DETERMINATIONS OF NATURALIZATION AND ITS IMPACT ON EARNINGS

Won Joong Kim, PhD
Kunkuk University, Korea
Yongseung Han, PhD
University of North Georgia
Myeong Hwan Kim, PhD
Indiana University-Purdue University Fort Wayne

Abstract

Many studies about the effects of naturalization on earnings do not account for the endogenous characteristic of naturalization. This study is concerned with the determinants of naturalization, its effect on earnings, and reversal causality between earnings and naturalization. Ordinary least squares (OLS), treatment effect model, and simultaneous equation model are used to estimate the effect of naturalization on earnings for U.S. immigrants. Using a treatment effect model, we find that naturalization has a much higher positive impact than the OLS method on earnings, but by employing a simultaneous equation model, the naturalization premium becomes closer to that of OLS. Also, we find that correcting the simultaneity problem between earnings and naturalization is much more important than orthogonalizing the naturalization variable using instruments.

Keywords: Naturalization; Earning; Treatment Effect Model; Simultaneous Equation Model

Introduction

The United States has long been known as “a land of immigrants,” not only because its founding fathers were immigrants but also because there has been a continuous influx of immigrants. Since 1820, more than 70 million immigrants have entered the United States. During the period of 1995 to 2000 alone more than 7 million foreign-born individuals entered, which set a record high in the U.S. history. Thus, foreign-born individuals make up 11.1 percent of the total population, as of 2000.

Immigration issues are complex. It has been both a crucial component of America’s growth and, particularly in recent years, one of the most contentious issues on the nation’s political agenda. For this reason, a

vast volume of studies have been performed to understand the role and impact of immigration on the economy and the society as well as to guide policy-makers, politicians and various groups of interest.

Economic consequences of immigration, simply speaking, provide an ideal source of labor and increase productivity growth. In this regard, many empirical studies focused on how immigrants perform in the U.S. labor market and whether immigration hurts employment opportunities of native workers. For the former question, there is a consensus that immigrants work hard for low pay. Particularly, the pioneering work of Chiswick (1978) shows that immigrants earn about 17 percent less than comparable natives at the time of arrival, but immigrant earnings overtake native earning within 15 years due to wage growth as immigrants assimilate to the host country. It has also been found that the specific wage gap between immigrants and native workers depends upon various factors: education, work experience, demographic characteristics, English proficiency as a proxy for degree of by assimilation, and so on.

There are various source of higher wage growth in the process of immigrant's assimilation. Higher wage growth can be ascribed to the characteristics of immigrants: ability and higher motivation (Chiswick 1978) or willingness to work longer and harder (Carliner 1980). It can also be ascribed to the accumulation of human capital as in economic growth literature. However, the result of higher wage growth may not be as robust as has been claimed. Borjas (1985) hypothesized that immigrants of 1970s may be intrinsically different from those of 1990s because of changes in immigration policy or changes in conditions in the source countries and the U.S. He then showed that there exist cohort effects, with more recent immigrant cohorts having relatively lower wage rates. Bratsberg *et al*. (2006) showed that higher wage growth and cohort effects are negligible for less educated immigrants.

Lower wages of immigrants increases profits for firms and lowers prices for consumers, therefore spurring economic growth. Borjas (2001) estimated that immigration into the U.S. depressed wages by $152 billion annually, but increased profits by $160 billions. A half century ago, Thomas (1954) clarified underlying patterns showing that the inflow of immigrants, coupled with foreign capital inflows, helped push the American economy in its upswings and slowed the growth phase in the source countries.

For the second question, Borjas (1994) found it unacceptable that the employment opportunities of native workers are strongly and adversely affected by immigration although immigration lowers native wages in the local labor markets.

Despite economic consequences, there exists criticism of immigration. Some fear the threat to the status quo posed by immigrants with

different social values and cultures and by the higher birth rate of immigrants. Some also criticize the burden of public assistance programs for immigrants. One private organization suggests that illegal immigration costs the state of Florida $1.7 billion per year for education, medical care and incarceration (Martin and Hehlman 2005). The issue of immigration becomes more contentious as the values and cultures of recent immigrants diverge more from European and Christian origins and as fiscal burdens of governments increase.

Immigrant naturalization or citizenship acquisition become more important as the criticism of immigration intensifies for the following reasons. First, immigrant naturalization helps reduces criticism about immigration because it signifies the immigrants' allegiance and commitment to the host country. Second, citizenship acquisition provides an effective way to have political influence on immigration policy as well as on other public policies, such as education and medical care. As Yang (1994) noted, the benefits of naturalization are more opportunities in employment (e.g., some jobs in federal government agencies), welfare assistance (e.g., social security benefits), and education (e.g., loans and fellowships). Some costs are an increase in obligations (e.g., serving in the military if drafted), a reduction in rights in the source countries, and an actual cost in the long and complicated application process of naturalization.

Given this benefit and cost of naturalization, some studies were devoted to find out what affects naturalization. Earlier studies emphasized the role of socioeconomic status and showed that immigrants with higher education and higher family income were more likely to obtain citizenship (Bernard 1936), whereas recent studies emphasized the role of assimilation (e.g., English competence) into the U.S. (Barken and Khokhlov 1980), country-of-origin characteristics (Woodrow-Lafield *et al*. 2004), personal ties (Street 2013), social context of immigrant reception (van Hook *et al*. 2006), and immigration policy (Rallu 2011). However, the empirical evidence is mixed, suggesting that the decision of naturalization is much complicated.

It is evident that the study on the performance of immigrants in the labor market must consider the values of naturalization when the criticism of immigration mounts, when naturalization is an effective way to counter the criticism, and when the decision of naturalization is affected by many factors, socioeconomic and non-socioeconomic. In this respect, the study of Bratsberg *et al*. (2002) is noteworthy: a higher wage growth occurs only after naturalization, although accumulation of human capital began before naturalization, because naturalization facilitates assimilation into the U.S. labor market. However, Bratsberg *et al*. (2002) did not consider endogenous characteristics of factors, such as period of immigration, county-of-origin

and English proficiency, that affect both the decision of naturalization and the wages of immigrants. Therefore, Bratsberg *et al*. (2002) may be subject to endogenous bias since the endogeneity of naturalization is not taken into account.

In this paper, we consider endogeneity of naturalization to determine the wages of naturalized immigrants. For this, we first consider simple ordinary least squares (OLS) estimation without considering this endogeneity and then perform a test to see whether there exists an endogeneity problem, confirming its existence in the simple OLS estimation. So we consider alternative way to estimate wage equation to tackle the problem. Treatment effect and simultaneous equation models are used to show that naturalized immigrants earn more than non-naturalized immigrants.

Model Specification and Data

OLS Estimation

Our basic interest can be described using a simple model:

$$\ln Y_i = N_i\gamma_1 + x_{i1}\beta_1 + \varepsilon_{i1} \quad (1)$$

where $\ln Y_i$ is a log of annual earnings of individual i, N_i is an index variable for naturalization, x_{i1} is a vector of control variables, ε_{i1} is the error term. If naturalization is purely exogenous, then OLS estimation is unbiased. However, if naturalization is endogenous, then coefficient estimate on naturalization is biased.

Tests for Endogeneity

To test for the endogeneity of the naturalization, we employ the Durbin-Wu-Hausman (DWH) test. Suppose naturalization is correlated with error term in Equation (1), and assume that the naturalization is a function of family income (total family income minus interviewee's total wages and salaries, and measured in thousand dollars) and other exogenous variables such as education, occupation, industry and country of origin such that

$$N_i = x_{i2}\beta_2 + \varepsilon_{i2} \quad (2)$$

where x_{i2} is a vector of instruments and ε_{i2} is the error term (for estimation purposes, we assume $\varepsilon_{2i} \sim N(0,\sigma^2)$ for OLS, and $\varepsilon_{2i} \sim N(0,1)$ for probit in Equation (2)). To test the null that N_i is uncorrelated with ε_{i1} in Equation (1), first get estimated residuals of Equation (2) and defined as naturalization index (N) minus estimated probability of being a naturalized, $\hat{\varepsilon}_{i2}$, then perform an augmented regression of Equation (1):

$$\ln Y_i = N_i\gamma_2 + \hat{\varepsilon}_{i2}\pi + x_{i1}\beta_3 + u_{i1} \quad (3)$$

The only difference between Equations (1) and (3) is that we add residuals from the probit instrumental variable (IV) Equation to (1). If the coefficient estimate on the residual $\hat{\pi}$ is significantly different from zero, then the null hypothesis that the OLS estimator is uncorrelated with error term will be rejected, hence OLS is not consistent.

Treatment Effect Model

There is a possibility that earnings may affect the naturalization decision. We, therefore, consider the following treatment effect model (See Maddala (1983) for more details.). Let

$$\ln Y_i^* = N_i^* \gamma_1 + x_{i1}\beta_1 + \varepsilon_{i1} \quad (4)$$

$$N_i^* = x_{i2}\beta_2 + \varepsilon_{i2} \quad (5)$$

where $\ln Y_i^*$ is the log of earnings, N_i^* is the naturalization status, x_{i1} and x_{i2} are vectors of exogenous variables, $\varepsilon_{i1} \sim N(0,\sigma^2)$, $\varepsilon_{i2} \sim N(0,1)$ and $Corr(\varepsilon_{i1},\varepsilon_{i2}) = \rho$.

$\ln Y_i^*$ and N_i^* are assumed to be observed as

$\ln Y_i = Y_i^*$ $N_i = 1$ if $N_i^* > 0$ and $N_i = 0$ otherwis

thus making naturalization exogenous by using instruments. The instruments for naturalization are family income, annual amount of return to house equity, and some of the explanatory variables used in the earnings equation.

Simultaneous Equation Model

There is a possibility that there is a simultaneity problem between earnings and naturalization. Therefore, we consider the following simultaneous equation model. Let

$$\ln Y_i^* = N_i^* \gamma_1 + x_{i1}\beta_1 + \varepsilon_{i1} \quad (6)$$

$$N_i^* = \ln Y_i^* \gamma_2 + x_{i2}\beta_2 + \varepsilon_{i2} \quad (7)$$

All the specifications in the simultaneous equation are the same as in the treatment effect model, except that we explicitly consider the earnings effect in the naturalization probit equation.

The procedure of obtaining robust standard errors is exactly the same as in Heckman (1978), except that we add the log of earnings into Equation (7). The procedures are as follow: First, from the structural equations of Equation (6) and (7), we derive reduced form of the earnings and the naturalization equations. Then, for earnings equation, we run OLS and get constructed earnings, and for the naturalization equation, we run a probit and get a constructed naturalization. After obtaining the constructed earnings and naturalization, we use these instead of endogenous variables on the right

hand side of Equations (6) and (7). Finally, using a suitable procedure, we derive robust standard errors (See Maddala (1983) for more details.).

Data

The data are primarily taken from the Current Population Survey (CPS) March Supplement for 1994-2002. Our research is confined to those periods because the CPS started collecting information on citizenship after 1994, and no data on race is available. Countries' native, official and common languages are from the World Factbook. By matching the country of birth, father's country of birth, and mother's country of birth with language, we are able to construct the language dummies that equal 1 if English is at least commonly used in the country of birth. Our data are limited to the full-time male worker whose age is between 18 and 65, and this consists of 31,482 total observations.

Our focus is on the effect of naturalization on earnings among immigrants, so we exclude the observations of U.S. native citizens. We use age as a proxy for experience. We could construct 'the potential experience' as suggested by Mincer (1974) to account for work experience premium. He defined 'potential experience' as 'age − years of schooling − 6' and educational attainment is categorical in the CPS. So, if we take the mean values of each category, we confront two problems. First, potential experience can have a negative value. Second, unlike the U.S., other countries may have different years of schooling system. Therefore, we use age as a proxy for experience. As Chiswick (1978) points out, ignoring the number of years since migration would mask important differences among foreign born immigrants. Therefore, we include 'years since migration (YSM) to US,' and it is defined as 'survey year – year of entry.' The years of entry is also categorical in the CPS, so we take the mean of each category and subtract from the interview year. Since we use observation for the age in between 18 and 65, for the category of 'before 1950', we calculate YSM as 'interview year – [(interview year-65)-1949]/2.' Also, by law, foreigners should stay, in general, in the United State at least five years. Therefore, observations with less than 5 years of entry to United States are excluded. For odd (even) year, observations with less than 6 (5) years of entry to U.S. are excluded. This is due to the limitation of CPS dataset.

Annual earnings with less than $1,000 are excluded from our research. We also exclude three top-coded observations in earnings. Regarding the annual amount of returns to house equity, we exclude bottom-coded observations (-9999). Family income is defined as (total family income – wages and salaries). Amount of returns to house equity and family income are measured in thousand dollars. Finally, we exclude countries if their sample size is less than 5 for each country. Deleted countries and their

sample size are Finland (3), Northern Island (3), Lithuania (4) and Bermuda (2).

Since many of the variables are in discrete forms in the CPS questionnaires, we regroup categories. We reduce marriage categories into 3 (married, separated/divorced, never married), education categories into 4 (less than high school, high school, junior college degree [including observations with college but no degree], college degree or above), and race categories into 3 (white, black, others).

We categorize industries into 12 groups, occupations into 13 groups, and classes of jobs into 3 groups. The definition and categorization of industries, occupations and class of jobs are illustrated in Table 1 and in Figure 1.

Table 1. Means and Definitions of Variables

Variable	All Group	Non-naturalized	Naturalized
Earnings	29814.96	24861.72	38258.46
Log of Earnings	10.0317	9.8636	10.3183
Family income (000's)	2.7934	1.9051	4.3077
Return to House Equity (000's)	2.3307	1.6938	3.4164
Years Since Migration (YSM)	17.9751	14.8310	23.3347
Age	38.7285	36.5274	42.4807
Inequality	0.0522	0.0490	0.0577
Naturalization dummies	0.3697		
Race dummies			
White	0.7441	0.7998	0.6492
Black	0.0629	0.0567	0.0735
Other (omitted)	0.1930	0.1435	0.2773
Education dummies			
Less than high school	0.3771	0.4817	0.1988
High school degree (omitted)	0.2463	0.2442	0.2499
Junior College	0.1598	0.1286	0.2131
College and over	0.2168	0.1455	0.3381
Marital status Dummies			
Married	0.7155	0.6846	0.7680
Separated/divorced	0.0733	0.0673	0.0835
Never married	0.2112	0.2481	0.1485
English fluency dummies			
English	0.1661	0.1273	0.2323
Official or common language			
English (Father)	0.1581	0.1218	0.2200
English (Mother)	0.1586	0.1222	0.2208
Industry dummies			
Ag./Forestry/Fisheries (omitted)	0.0586	0.0777	0.0261
Mining	0.0037	0.0034	0.0044

Construction	0.1198	0.1408	0.0842
Manufacturing	0.2355	0.2382	0.2309
Transport/Comm./Utility	0.0745	0.0605	0.0983
Wholesale/Retail Trade	0.2137	0.2265	0.1919
Finance/Ins./Real Estate	0.0396	0.0315	0.0534
Personal Services	0.0387	0.0401	0.0364
Business/Repair Services	0.0840	0.0906	0.0726
Entertainment/Rec. Services	0.0150	0.0140	0.0166
Professional Services	0.0989	0.0702	0.1479
Public Administration	0.0179	0.0067	0.0372
Occupation dummies			
Executive, Admin. & Managerial (omitted)	0.0919	0.0629	0.1412
Professional Specialty	0.0978	0.0652	0.1533
Technicians/Related Support	0.0218	0.0143	0.0346
Sales Occupations	0.0677	0.0547	0.0898
Admin. Support, Incl. Clerical	0.0482	0.0402	0.0618
Private Household Occ.	0.0009	0.0010	0.0008
Protective Service Occ.	0.0116	0.0082	0.0174
Service, Excl. Protective Hhld.	0.1367	0.1568	0.1023
Precision Prod., Craft & Repair	0.1985	0.2106	0.1779
Machine Opers, Assmblers & Inspectors	0.1146	0.1326	0.0838
Transportation and Meterial Moving	0.0688	0.0735	0.0607
Handlers, Equip. Cleaners, Helpers, Laborers	0.0795	0.0981	0.0478
Farming, Forestry and Fishing	0.0621	0.0817	0.0286
Class of Job dummies			
Government	0.0622	0.0323	0.1132
Private wage and salary sector	0.8866	0.9355	0.8031
Self-employed (omitted)	0.0512	0.0322	0.0837
Year dummies			
Year – 1994 (omitted)	0.0843	0.0900	0.0747
Year - 1995	0.0945	0.0981	0.0883
Year - 1996	0.0965	0.1034	0.0849
Year - 1997	0.1011	0.1000	0.1029
Year - 1998	0.1080	0.1042	0.1144
Year - 1999	0.1056	0.1042	0.1082
Year - 2000	0.1204	0.1221	0.1176
Year - 2001	0.1159	0.1127	0.1212
Year - 2002	0.1736	0.1652	0.1878
Observation	31,482	19,842	11,640

Finally and most interestingly, we employ a measure of inequality to investigate the effect of inequality on naturalization. We define inequality as 'standard deviation of log of earnings for people with the same educations, in the same industries, and with the same occupations.' We delete observations

with zero inequality. In our sample, zero inequality means that sample size within a cell is less than one in the relevant cells. A person may easily compare his earnings with the earnings of his colleagues' earnings with the same educational background in the same position within the same company. Therefore, our measure of inequality may be a good instrument for naturalization decision.

Figure 1. Average Earnings and Family Income by Naturalization Status

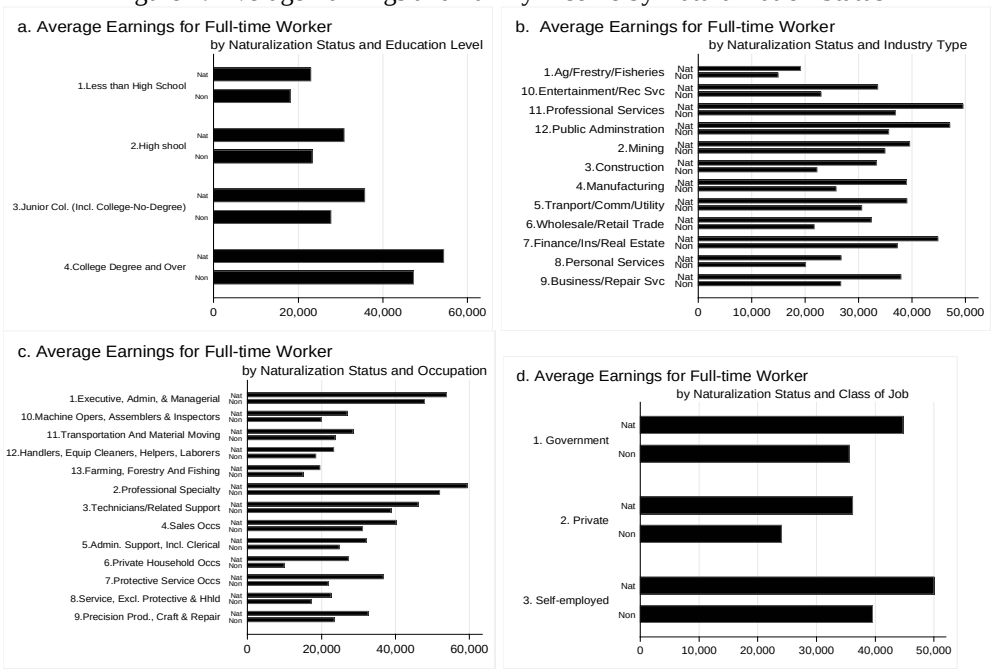
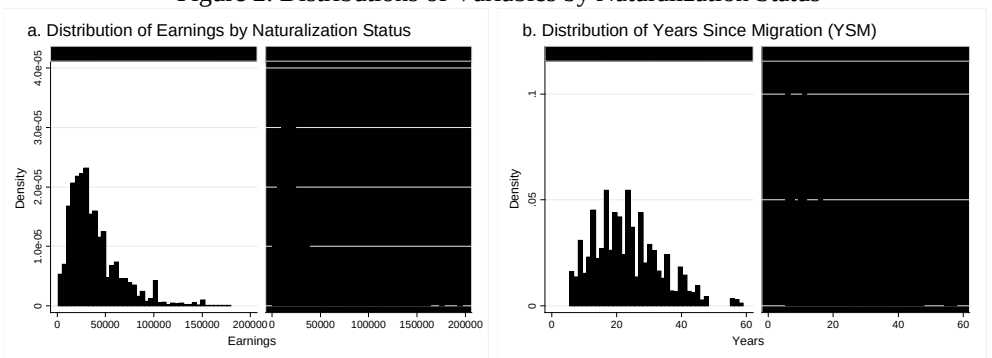
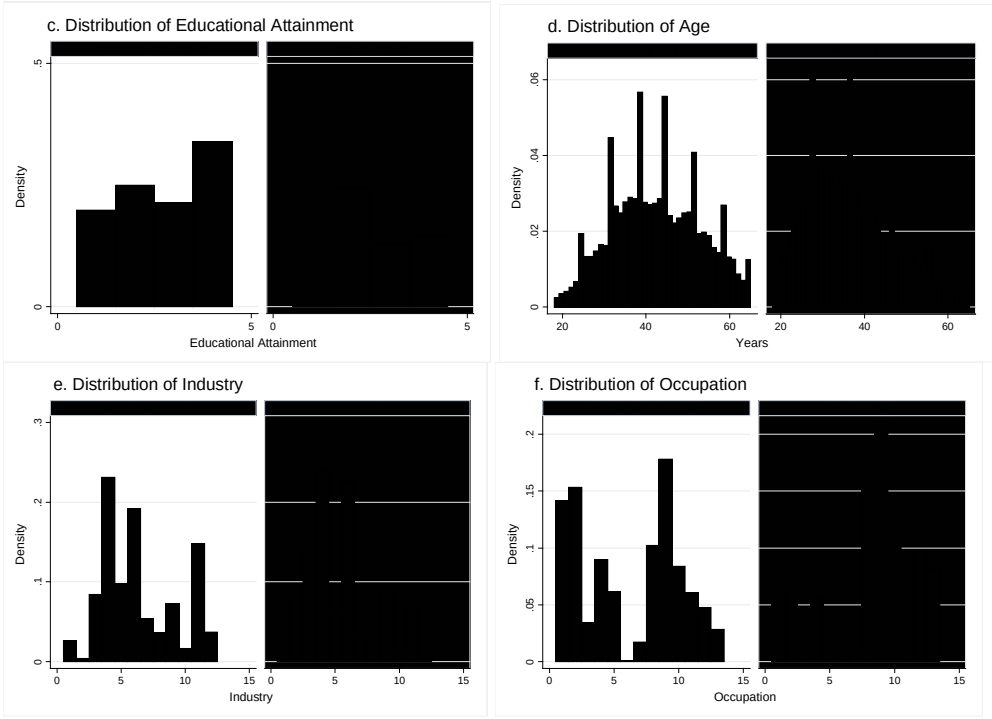


Figure 2. Distributions of Variables by Naturalization Status





Empirical Results

OLS Estimation Results

The OLS estimation results are shown in Table 2 and Table 3. Table 2 presents the estimation results without interaction terms. There are some differences in the coefficient estimate of naturalization, depending on the specification of the model, but all show positive and significant effects on earnings. The coefficient estimate is the greatest (0.130) in column 1, and the smallest (0.081) in column 4 of Table 2. As Chiswick (1978) points out, the YSM variable is very important and correlated with naturalization. By including YSM, the coefficient estimate of naturalization drops significantly. In column 3, we include an English dummy and find that, if he is from a country where English is either an official or commonly used language, there is a 0.8 percent earnings premium. Finally, in column 4, we include country dummies, and the naturalization premium is 8.1 percent, which is the smallest premium among four different specifications.

Table 2. OLS Results for Earnings Equation

Explanatory Variable	Dependent Variable: Log of Earnings			
	(1)	(2)	(3)	(4)
Naturalization	0.130^{***}	0.083^{***}	0.084^{***}	0.081^{***}
	(0.008)	(0.008)	(0.008)	(0.008)
Age	0.058^{***}	0.055^{***}	0.055^{***}	0.054^{***}
	(0.002)	(0.002)	(0.002)	(0.002)

Age Squared	-0.001***	-0.001***	-0.001***	-0.001***
	(2.78e-5)	(2.77e-5)	(2.77e-5)	(2.76e-5)
Education (High School Degree omitted)				
Less than High School	-0.190***	-0.188***	-0.187***	-0.157***
	(0.009)	(0.009)	(0.009)	(0.009)
Junior College	0.065***	0.059***	0.059***	0.058***
	(0.011)	(0.011)	(0.011)	(0.011)
College Degree and Over	0.239***	0.255***	0.254***	0.238***
	(0.012)	(0.012)	(0.012)	(0.013)
Marital Status (Separate/Divorced omitted)				
Married	0.122***	0.127***	0.128***	0.135***
	(0.013)	(0.013)	(0.013)	(0.013)
Never Married	-0.054***	-0.055***	-0.055***	-0.052***
	(0.015)	(0.015)	(0.015)	(0.015)
Race (Omitted: other than white or black)				
White	0.026***	-0.003	-4.40E-04	-0.027
	(0.010)	(0.010)	(0.010)	(0.018)
Black	-0.082***	-0.086***	-0.085***	-0.066**
	(0.016)	(0.016)	(0.016)	(0.027)
English (Father)	0.081**	0.084**	0.038	0.050
	(0.036)	(0.036)	(0.038)	(0.038)
English (Mother)	0.022	0.020	-0.035	-0.033
	(0.036)	(0.036)	(0.038)	(0.039)
English			0.105***	
			(0.028)	
Year Since Migration (YSM)		0.008***	0.008***	0.007***
		(4.41e-4)	(4.41e-4)	(4.58e-4)
Constant	8.731***	8.727***	8.724***	8.924***
	(0.062)	(0.062)	(0.062)	(0.127)
Fixed Effects				
Industry	YES	YES	YES	YES
Occupation	YES	YES	YES	YES
Class of Job	YES	YES	YES	YES
Year	YES	YES	YES	YES
Country	NO	NO	NO	YES
R^2	0.38	0.39	0.39	0.40

Note: Standard errors in parentheses. *, ** and *** denote significance at the 10, 5, and 1 percent levels, respectively.

The results in Table 3 are the extension of the specification in column 4 of Table 2, except we include the interaction terms. By including interaction terms, we find that the naturalization premium varies depending on situations. The coefficient estimate on naturalization is statistically insignificant at the conventional levels. For those who have less than a high school degree, becoming a citizen lowers the earnings by 3.7 percent. The citizenship premium is more apparent in occupational categories. If he has a sales occupation, being a citizen raises earnings by 7.8 percent, and if he has a private household occupation, being a citizen raises earnings by 58.5 percent. In sum, Table 3 shows that the naturalization premium exists in different occupations.

Table 3. OLS Results for Earnings Equation with Interaction Terms

Explanatory Variable	Dependent Variable: Log of Earnings	
	Level	Interaction Terms with Naturalization
Naturalization	0.035	
	(0.140)	
Age	0.057***	-4.47E-04
	(0.003)	(0.005)
Age Squared	-0.001***	4.46E-05
	(3.62e-5)	(5.98E-05)
Education (High School Degree omitted)		
Less than High School	-0.147***	-0.037*
	(0.011)	(0.020)
Junior College	0.048***	0.017
	(0.015)	(0.022)
College Degree and Over	0.239***	-0.003
	(0.018)	(0.025)
Marital Status (Separate/Divorced omitted)		
Married	0.121***	0.031
	(0.017)	(0.027)
Never Married	-0.045**	-0.038
	(0.019)	(0.032)
Race (Omitted: other than white or black)		
White	-0.023	-0.012
	(0.019)	(0.020)
Black	-0.065**	-0.006
	(0.030)	(0.032)
English (Father)	0.073	-0.045
	(0.053)	(0.071)
English (Mother)	-0.008	-0.051

	(0.053)	(0.071)
Year Since Migration (YSM)	0.007***	-0.001
	(0.001)	(0.001)
Industry (Ag./Forestry/Fisheries omitted)		
Mining	0.405***	-0.079
	(0.081)	(0.134)
Construction	0.088**	0.020
	(0.038)	(0.079)
Manufacturing	0.179***	0.008
	(0.037)	(0.077)
Transport/Comm./Utility	0.231***	-0.007
	(0.040)	(0.079)
Wholesale/Retail Trade	0.034	-0.025
	(0.036)	(0.077)
Finance/Ins./Real Estate	0.164***	-0.005
	(0.042)	(0.082)
Personal Services	0.069*	-0.010
	(0.040)	(0.083)
Business/Repair Services	0.039	-0.015
	(0.038)	(0.079)
Entertainment/Rec. Services	-0.044	0.089
	(0.047)	(0.087)
Professional Services	0.020	0.069
	(0.040)	(0.079)
Public Administration	0.048	0.099
	(0.069)	(0.103)
Occupation (Executive, Admin. & Managerial omitted)		
Professional Specialty	0.048*	0.015
	(0.026)	(0.034)
Technicians/Related Support	-0.138***	0.077
	(0.040)	(0.052)
Sales Occupations	-0.275***	0.078**
	(0.026)	(0.036)
Admin. Support, Incl. Clerical	-0.390***	0.045
	(0.028)	(0.040)
Private Household Occ.	-1.165***	0.585**
	(0.136)	(0.243)
Protective Service Occ.	-0.593***	0.389***
	(0.051)	(0.070)
Service, Excl. Protective Hhld.	-0.541***	0.004
	(0.022)	(0.034)
Precision Prod., Craft & Repair	-0.364***	0.087***

	(0.022)	(0.031)
Machine Opers, Assmblers & Inspectors	-0.501***	0.071*
	(0.024)	(0.037)
Transportation and Meterial Moving	-0.408***	-0.004
	(0.025)	(0.039)
Handlers, Equip. Cleaners, Helpers, Laborers	-0.546***	0.004
	(0.024)	(0.040)
Farming, Forestry and Fishing	-0.597***	0.087
	(0.039)	(0.077)
Class of Job (Self-employed omitted)		
Government Sector	0.089**	-0.050
	(0.038)	(0.048)
Private wage and salary sector	0.041*	-0.021
	(0.025)	(0.033)
Year Fixed Effects / Country Fixed Effects	YES/YES	
R^2	0.40	

Note: Standard errors in parentheses. *, ** and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Tests for Endogeneity

To test for endogeneity using Durbin-Wu-Hausman (DWH) test suggested by Davidson and MacKinnon (1993), we need to first estimate IV equation.

Let's consider the following three equations. To test for endogeneity, first we obtain residuals of Equation (9) from either a regression or a probit of naturalization on other explanatory variables and then add the estimated residual to Equation (8). An *F*-test on $\hat{\varepsilon}_{i2}$ in Equation (10) determines the validity of the null that naturalization is not correlated with error term in Equation (8). The specification of Equation (8) corresponds to the second column of Table 2. Also instruments for naturalization equation [Equation (9)] are the family income, the annual amount of return to house equity, a measure of inequality, year since migration, educational dummies, marital status dummies, race dummies, industry dummies, occupation dummies and country dummies.

$$\ln Y_i^* = N_i^* \gamma_1 + x_{i1}\beta_1 + \varepsilon_{i1} \quad (8)$$

$$N_i^* = x_{i2}\beta_2 + \varepsilon_{i2} \quad (9)$$

$$\ln Y_i^* = \hat{\varepsilon}_{i2}\pi + N^*\gamma_3 + x_{i1}\beta_3 + u_{i3} \quad (10)$$

First we use linear probability model on Equation (9) and test for endogeneity. The *F*-statistic on $\hat{\varepsilon}_{i2}$ is 30.71 and significant at one percent significance level. Using a probit model on Equation (9) produces a similar

result: the F-statistic on $\hat{\varepsilon}_{i2}$ is 44.50 and also statistically significant at one percent significance level. Therefore, a null hypothesis that naturalization is not correlated with error term is rejected.

Treatment Effect Model Estimation Results

The treatment effect estimation results are summarized in Table 4. The first column in Table 4 shows the estimation results using instruments. The coefficient estimate on constructed naturalization is 0.272, which is much higher compared with that from OLS.

Table 4. Treatment Effect Model

Explanatory Variable	Dependent Variable		
	Earnings	Naturalization	
		Estimation Results	Marginal Effect
Constructed Naturalization	0.272***		
	(0.033)		
Family Income		0.002**	0.001
		(0.001)	
Return to House Equity		0.023***	0.007
		(0.002)	
Inequality		-0.429**	-0.081
		(0.186)	
Year Since Migration (YSM)	0.004***	0.059***	0.022
	(0.001)	(0.001)	
Age	0.055***	-4.88E-04	-1.30E-05
	(0.002)	(0.001)	
Age Squared	-0.001***		
	(2.95e-5)		
Education (High School Degree omitted)			
Less than High School	-0.165***	-0.307***	-0.113
	(0.010)	(0.023)	
Junior College	0.049***	0.107***	0.039
	(0.011)	(0.026)	
College Degree and Over	0.238***	0.196***	0.077
	(0.014)	(0.029)	
Marital Status (Separate/Divorced omitted)			
Married	0.124***	0.044	0.017
	(0.014)	(0.033)	
Never Married	-0.051***	-0.142***	-0.049
	(0.016)	(0.038)	
Race (Omitted: other than white or black)			

White	0.025**	-0.169***	-0.056
	(0.011)	(0.042)	
Black	-0.073***	-0.043	-0.015
	(0.017)	(0.063)	
English (Father)	0.084**		
	(0.041)		
English (Mother)	0.023		
	(0.041)		
Constant	8.675***	-0.834***	
	(0.068)	(0.287)	
Fixed Effects			
Industry	YES	YES	
Occupation	YES	YES	
Class of Job	YES	YES	
Year	YES	NO	
Country	NO	YES	
ρ	-0.1940		
σ	0.6027		
λ	-0.1169		
Wald Test of Independent Equations (ρ=0)	χ^2 (1) = 36.84	Probability > χ^2 = 0.000	

Note: Standard errors in parentheses. *, ** and *** denote significance at the 10, 5, and 1 percent levels, respectively.

The second column and the third column in Table 4 show the estimation results and marginal effects for treatment equation of naturalization. We use family income, annual amount of return to house equity, and inequality as instruments, in addition to some variables used in earnings equation. We find that family income and return to house equity has positive effect on naturalization. A $10,000 increase in family income raises the probability of naturalization by one percent, and a $10,000 increase in annual amount of return to house equity increase the probability of naturalization by 7 percent.

One might expect that, if there is a higher inequality within the same education level in the same industry with the same occupation, he will be more likely to try to become a citizen. The result shown in Table 4 is quite opposite: a 10 percent increase in inequality decreases the probability of naturalization by 0.8 percent. One possible explanation to this is that naturalization is not determined solely by applicants' will: U.S. government may ask for a certain applicant requirement, such as English fluency. Also, if the inequality is higher in a lower education level, less human capital-intensive occupation, and less skill-intensive industry, then the negative effect of inequality on naturalization might be due to the requirements for naturalization by the U.S. government.

Simultaneous Equation Estimation Results

Table 5 presents the simultaneous equation estimation results. The first column shows estimates of the earnings equation. Here, the coefficient estimate on constructed naturalization is 0.084, which happens to be very similar to the result from OLS. Compared with the treatment effect model, the coefficient estimate on constructed naturalization in the simultaneous equation model is much lower. It suggests that naturalization is not only correlated with the error term in OLS, but there is also a reverse causality between naturalization and earnings.

Table 5. Simultaneous Equation Model

Explanatory Variable	Dependent Variable		
	Earnings	Naturalization	
		Estimates	Marginal Effects
Constructed Naturalization	0.084***		
	(0.012)		
Constructed Log of Earnings		0.374***	0.137
		(0.061)	
Family Income		0.002**	0.001
		(0.001)	
Return to House Equity		0.012***	0.004
		(0.002)	
Inequality		0.055	0.020
		(0.189)	
Year Since Migration (YSM)	0.004***	0.057***	0.021
	(0.001)	(0.001)	
Age	0.054***	-0.001	3.47E-04
	(0.002)	(0.001)	
Age Squared	-0.001***		
	(2.79e-5)		
Education (High School Degree omitted)			
Less than High School	-0.162***	-0.249***	-0.090
	(0.010)	(0.026)	
Junior College	0.050***	0.080***	0.030
	(0.011)	(0.027)	
College Degree and Over	0.240***	0.110***	0.041
	(0.013)	(0.034)	
Marital Status (Separate/Divorced omitted)			
Married	0.124***	0.002	5.75E-04
	(0.013)	(0.033)	
Never Married	-0.048***	-0.087**	-0.031
	(0.016)	(0.038)	

Race (Omitted: other than white or black)			
White	0.025**	-0.148***	-0.055
	(0.011)	(0.042)	
Black	-0.073***	-0.039	-0.140
	(0.016)	(0.065)	
English (Father)	0.083**		
	(0.036)		
English (Mother)	0.026		
	-0.036		
Constant	8.832***	-4.690***	
	(0.064)	(0.693)	
Fixed Effects			
Industry	YES	YES	
Occupation	YES	YES	
Class of Job	YES	YES	
Year	YES	NO	
Country	NO	YES	

Note: Standard errors in parentheses. *, ** and *** denote significance at the 10, 5, and 1 percent levels, respectively.

The second and third columns in Table 5 show the estimates and marginal effects of the naturalization equation. Interestingly, in a simultaneous equation model, the coefficient estimate on inequality becomes insignificant. The only difference between the treatment effect model and the simultaneous equation model in the naturalization equation is that we include constructed earnings in the simultaneous equation. We, therefore, see that the inequality is negatively correlated with earnings. Therefore, one possible explanation on the negative effect of inequality on naturalization is more persuasive: a person with higher earnings can more easily become a citizen if he wants to, whereas a person with lower earnings may not, due to some requirements from U.S. government, hence inequality persists. A one percent increase in earnings increases the probability of naturalization by 13.7 percent. We also find that an additional year after immigration increases the probability of naturalization by 2 percent.

Conclusion

In this paper, we examine the relationship between naturalization and earnings. Using OLS without interaction terms, we find that there exists a naturalization premium ranging from 0.081 to 0.130. By including interaction terms, we find that the naturalization premium itself does not exist, and it varies depending on situations. The citizenship premium is more apparent in certain occupation categories. If he has a sales occupation, being

a citizen raises earnings by 7.8 percent, and if he has a private household occupation, being a citizen raises earnings by 58.5 percent. In sum, our findings show that the naturalization premium exists in different occupations.

Using the treatment effect model, we find that the coefficient estimate on constructed naturalization is 0.272, which is much higher compared with that from OLS. Using family income, the annual amount of return to house equity, and inequality as instruments, in addition to some variables used in earnings equation, we find that family income and return to house equity has a positive effect on naturalization. A $10,000 increase in family income raises the probability of naturalization by one percent and a $10,000 increase in annual amount of return to house equity increase the probability of naturalization by 7 percent.

Compared with the treatment effect model, the coefficient estimate on constructed naturalization in the simultaneous equation model is much lower. It suggests that naturalization is not only correlated with the error term in OLS, but there is also a reverse causality between naturalization and earnings. Interestingly, in the simultaneous equation model, the coefficient estimate on inequality becomes insignificant. The only difference between the treatment effect model and the simultaneous equation model in our naturalization equation is that we include constructed earnings in simultaneous equation. We see that inequality is negatively correlated with earnings. Therefore, our one possible explanation of the negative effect of inequality on naturalization is more persuasive: a person with higher earnings can more easily become a citizen if he wants to, but a person with lower earnings may not easily become a citizen because of some requirements from the U.S. government, hence inequality persists.

References:

Barken, E. R., & Khokhlov, N. (1980). Socioeconomic data as indices of naturalization patterns in the United States: a theory revisited. *Ethnicity*, 7, 151-190.

Bernard, W. S. (1936). Cultural determinants of naturalization. *American Sociological Review*, 1 (6), 943-953.

Borjas, G. J. (1985). Assimilation, changes in cohort quality, and the earnings of immigrants. *Journal of Labor Economics*, 3 (4), 463-489.

Borjas, G. J. (1994). The economics of immigration. *Journal of Economic Literature*, 32 (4), 1667-1717.

Borjas, G. J. (2001). *Heaven's door: immigration policy and the American economy*. Princeton University Press, Princeton, NJ.

Bratsberg, B., & Barth, E., & Raaum, O. (2006). Local unemployment and the relative wages of immigrants: evidence from the current population surveys. *Review of Economics and Statistics*, 88 (2), 243-264.

Bratsberg, B., & Ragan, J. F., & Nasir, Z. M. (2002). The effect of naturalization on wage growth: a panel study of young male immigrants. *Journal of Labor Economics*, 20 (3), 568-597.
Carliner, G. (1980). Wages, earnings, and hours of first, second and third generation American males. *Economic Inquiry*, 18 (1), 87-102.
Chiswick, B. (1978). The effect of Americanization on the earnings of foreign-born men. *Journal of Political Economy*, 86 (5), 897-921.
Davidson, R., & MacKinnon, J. G. (1993). *Estimation and inference in econometrics*. Oxford University Press, New York, NY.
Heckman, J. J. (1978). Dummy endogenous variables in simultaneous equation system. *Econometrica*, 46 (4), 931-959.
Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press, Cambridge, MA.
Martin, J., & Mehlman, I. (2005). *The costs of illegal immigration to Flori*dians. Federation for American Immigration Reform, Washington, DC.
Mincer, J. (1974). *Schooling, experience and earnings*. Columbia University Press, New York, NY.
Rallu, J. L. (2011). Naturalization policies in France and the USA and their impact on migrants' characteristics and strategies. *Population Review*, 50, 40-61.
Street, A. (2013). Naturalization dynamics in immigrant families. *Comparative Migration Studies*, 1 (1), 23-44.
Thomas, B. (1954). *Migration and economic growth*. Cambridge University Press, Cambridge, England.
van Hook, J., Brown, S. K., & Bean, F. D. (2006). For love or money? Welfare reform and immigrant naturalization. *Social Forces*, 85 (2), 643-666.
Woodrow-Lafield, K. A., Xu, X., Kersen, T., & Poch, B. (2004). Naturalization of U.S. immigrants: highlights from ten countries. *Population Research and Policy Review*, 23 (3), 187-218.
Yang, P. Q. (1994). Explaining immigrant naturalization. *International Migration Review*, 28 (3), 449-477.