Economics Bulletin

Volume 35, Issue 4

An empirical analysis of the role of immigrant networks on legalization in the United States

Gihoon Hong
Indiana University South Bend

Soyoung Lee
Indiana Wesleyan University

Abstract

This paper examines the importance of immigrant networks on the likelihood of obtaining a U.S. immigrant visa. In order to control for the endogeneity arising from the formation of migration networks, we exploit variation in historical settlement patterns of earlier immigrants. Using panel data from the Mexican Migration Project, we find that the size of immigrant network at the destination is strongly positively related to the propensity to apply for a U.S. visa. In addition, estimating a hazard model of visa approval reveals that the wait time for visa approval is inversely associated with the size of the visa applicant's migration network at the destination in the case of employment-based applications. The results indicate that in addition to improving its members' earning potentials directly in the labor market, migration networks confer non-pecuniary benefits attached to legal status.

We thank an anonymous referee for constructive comments and suggestions. All remaining errors are our own.
Citation: Gihoon Hong and Soyoung Lee, (2015) "An empirical analysis of the role of immigrant networks on legalization in the United States", *Economics Bulletin*, Volume 35, Issue 4, pages 2317-2325
Contact: Gihoon Hong - honggi@iusb.edu, Soyoung Lee - soyoung.lee@indwes.edu.
Submitted: October 17, 2014. **Published:** November 20, 2015.

1. Introduction

Migration networks, defined as the size of migrant population at the destination originating from the same source country, are considered to be important determinants of location decisions of new immigrants. Previous studies have identified two different routes as to how existing networks influence the value of a location choice for potential immigrants. First, an immigrant's network improves labor market outcomes at the destination in terms of higher wages or likelihood of employment (Edin et al., 2003; Hong, 2015; Ioannides and Loury, 2004; Munshi, 2003). The second is the non-pecuniary migration costs channel through which the networks help lower the costs of assimilation into the host country (Carrington et al., 1996). Although extensive research has been conducted focusing on these direct impacts of migration networks on immigration, little is known about the indirect influences migration networks may exert in the visa application process. Established immigrants at the destination can help potential migrants to overcome legal barriers to immigration through either family reunification or employment-based sponsorship. By increasing the likelihood of obtaining legal status, migration networks may confer numerous monetary/non-monetary benefits on immigrants, thereby affecting immigrants' location decisions as well as the duration of their stay at the new location. The issue is particularly important because of its potential implication on migrants' self-selection (McKenzie and Rapoport, 2007; Beine et al., 2008). In the context of the U.S.–Mexico immigration, for example, the likely stronger effects of migration networks on unskilled workers may help explain the observed pattern of migrant selection from Mexico in the United States.[1] Isolating the additional benefits migration networks offer through the legalization process may inform U.S. policy makers as to how to best reform the current immigration system to neutralize the negative-selection effects, if any, of migration networks.

This paper develops and estimates an empirical model of visa application decisions to investigate the role of immigrant networks in the visa application process in the United States. The model accounts for individual heterogeneity in terms of visa qualification. However, there are some important challenges that need to be addressed in estimation. First, an empirical examination of network effects in the visa process requires complete information on the visa category each application falls in. This is an important consideration because visa eligibility and admission criteria vary with visa categories. Moreover, to the extent that immigrants' location decisions are correlated with the likelihood of getting a sponsorship for the visa process, migration networks may be endogenous to the visa application decisions.[2] To circumvent these issues, we use large micro panel data on Mexican immigrants in the United State s combined with a novel instrumental variable for the size of network at the destination that exploits variation in the historical migration pattern from Mexico to the United States.

The estimation results provide the first empirical evidence of strong network effects on U.S. visa application decisions among Mexican workers regardless of visa categories. In addition, we find that the size of a migrant's network is inversely related with the length of waiting for employer-based visa applications, indicating the presence of network effects in employment-based immigration. The results are found to be robust to the inclusion of unobserved heterogeneity and the use of an alternative instrument.

The rest of the paper is organized as follows. The next section describes data. Section 3 discusses the empirical strategy and presents the estimation results. We conclude in Section 4.

---

[1] Although there is little consensus in the literature on the skill composition of Mexican immigrants in the United States, Fernández-Huertas (2011) using recent household survey data shows that Mexican migrants in the United States are less skilled compared to nonmigrant Mexicans, providing evidence of negative selection.

[2] For instance, it is conceivable that potential immigrants in Mexico base their decision to move on the prospect of obtaining an immigration document, thereby being *attracted* to U.S. destinations with better prospects of getting a visa sponsor.

2. Data

The data for this study comes from the Mexican Migration Project (MMP). [3] Each year since 1982, the MMP has randomly selected households in communities throughout Mexico and collected U.S. migration information as well as life histories of a sample of household heads. In particular, the MMP records the year a visa application is submitted and approved, which allows us to observe the duration of the legalization process. The most recent version of the project, MMP143, contains 23,645 households with 1,151,489 person-year observations. After excluding observations with a non-working age (age<20), born in the U.S., or missing information on educational attainment, we are left with 666,864 person-year observations.

In addition, in order to address the endogeneity issue associated with the size of immigrant networks in the visa process, we adopt an instrumental variable from the recent literature. Examining the impact of migration networks on the development of microenterprises in Mexico, Woodruff and Zenteno (2007) propose the use of geographical distance from the capital in each Mexican state to the nearest station on the north/south rail lines in order to proxy for the size of network. Because these rail lines were intensively used for transportation with the implementation of the guest worker program during the 1910s, it is believed that the proximity to the rail lines was strongly correlated with the geographic pattern of the massive migration flows from Mexico. To the degree that the historical migration pattern has been persistent through the self-enforcing network effects, the distance data provide plausibly exogenous variation that identifies the effect of the network (Demirgüç-Kunt et al., 2011).

Table I: Descriptive Statistics

	Mean	Std. Dev.
Panel A : Sample of 23,192 individuals		
Educational attainment (in years)	6.04	4.55
Years in the sample	28.75	15.24
Never been to the U.S.	0.68	0.46
Distance to rail line (in kilometers)	144.86	308.49
Panel B : Sample of 7,302 individuals with U.S. experience		
Educational attainment (in years)	5.47	3.99
Distance to rail line (in kilometers)	78.43	239.71
Years in the U.S.	6.14	7.71
Applied for a U.S. visa	0.08	0.28
Panel C : Sample of U.S. 629 visa applicants		
Educational attainment (in years)	6.03	4.02
Distance to rail line (in kilometers)	63.58	216.73
Family-based applicant	0.273	0.446
Employment-based applicant	0.182	0.386
Application processing time (in years)	3.034	3.755

Table I presents summary statistics for the MMP sample. A brief comparison between the total sample and the U.S. migrant sample reveals that a typical Mexican worker with U.S.

[3] The MMP is publicly available at http://mmp.opr.princeton.edu. For more information on the MMP, see Massey and Espinosa (1997).

migration experience, on average, are less-educated and were born from a state nearer to the rail lines compared to the general Mexican population. 8.6% of the U.S. migrants are shown to have applied for a U.S. visa. Panel C reports that family-based and employment-based visa applicants comprise 27% and 18% of the total applications, respectively; it takes about 3 years, on average, for a case to be adjudicated. The standard deviation, however, suggests that there is a large variation in the visa processing time.

3. Empirical Specification and Results

The empirical analysis involves two separate estimation exercises. First, we use the logit model to explore the impact of networks on the decision to apply for a U.S. immigration document. Next, the Cox proportional hazard model is employed to analyze how immigrant networks may affect U.S. visa wait times.

3.1 Logistic Regression of Visa Application Decision

A potential migrant may be eligible for an immigration document if he has someone in the United States who may agree to be a sponsor as an employer or as a close relative who is either a U.S. citizen or permanent resident. Although eligibility for legal immigration is not observed in the data, we can still address heterogeneity in eligibility considering the propensity to apply for an immigration document as a function of observable person-specific qualification measures. In order to study the importance of a migrant's network in the visa application decision, we start with the following generic logit model:

$$\text{Prob}(Y_i = 1|X_{1i}) = \frac{\exp(X_{1i}\beta)}{1 + \exp(X_{1i}\beta)}, \quad (1)$$

where Y_i is a dummy variable that takes a value of one if an individual i applies for an immigration document and a value of zero otherwise; X_{1i} is a vector of time-constant explanatory variables that include educational attainment and a dummy that indicates the existence of family in the United States as well as the network instrument[4]; β is a vector of the parameters of interest. Then information on visa application decisions in the MMP allows us to construct the choice probabilities. The parameters of interest in β are estimated using the Maximum likelihood Method. Additionally, we employ the data on the type of sponsor (employment-based vs. family-based) in the MMP and explore the determinants of the visa application decision for different immigration categories.

3.2 Cox Proportional Hazard Model of Visa Wait Time

The second estimation exercise concerns exploring the effects of immigrant networks on the wait time for the visa applicants. Having data on the time-to-approval for each applicant, a natural approach to capture the network effects on the wait time is to apply survival analysis. Although the wait time varies greatly depending on the category (employment-based or family-based) the application falls under, the MMP reports the type of sponsor for each visa application. Therefore, we choose to estimate the hazard model separately for different sponsor categories. One important consideration in estimating visa wait time is that the application process for immigration documents within the United States, called "Adjustment of Status (AOS)," is generally different from the process outside of the country, called "Consular Processing (CP)."

[4] Although our main dataset is considered a panel, we cannot use all the person-year observations because the network instrument is time-constant and it cannot be separately identified from fixed effects. Therefore, all the variables are collapsed using the mean values of each variable and defined at the individual level. This implies that time-varying covariates such as age, work experience and marital status are not included in the regression.

Given the small number of the CP cases in the MMP,[5] we choose to focus on the visa applicants within the United States in the estimation. Then, considering the visa wait time as a random variable W that is determined by a discrete hazard function, we propose the following hazard model based on Cox (1972):

$$\rho(w|X_{2i}) = \Pr(W = w|X_{2i}, W \geq w) = \lambda(w)\exp(X_{2i}\beta), \qquad (2)$$

Where ρ is the hazard rate of exiting the waiting spell upon application approval at period w given that the application has not been approved for w-1 periods; $\lambda(w)$ is the baseline hazard; X_{2i} includes, in addition to the time-constant variables in X_{1i} in (1), time-varying covariates that may affect the length of the waiting period such as age, marital status, and work experience in Mexico.[6] In addition, year fixed effects are included to control for aggregate fluctuation in visa availability. The Cox Proportional Hazard model presented in (2) is well suited in our study because it accounts for right-censored observations (i.e., visa applications observed to be pending throughout the sample periods). Moreover, the model allows for a set of time-varying covariates in X_{2i} including the size of network without having to rely on restrictive distributional assumptions on $\lambda(w)$ for identification.

Although the semi-parametric approach described above is appealing because it eliminates the need to impose a strong parametric assumption on the baseline hazard, one concern arises from unobserved heterogeneity. As Lancaster (1990) points out, when person-specific unobserved effects are present in the data, estimating the baseline model in (2) would understate the baseline hazard rate. Intuitively, ignoring unobserved heterogeneity would cause the esimated *population* baseline hazard to decline with duration even when the true *individual* hazard rate is constant, because individuals with lower baseline hazard rates are expected to be overrepresented in the sample as duration increases. To guard against the issue of unobserved heterogeneity, we also consider a parametric hazard model with gamma distributed heterogeneity (Lancaster, 1990; Meyer, 1990) and compare the results with the baseline results.

Table II: Network Effects on Visa Application Decisions (Logit Model)

	[1] All Applicants	[2] Family-based	[3] Employment-based
Educational attainment (in years)	0.0177	0.0488**	0.0462*
	(0.0118)	(0.0204)	(0.0239)
Distance to Rail lines (in kilometers)	-0.0011***	-0.0006*	-0.0007*
	(0.0002)	(0.0004)	(0.0004)
Family in the U.S.	1.3803***	1.9085***	0.2659
	(0.1203)	(0.1708)	(0.2812)
Constant	-1.7235***	-3.6238***	-3.6422***
	(0.0887)	(0.1704)	(0.1901)
Number of observations	3514	3514	3514
Log likelihood	-1564.89	-629.34	-502.26

Notes: Standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

[5] 851 out of 1,059 applicants (80.3%) observed in the MMP submitted the visa application within the U.S.

[6] In order to capture possible nonlinearities, the squared values of age and work experience as well as their interaction term are also included in the regression.

3.3 The Effects of Immigrant Networks on Visa application Process

The maximum-likelihood estimates of the logit model in (1) are presented in Table II. Overall, the estimated coefficients are shown to have expected signs, indicating positive impacts of educational attainment and existence of family in the United States on U.S. visa application decisions. More importantly, the coefficient estimates of the network instrument, *Distance to Rail lines*, are negative and statistically significant, regardless of the category the visa application falls in. This provides the first evidence of positive network effects on visa application decisions, suggesting that immigrants are more likely to apply for an immigrant visa as the size of their immigrant network increases.[7] For example, the estimated coefficient associated with *Distance to Rail lines* in Column [1] shows that, holding all other variables constant, the estimated odds in favor of applying for a potential migrant who was born in a Mexican state in which the rail line passed through is 3.00 ($= \exp(-1000 \cdot -0.0011)$) times greater than the odds of applying for a Mexican worker who is from a state which was 1,000 kilometers away from the rail line.

Table III: Network Effects on Visa Wait Time (Cox Proportional Hazard Model)

	[1] All Applicants	[2] Family-based	[3] Employment-based
Educational attainment (in years)	-0.0333***	-0.1422***	-0.0147
	(0.0108)	(0.0319)	(0.0322)
Distance to Rail lines (in kilometers)	0.0001	0.0008	-0.0025***
	(0.0002)	(0.0010)	(0.0007)
Family in the U.S.	-0.2116*	0.0226	-0.3168
	(0.1106)	(0.3261)	(0.4229)
Age	0.0228	0.1246	0.2256**
	(0.0334)	(0.1014)	(0.0965)
Age sqaured	0.0001	-0.0013	-0.0022
	(0.0005)	(0.0013)	(0.0015)
Work experience in Mexico	-0.0140	-0.0779	-0.0767
	(0.0245)	(0.0480)	(0.0865)
Work experience in Mexico squared	0.0031***	0.0054***	0.0005
	(0.0007)	(0.0015)	(0.0036)
Marital Status	-0.0026***	-0.0035**	0.0004
	(0.0009)	(0.0017)	(0.0042)
*Age * Work experience in Mexico*	-0.0140	-0.0779	-0.0767
	(0.0245)	(0.0480)	(0.0865)
Year fixed effects	Yes	Yes	Yes
Number of subjects	742	162	144
Time at risk	2642	812	512
Log likelihood	-4243.95	-651.45	-568.48

Notes: Standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

[7] Because the instrument measures distances from the capital of each state to the nearest railroad, a negative sign of the estimated coefficient indicates positive network effects.

Having identified the effects of immigrant networks on visa application decisions, we examine whether the size of networks affects the wait time, as defined to be the time to take until the visa application is adjudicated. Table III presents results from estimating the proportional hazard model in (2). We find the estimated coefficient of *Distance to Rail lines* to be negative and statistically significant only for the employment-based visa applications. This indicates the size of immigrant network, as proxied by the inverse of the distance between the state of birth and the rail lines, is positively correlated with the hazard rate of exiting the waiting spell upon application approval only in the case of employment-based application. Despite the unobserved nature of how the network functions in the labor market, it seems plausible that the same applicant is more likely to have personal connections with the sponsoring employer when (s)he comes from an exogenously larger network, which can affect the application process favorably and possibly shorten the wait time for visa approval. Interestingly, network effects do not seem to be present in determining wait times for other visa application categories including family-based applications. Besides network effects, we find that *Age* is also positively associated with the hazard rate of visa approval for employment-based visa applicants, which demonstrates the importance of applicants' general human capital in the employment-based immigration process.

Table IV: Alternative Instrument for the Size of Immigrant Network (Historical State Migration Rate)

	[1] All Applicants	[2] Family-based	[3] Employment-based
Educational attainment (in years)	0.0266**	0.0632***	0.0464*
	(0.0119)	(0.0204)	(0.0241)
Historical state migration rate in 1924	4.3910***	5.7585***	1.4418
	(0.7047)	(1.2971)	(1.4424)
Family in the U.S.	1.3356***	1.8090***	0.2760
	(0.1211)	(0.1723)	(0.2836)
Constant	-2.2812***	-4.3201***	-3.8437***
	(0.1196)	(0.2306)	(0.2429)
Number of observations	3514	3514	3514
Log likelihood	-1562.69	-502.26	-503.60

Notes: Standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Overall, the empirical exercise confirms the significant role of immigrant networks in the visa application process. However, it is possible that the results are sensitive to the choice of the network instrument. In order to ensure robustness of the main findings, we estimate the logit model in (1) using an alternative instrument variable for the size of network, exploiting variation in historical state-specific migration rates from 1924.[8] Although one may argue that the historical migration pattern may still be endogenous to current migration decisions due to unobserved state effects, we expect it to be strongly correlated with the size of migration network and, therefore, allows us to check the validity of the main instrument. Table IV shows that the results are qualitatively similar to the original logit estimates in Table II, in that the

[8] The large scale Mexican migration to the U.S. began with the implementation of the guest worker program during the 1910s and 1920s (Woodruff and Zenteno, 2007). Data on the 1920's state migration rates is available at http://www2.warwick.ac.uk/fac/soc/economics/staff/academic/woodruff/data/mexico_migration/.

odds ratio of applying for visa increases with the size of immigrant network except for the employment based visa applicants.

Finally, despite the flexibility which does not require any strong functional form for the baseline hazard, the Cox Proportional Hazard model used to analyze the determinants of visa wait time is known to suffer from biases in the presence of unobserved heterogeneity. To address the issue, we follow Lancaster (1990) and estimate a parametric Mixed Proportional Hazard (MPH) model to account for unobserved person-specific random effects that may influence the baseline hazards. In particular, we consider a Weibull baseline hazard with gamma-distributed random effects for its computational and expositional convenience (Lancaster, 1990; Abbring and Van Den Berg, 2007). Table V shows that age as well as the size of migration network are positively associated with the hazard rate of visa approval only in the case of employment-based visa applicants, consistent with the main findings in Table III.

Table V: Parametric (Weibull) Hazard Model with Gamma-distributed Random Effects

	[1] All Applicants	[2] Family-based	[3] Employment-based
Educational attainment (in years)	-0.0013	-0.1840***	-0.0345
	(0.0234)	(0.0328)	(0.0353)
Distance to Rail lines (in kilometers)	0.0002	0.0015	-0.0042***
	(0.0004)	(0.0010)	(0.0008)
Family in the U.S.	0.0894	0.2137	-0.4880
	(0.2024)	(0.3558)	(0.4552)
Age	0.0797	0.1342	0.4119***
	(0.0614)	(0.1181)	(0.1033)
Age^2	0.0002	-0.0012	-0.0043***
	(0.0009)	(0.0015)	(0.0016)
Work experience in Mexico	0.0073	-0.1266**	-0.1454
	(0.0437)	(0.0498)	(0.0920)
Work experience in Mexico squared	0.0078***	0.0079***	0.0003
	(0.0015)	(0.0015)	(0.0040)
Marital Status	0.7720***	0.5588**	-0.1590
	(0.1809)	(0.2796)	(0.2900)
*Age * Work experience in Mexico*	-0.0073***	-0.0049***	0.0013
	(0.0017)	(0.0017)	(0.0046)
Year fixed effects	Yes	Yes	Yes
Individual-specific random effects	Yes	Yes	Yes
Number of subjects	742	162	144
Time at risk	2642	812	512
Log likelihood	-693.89	-125.70	-86.23

Notes: Standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

4. Conclusion

In this paper, we have developed a simple model of legalization in the United States which incorporates migration networks at the U.S. destination as a determinant of visa application decisions and estimated the model parameters with retrospective panel data from the Mexican Migration Project. After accounting for the endogeneity issue regarding the formation of migration networks, we find that the size of immigrant network at the destination is strongly positively related to the propensity to apply for a U.S. visa. In addition, estimating a hazard model of visa approval reveals that the wait time for visa approval is inversely associated with the size of the visa applicant's migration network at the destination in the case of employment-based applications. The results indicate that in addition to improving its members' earning potentials directly in the labor market, migration networks confer non-pecuniary benefits attached to legal status. To the best of our knowledge, this is the first study that identifies the indirect channel through which immigrant networks benefit potential migrants.

References

Abbring, H. and Van Den Berg, J. G. (2007), "The unobserved heterogeneity distribution in duration analysis," *Biometrika*, Vol. 94(1), 87-99.

Beine, M., Docquier, F., and Ozden, C. (2008), "Diasporas", paper presented at the Conference on Migration and Development, June, Lille.

Borjas, G. J. (1999), *Heaven's Door: Immigration Policy and the American Economy*, Princeton University Press, Princeton, NJ.

Carrington, W. J., E. Detragiache, and T. Vishwanath (1996), "Migration with Endogenous Moving Costs," *American Economic Review*, 86(4), 909-930.

Cox, D. R. (1972), "Regression models and life tables (with discussion)," *Journal of the Royal Statistical Society*, 34, 187-220.

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

Chiswick, B. (1999), "Are Immigrants Favorably Self-Selected?" *American Economic Review* 89, 181–85.

Demirgüç-Kunt, A., E. Córdova, M. Pería, and C. Woodruff. 2011. "Remittances and Banking Sector Breadth and Depth. Evidence from Mexico." *Journal of Development Economics*, 95(2), 229–241.

Edin, P.-A., P. Fredriksson, and O. Slund (2003), "Ethnic Enclaves And The Economic Success Of Immigrants - Evidence From A Natural Experiment," *The Quarterly Journal of Economics*, 118(1), 329–357.

Fernández-Huertas Moraga, J. (2011), "New Evidence on Emigrant Selection," *The Review of Economics and Statistics*, 93(1), 72-96.

Hong, G. (2015), “Examining the U.S. Labor Market Performance of Immigrant Workers in the Presence of Network Effects,” *Journal of Labor Research*, 36(1), 9-26.

Ioannides, Y. M., and L. D. Loury (2004), “Job Information Networks, Neighborhood Effects, and Inequality,” *Journal of Economic Literature*, 42(4), 1056–1093.

Lancaster, T. (1990), “The Econometric Analysis of Transition Data,” Cambridge, Cambridge University Press.

McKenzie, D. and H. Rapoport (2010), "Self-Selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks," *The Review of Economics and Statistics*, 92(4), 811-821, November.

Massey, D., and K. Espinosa (1997), “What’s Driving Mexico-U.S. Migration? A Theoretical,
Empirical, and Policy Analysis,” *American Journal of Sociology*, 102, 939-999.

Meyer, B. D. (1990), “Unemployment Insurance and Unemployment Spells,” *Econometrica*, 58, 757-782.

Munshi, K. (2003), “Networks in the Modern Economy: Mexican Migrants in the US Labor Market,” *Quarterly Journal of Economics*, 118, 549–597.

Woodruff, C., and R. Zenteno (2007), “Migration Networks and Microenterprises in Mexico,” *Journal of Development Economics*, 82, 509–528.