & Barbuda from 1973-1981 are assigned the top value, of 1 as it was a British colony.
14) Bahamas was a British colony until 1973, hence a value of 1 for 1973.
15) Data for Dominica from 1973-1978 are assigned the top value, of 1 as it was a British colony.
16) Data for Grenada from 1973-1974 are assigned the top value, of 1 as it was a British colony until full independence in 1974.
17) Netherlands Antilles remains an autonomous country within the Netherlands, and a value of 1 was assigned.
18) Data for St Kitts-Nevis from 1973-1981 are assigned the top value, of 1 as it was a British colony.
19) Data for St Lucia from 1973-1979 are assigned the top value, of 1 as it was a British colony.
20) Data for St Vincent & the Grenadines from 1973-1979 are assigned the top value, of 1 as it was a British colony.
21) Suriname was a Netherlands colony until 1975. The first post-independence value available was used for the pre-independence years 1973-1975.

22) Angola gained independence from Portugal in 1975, but the transition to independence was difficult. The first post-independence value available was used for the pre-independence years 1973-1975.
23) Cape Verde was a colony of Portugal until 1975. The first post-independence value available was used for the pre-independence years 1973-1975.
24) Comoros was a colony of France until 1975. The first post-independence value available was used for the pre-independence years 1973-1975.
25) Djibouti was a colony of France until 1977. The first post-independence value available was used for the pre-independence years 1973-1977.
26) Data for Eritrea from 1973-1992 are assigned the corresponding values for Ethiopia.
27) Mozambique was a colony of Portugal until 1975. The first post-independence value available was used for the pre-independence years 1973-1975.
28) Namibia was occupied by South Africa until 1988. The values for 1976-1988 were imputed using the values for 1975.
29) Sao Tome & Principe gained independence from Portugal in 1975. The first post-independence value available was used for the pre-independence years 1973-1975.
30) Seychelles gained independence from UK in 1976. The first post-independence value available was used for the pre-independence years 1973-1976.
31) Datum for South Africa for 1973 is assigned based on the value of 1974, as there is an absence of significant events between 1973 and 1974 to suggest otherwise.
32) Kiribati had self-rule 1971-1979 and gained full independence from the UK since 1979. The first post-independence value available was used for the pre-independence years 1973-1978.
33) Micronesia was a UN Trust Territory under US administration until independent in 1986. The first post-independence value available was used for the pre-independence years 1973-1990.
34) Palau was a UN Trust Territory under US administration until independent in 1978. The first post-independence value available was used for the pre-independence years 1973-1993.
35) Papua New Guinea was administered by Australia until its independence in 1975. The modal post-independence value available was used for the pre-independence years 1973-1975.

Civil Liberties:

1) Data for Czechoslovakia from 1993-2000 are assigned the corresponding values for Czech Republic.
2) Data for Azores Island are assigned the corresponding values for Portugal.
3) Data for England, Scotland, Wales, Northern Ireland, Guernsey, Jersey and Isle of Man are assigned the corresponding values for United Kingdom.
4) Data for Czech Republic from 1973-1992 are assigned the corresponding values for Czechoslovakia.
5) Data for Slovakia (1973-1992), Bosnia & Herzegovina (1973-1991), Croatia (1973-1990), Macedonia (1973-1991), Slovenia (1973-1990), Serbia (1973-2000) and Kosovo (1973-2000) are assigned the corresponding values for Yugoslavia.
6) Data for Estonia, Latvia, Lithuania, Armenia, Azerbaijan, Belarus, Georgia, Moldova, Russia, Ukraine, Kazakhstan, Tajikistan, Turkmenistan and Uzbekistan from 1973-1990 are assigned the corresponding values for USSR. Data for Kyrgyzstan are assigned the corresponding values for USSR from 1973-1991 and for Russia from 1992-2000.
7) Data for USSR from 1992-2000 are assigned the corresponding values for Russia.

8)   Data for Hong Kong are assigned the value of 1 based on its economic freedom.
9)   Data for Korea are assigned the corresponding values for South Korea.
10)   Data for Macau are assigned a middle ranking of 3 throughout the data period. It was a colony of Portugal until 1999 when it became a Special Administrative Region of China.
11)   Data for Bermuda are assigned the corresponding values for United Kingdom throughout the data period, as it was a British colony.
12)   Data for Belize from 1973-1981 are assigned the corresponding values for United Kingdom, as it was a British colony.
13)   Data for Antigua & Barbuda from 1973-1981 are assigned the corresponding values for United Kingdom, as it was a British colony.
14)   Bahamas was a British colony until 1973, hence a value of 1 for 1973.
15)   Data for Dominica from 1973-1978 are assigned the corresponding values for United Kingdom, as it was a British colony.
16)   Data for Grenada from 1973-1974 are assigned the corresponding values for United Kingdom, as it was a British colony.
17)   Netherlands Antilles remains an autonomous country within the Netherlands, and a value of 1 was assigned.
18)   Data for St Kitts-Nevis from 1973-1981 are assigned the corresponding values for United Kingdom, as it was a British colony.
19)   Data for St Lucia from 1973-1979 are assigned the corresponding values for United Kingdom, as it was a British colony.
20)   Data for St Vincent & the Grenadines from 1973-1979 are assigned the corresponding values for United Kingdom, as it was a British colony.
21)   Suriname was a Netherlands colony until 1975. The first post-independence value available was used for the pre-independence years 1973-1975.
22)   Angola gained independence from Portugal in 1975, but the transition to independence was difficult. The first post-independence value available was used for the pre-independence years 1973-1975.
23)   Cape Verde was a colony of Portugal until 1975. The first post-independence value available was used for the pre-independence years 1973-1975.
24)   Comoros was a colony of France until 1975. The first post-independence value available was used for the pre-independence years 1973-1975.
25)   Djibouti was a colony of France until 1977. The first post-independence value available was used for the pre-independence years 1973-1977.
26)   Data for Eritrea from 1973-1992 are assigned the corresponding values for Ethiopia.
27)   Mozambique was a colony of Portugal until 1975. The first post-independence value available was used for the pre-independence years 1973-1975.
28)   Namibia was occupied by South Africa until 1988. The values for 1976-1988 were imputed using the values for 1975.
29)   Sao Tome & Principe gained independence from Portugal in 1975. The first post-independence value available was used for the pre-independence years 1973-1975.
30)   Seychelles gained independence from UK in 1976. The first post-independence value available was used for the pre-independence years 1973-1976.
31)   Datum for South Africa for 1973 is assigned based on the value of 1974, as there is an absence of significant events between 1973 and 1974 to suggest otherwise.
32)   Kiribati had self-rule 1971-1979 and gained full independence from the UK since 1979. The first post-independence value available was used for the pre-independence years 1973-1978.
33)   Micronesia was a UN Trust Territory under US administration until independent in 1986. The first post-independence value available was used for the pre-independence years 1973-1990.

34) Palau was a UN Trust Territory under US administration until independent in 1978. The first post-independence value available was used for the pre-independence years 1973-1993.

35) Papua New Guinea was administered by Australia until its independence in 1975. The modal post-independence value available was used for the pre-independence years 1973-1975.

Economic Freedom:

1) Data for Albania from 1970-1985 are assigned the value for 1990.
2) Data from Bulgaria from 1970-1980 are assigned the value for 1985.
3) Data for Czechoslovakia from 1970-1990 are assigned the value for 1995.
4) Data for Hungary from 1970-1975 are assigned the value for 1980.
5) Data for Malta from 1970-1975 are assigned the value for 1980.
6) Data from Poland from 1970-1980 are assigned the value for 1985.
7) Data from Romania from 1970-1980 are assigned the value for 1985.
8) Data for Czech Republic from 1970-1990 are assigned the value for 1995.
9) Data for Slovakia from 1970-1990 are assigned the value for 1995.
10) Data for Croatia from 1970-1990 are assigned the value for 1995.
11) Data for Slovenia from 1970-1990 are assigned the value for 1995.
12) Data for Estonia from 1970-1990 are assigned the value for 1995.
13) Data for Latvia from 1970-1990 are assigned the value for 1995.
14) Data for Lithuania from 1970-1990 are assigned the value for 1995.
15) Data for Russia from 1970-1990 are assigned the value for 1995.
16) Data for Ukraine from 1970-1990 are assigned the value for 1995.
17) Data for USSR from 1970-1990 are assigned the value for 1995.
18) Data for Bahrain from 1970-1975 are assigned the value for 1980.
19) Data for Bangladesh from 1970 are assigned the value for 1975.
20) Data for China from 1970-1975 are assigned the value for 1980.
21) Data for Cyprus for 1970 is assigned the value for 1975.
22) Data for Jordan for 1970 is assigned the value for 1975.
23) Data for Kuwait from 1970-1975 are assigned the value for 1980.
24) Data for Nepal from 1970-1975 are assigned the value for 1980.
25) Data for Oman from 1970-1980 are assigned the value for 1985.
26) Data for Sri Lanka from 1970-1975 are assigned the value for 1980.
27) Data for United Arab Emirates from 1970-1975 are assigned the value for 1980.
28) Data for Belize from 1970-1975 are assigned the value for 1980.
29) Data for Costa Rica for 1970 is assigned the value for 1975.
30) Data for El Salvador from 1970-1975 are assigned the value for 1980.
31) Data for Honduras from 1970-1975 are assigned the value for 1980.
32) Data for Nicaragua from 1970-1975 are assigned the value for 1980.
33) Data for Panama for 1970 is assigned the value for 1975.
34) Data for Bahamas for 1970 is assigned the value for 1975.
35) Data for Barbados for 1970 is assigned the value for 1975.
36) Data for Dominican from 1970-1975 are assigned the value for 1980.
37) Data for Haiti from 1970-1975 are assigned the value for 1980.
38) Data for Jamaica from 1970-1975 are assigned the value for 1980.
39) Data for Trinidad and Tobago for 1970 is assigned the value for 1975.
40) Data for Bolivia from 1970-1975 are assigned the value for 1980.

41) Data for Guyana from 1970-1975 are assigned the value for 1980, while data for 1990 is assigned the value for 1995.
42) Data for Paraguay from 1970-1975 are assigned the value for 1980.
43) Data for Uruguay from 1970-1975 are assigned the value for 1980.
44) Data for Algeria from 1970-1975 are assigned the value for 1980.
45) Data for Benin from 1970-1975 are assigned the value for 1980.
46) Data for Botswana from 1970-1975 are assigned the value for 1980.
47) Data for Burundi for 1970 is assigned the value for 1975.
48) Data for Cameroon from 1970-1975 are assigned the value for 1980.
49) Data for Central Africa from 1970-1975 are assigned the value for 1980.
50) Data for Chad from 1970-1980 are assigned the value for 1985.
51) Data for Congo from 1970-1975 are assigned the value for 1980.
52) Data for Egypt for 1970 is assigned the value for 1975.
53) Data for Gabon from 1970-1975 are assigned the value for 1980.
54) Data for Ghana for 1970 is assigned the value for 1975.
55) Data for Madagascar for 1975 is assigned the value for 1980.
56) Data for Malawi for 1970 is assigned the value for 1975.
57) Data for Mali for 1970 is assigned the value for 1975.
58) Data for Namibia from 1970-1985 are assigned the value for 1990.
59) Data for Niger for 1970 is assigned the value for 1975.
60) Data for Rwanda for 1970 and 1985 are assigned the value for 1975 and 1990 respectively.
61) Data for Senegal from 1970-1975 are assigned the value for 1980.
62) Data for Sierra Leone for 1970 is assigned the value for 1975.
63) Data for Togo from 1970-1975 are assigned the value for 1980.
64) Data for Uganda from 1970-1975 are assigned the value for 1980.
65) Data for Zambia for 1970 is assigned the value for 1975.
66) Data for Zimbabwe from 1970-1975 are assigned the value for 1980.
67) Data for Fiji for 1970 is assigned the value for 1975.
68) Data for Papua New Guinea from 1970-1980 are assigned the value for 1985.

Gross National Product:

A. Data missing for part of the period 1950-2000.
In these instances, the data are imputed using the product of the US GNP for the corresponding missing years and the ratio of the last available GDP data for the particular country to that of US GNP. That is, if data are missing for country X for 1950-1960, the missing data are imputed by

$$[\frac{\text{country X 1961 GNP}}{\text{US 1961 GNP}}]*\text{US GNP}_t , \quad t = 1950, \ldots, 1960.$$

This approach is taken for:
1) Albania for 1950-1989.
2) Bulgaria for 1950-1990.
3) Germany for 1950-1969.

4)   Greece for 1950.
5)   Hungary for 1950-1969.
6)   Malta for 1950-1969.
7)   Poland for 1950-1969.
8)   Romania for 1950-1959.
9)   Czech Republic for 1950-1989.
10)   Slovakia for 1950-1986.
11)   Bosnia & Herzegovina for 1950-1989.
12)   Croatia for 1950-1989.
13)   Macedonia for 1950-1989.
14)   Slovenia for 1950-1989.
15)   Serbia for 1950-1989.
16)   Estonia for 1950-1989.
17)   Latvia for 1950-1992.
18)   Lithuania for 1950-1992.
19)   Armenia for 1950-1994.
20)   Azerbaijan for 1950-1993.
21)   Belarus for 1950-1994.
22)   Georgia for 1950-1991.
23)   Moldova for 1950-1991.
24)   Russia for 1950-1991.
25)   Ukraine for 1950-1992.
26)   Afghanistan for 1950-1969.
27)   Bahrain for 1950-1969.
28)   Bangladesh for 1950-1971.
29)   Brunei for 1950-1969.
30)   Cambodia for 1950-1969.
31)   China for 1950-1951.
32)   Cyprus for 1950-1969.
33)   Hong Kong for 1950-1959.
34)   Indonesia for 1950-1959.
35)   Iran for 1950-1954.
36)   Iraq for 1950-1969.
37)   Korea and South Korea for 1950-1952.
38)   Kazakhstan for 1950-1989 (data of GDP per capital, current international dollars (PPPs) based on World Bank estimates).
39)   Kyrgyzstan for 1950-1986 (data of GDP per capital, current international dollars (PPPs) based on World Bank estimates).
40)   North Korea for 1950-1969.
41)   Kuwait for 1950-1969.
42)   Laos for 1950-1969.
43)   Lebanon for 1950-1990.
44)   Macau for 1950-1969.
45)   Malaysia for 1950-1954.
46)   Mongolia for 1950-1969.
47)   Nepal for 1950-1959.
48)   Oman for 1950-1969.
49)   Qatar for 1950-1969.
50)   Saudi Arabia for 1950-1969.
51)   Singapore for 1950-1959.
52)   Syria for 1950-1959.

53) Taiwan for 1950.
54) Tajikistan for 1950-1992.
55) Turkmenistan for 1950-1992.
56) United Arab Emirates for 1950-1969.
57) Uzbekistan for 1950-1989.
58) Vietnam for 1950-1988.
59) Yemen for 1950-1988.
60) Bermuda for 1950-1969.
61) Belize for 1950-1969.
62) Antigua and Barbuda for 1950-1969.
63) Bahamas for 1950-1969.
64) Barbados for 1950-1959.
65) Cuba for 1950-1969.
66) Dominica for 1950-1969.
67) Dominican Republic for 1950.
68) Grenada for 1950-1969.
69) Haiti for 1950-1969.
70) Jamaica for 1950-1952.
71) Netherlands Antilles for 1950-1969.
72) St Kitts-Nevis for 1950-1969.
73) St Lucia for 1950-1969.
74) St Vincent & the Grenadines for 1950-1969.
75) Chile for 1950.
76) Ecuador for 1950.
77) Guyana for 1950-1999.
78) Paraguay for 1950.
79) Suriname for 1950-1969.
80) Algeria for 1950-1959.
81) Angola for 1950-1999.
82) Benin for 1950-1958.
83) Botswana for 1950-1969.
84) Burkina Faso for 1950-1958.
85) Burundi for 1950-1959.
86) Cameroon for 1950-1959.
87) Cape Verde for 1950-1959.
88) Central African Republic for 1950-1969.
89) Chad for 1950-1959.
90) Comoros for 1950-1959.
91) Congo for 1950-1959.
92) Djibouti for 1950-1978.
93) Equatorial Guinea for 1950-1959.
94) Eritrea for 1950-1991.
95) Gabon for 1950-1959.
96) Gambia for 1950-1959.
97) Ghana for 1950-1954.
98) Guinea for 1950-1958.
99) Lesotho for 1950-1959.
100) Liberia for 1950-1969.
101) Libya for 1950-1999.
102) Madagascar for 1950-1959.
103) Malawi for 1950-1953.

104) Mali for 1950-1959.
105) Mauritania for 1950-1969.
106) Mozambique for 1950-1959.
107) Namibia for 1950-1969.
108) Niger for 1950-1959.
109) Rwanda for 1950-1959.
110) Sao Tome & Principe for 1950-1969.
111) Senegal for 1950-1959.
112) Seychelles for 1950-1999.
113) Sierra Leone for 1950-1969.
114) Somalia for 1950-1969.
115) Sudan for 1950-1969.
116) Swaziland for 1950-1969.
117) Tanzania for 1950-1959.
118) Togo for 1950-1959.
119) Tunisia for 1950-1960.
120) Democratic Republic of Congo for 1950-1969.
121) Zambia for 1950-1954.
122) Zimbabwe for 1950-1953.
123) Fiji for 1950-1969.
124) Kiribati for 1950-1969.
125) Micronesia for 1950-1969.
126) Palau for 1950-1979.
127) Papua New Guinea for 1950-1969.
128) Tonga for 1950-1969.
129) Samoa for 1950-1969.

B. Data missing for all of the period 1950-2000.

1) Data for Czechoslovakia are given by the average of the Czech Republic and Slovakia.
2) Data for Azores Island are imputed by taking half of the value of GDP for Portugal.
3) Yugoslavia and Kosovo are assigned the corresponding values for Serbia.
4) Data for USSR are given by the weighted average of the former republics (Russia, Ukraine, Armenia, Azerbaijan, Belarus, Georgia, Moldova, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan and Uzbekistan) with weights given by US Census 2000 population count.
5) Data for England, Scotland, Wales, Northern Ireland, Guernsey, Jersey and Isle of Man are assigned the corresponding values for United Kingdom.