Why Are Some Immigrant Groups More Successful Than Others?

Edward P. Lazear, *Stanford University*

The composition of immigrants depends not only on immigrant choice but also on immigration policy, because slots are rationed. Policy determines immigrant attainment, as evidenced by immigrants from Algeria having higher educational attainment than those from Israel or Japan. Theory predicts and evidence confirms that immigrant attainment is inversely related to the number admitted from a source country and positively related to population and education levels at home. A parsimonious specification has only two variables yet explains a majority of the variation in educational attainment of US immigrant groups. The theory and predictions are bolstered by Swedish data.

I. Introduction

Algeria, Israel, and Japan, along with more than 100 other countries, are source countries of migrants to the United States. Average education levels in Algeria are well below those in Israel and Japan, yet US immigrants from Algeria have higher levels of educational attainment than those from Israel

I am grateful to Lars Oxelheim and Charlotta Olofsson, who provided the data from Sweden, and to Melissa Kearney for discussant comments at the Utah Business Conference. My thanks also to Caroline Hoxby for extensive and thoughtful comments and to Juan Saavedra for help with the International Public Use Microdata Series data. Additionally, thanks to Enrico Moretti, Robert Pollak, Kathryn Shaw, and others at the National Bureau of Economic Research (NBER) Summer Institute for helpful comments. This paper is a shortened version of NBER Working Paper no. 23548 (https://www.nber.org/papers/w23548). Contact the author at lazear@stanford.edu. Information concerning access to the data used in this paper is available as supplemental material online.

[*Journal of Labor Economics*, 2021, vol. 39, no. 1]
 0734-306X/2021/3901-0004$10.00
Submitted March 12, 2019; Accepted October 15, 2019; Electronically published October 15, 2020

or Japan. Also true is that those from Algeria are a much smaller share of the immigrant population than those from Israel or Japan.

The attainment of immigrants in the United States varies greatly by country of origin. The more underrepresented a source country, the higher is the attainment of immigrants from that country. Two of the largest source countries, Mexico and India, provide cases in point. Migrants to the United States from Mexico have a mean educational attainment of 9 years, while those from India have a mean of 16 years. Yet average educational attainment in Mexico is almost twice that of average educational attainment in India. Mexico is overrepresented among US immigrants relative to its world population by a factor of 15, while India is highly underrepresented.

Immigrant selection implicit in the policies of recipient countries is mostly ignored in the earlier literature, which focuses on choices made by migrants.[1] Between 2009 and 2014, there were many times as many applicants to the United States as the number granted permanent residency, which implies that immigrant composition depends not only on those who are willing to come but also on how policy allocates visas[2] among a heterogeneous pool of applicants.[3]

Here, an alternative (albeit extreme) approach is adopted. A desirable destination country is assumed to choose, explicitly or implicitly, the number of immigrants admitted from each of the world's source countries. Selection is from the top of that country's attainment distribution down, either because the destination country uses that criterion of admission or because those with higher levels of attainment are more likely to migrate.[4] As a consequence, the education distribution in the origin country, the country's

[1] Some earlier literature that focuses on policy and demand by recipient countries include Clark, Hatton, and Williamson (2007), who look at the effect of quotas and refugee policy; Mayda (2010), who examines migration to 14 Organization for Economic Cooperation and Development (OECD) countries; and Ortega and Peri (2013), whose data include time-varying entry regulations.

[2] See US Department of State (2015) and US Department of Homeland Security (2016). In 2015–17, about 1.1 million per year obtained lawful permanent residence status in the United States (https://www.dhs.gov/immigration-statistics/yearbook/2017/table1). In 2017, the Pew Foundation reports that there were 22.4 million applications for entry through the diversity lottery, which allocates 50,000 slots per year (Connor 2018). Additionally, the US Department of State (2018) reports that 3.792 million were on the family- or employer-based wait list in 2018.

[3] Among those entering the United States in the last 2 years of the survey date, the standard deviation of educational attainment was 4.9 years, and that of age was 15.4 years.

[4] Selection can be positive even on the supply side if the costs of migration are less than proportional to earnings. Borjas and Friedberg (2009) argue that selection rules explain the rise in wages of immigrants that occurred between 1990 and 2000. Kerr and Lincoln (2010) examine the effect of the rise in H-1B visas on innovation. Kerr, Kerr, and Lincoln (2015) find that skilled immigrant labor is an important

population, and, importantly, the number of immigrants admitted from that country determine the composition and attainment of immigrants.

The focus is the United States because the slot rationing applies to the United States better than to most countries. Data from Sweden, another desirable destination country, support the hypotheses. The model, which implies that attainment should depend only on the number of immigrants admitted, the population, and the average attainment in the origin country, explains a large fraction of immigrant educational attainment and economic success in the United States.

II. Theoretical Predictions

Figure 1 illustrates the main argument. Consider two countries, 1 and 2, with equal population sizes and with educational distributions as shown. Country 2's distribution is a rightward displacement of country 1's distribution. Suppose that the United States were to decide that 3% of immigrants will come from country 1 but that 30% will come from country 2, and suppose further that all who are offered US residency accept. If the United States also allows the most educated immigrants in first from each country or, alternatively, if the most educated in each country are most attracted to the United States or most able to navigate their way through the immigration process, then the upper tail of each distribution will migrate to the United States. In this example, because the United States targets 3% from country 1 and 30% from country 2, the minimum cutoff level of education in each country is A_1^* and A_2^*, respectively, which yields average levels of education of A_1 and A_2. The educational attainment of immigrants from country 1 exceeds that from country 2 even though country 1's education level at home is below that of country 2 at home.

More formally, a destination country, here the United States, allows I_i to enter from country i and admits the top of the attainment or ability distribution first. Let N_i be the population of country i, and let $F_i(A)$ be

determinant of the overall increase in the skilled workforce of US firms. H-1Bs are important because the H-1B program is always oversubscribed. Two Hunt (2011, 2015) papers examine the performance of highly skilled immigrants in the United States. Another measure of ability to perform in the United States relates to English fluency, which is explored in Lazear (1999). Lewis (2013) discusses the ability of immigrants to substitute for native-born US labor and how that relates to language skills. Peri, Shih, and Sparber (2015) estimate the effect of STEM immigration on productivity and find that it is substantial, using city differences. None of these papers speaks to the empirical validity of the assumption that selection is from the top of an origin country's distribution, but the fact that many of the origin countries have low educational levels implies that the high-skilled ones are selected from the top.

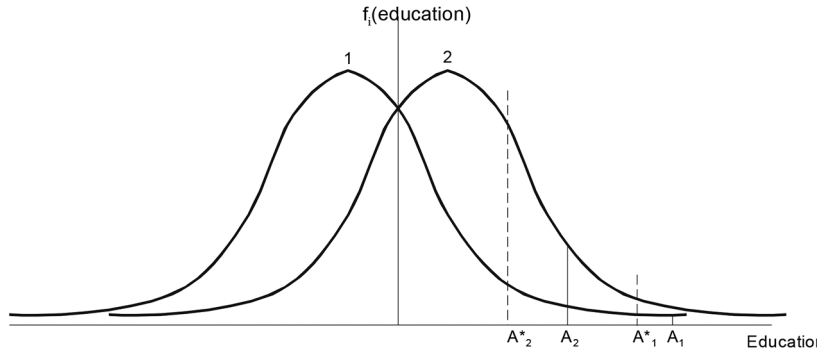

FIG. 1.—Illustration of the main argument.

the distribution of education or some other measure of ability or attainment, A, in country i.

Given I_i, A_i^* is the cutoff ability level of immigrants from country i determined such that

$$N_i\left[1 - F_i\left(A_i^*\right)\right] = I_i. \tag{1}$$

The expected level of attainment among those from country i in the United States is simply the conditional expectation, or

$$\bar{A}_i = \frac{1}{1 - F_i\left(A_i^*\right)} \int_{A_i^*}^{\infty} A f_i(A)\, dA. \tag{2}$$

For a given distribution F, the higher is the cutoff, A^*, the higher is the conditional expectation. Three predictions follow directly, which are stated formally and proved in an earlier version of this paper (see Lazear 2017).

First, the larger is the number of immigrants admitted from country i, the lower is their expected level of attainment, $\bar{A}_i$, because this pushes A^* down. When a country is permitted to send only a small number of immigrants and when selection is from the top, those who enter the United States will be highly talented. Second, for any given number of immigrants, I_i, the larger is the population in country i, the higher is the expected level of attainment of immigrants $\bar{A}_i$ from that country. It is easier to fill a quota of any size when drawing from a large population than when drawing from a small population. As a consequence, the cutoff shifts up with the population of the source country, raising the conditional expectation of attainment. Consider selecting the most highly educated 1 million people from a tiny country like Laos with 7 million people versus 1 million from India with 1.3 billion. If the distribution of underlying ability were the same in both countries, A^* for India would be higher than A^* for Laos because 1 million people comprise a much smaller fraction of the top tail of the distribution in India than

in Laos. Third, when the distribution of attainment in the source country is displaced positively, the conditional expectation rises by exactly the same amount.

Summarizing, $\partial \bar{A}_i/\partial I_i < 0$, $\partial \bar{A}_i/\partial N_i > 0$, and $\partial \bar{A}_i/\partial \mu_i > 0$, with $\partial \bar{A}_i/\partial \mu_i = 1$ when all sources countries have the same underlying distribution of attainment up to a shift in the mean.

It is also possible to express the predictions in terms of the "representation ratio," defined as

$$R_i \equiv \frac{I_i/\sum I_i}{N_i/\sum N_i}.$$

When R_i equals 1, immigrants are represented in the United States in accordance with world population proportions. Then, given a fixed number of immigrants, attainment of immigrants falls in the representation ratio, or $\partial \bar{A}_i/\partial R_i < 0$.

These predictions are tested below using data sets from two countries that have an excess supply of those wanting to migrate, namely, the United States and Sweden. Without exception, all predictions are borne out for three measures of attainment—education, wages, and earnings. The predictions are valid in both the United States and Sweden. Additionally, the number of immigrants, the population of the origin country, and the mean attainment in the origin country explain between 53% and 68% (depending on specification) of the variation in immigrant educational attainment in the United States and up to 60% in Sweden.

III. Data

The American Community Survey (ACS) from years 2011 through 2015, a series of five consecutive cross-sectional data sets, is the primary data source. By combining years, a larger sample is created so that more precision can be obtained. It is straightforward to adjust standard errors for population weights and to correct wage data for inflation to turn nominal wages and earnings into real values. Given the individual data, the means of education, wages, and annual earnings can be computed by country of origin. Most of the results reported below use the unweighted specifications. Weighting improves the fit but does not fundamentally change the results.

The independent variables are either available directly with minor transformation from the ACS (or Swedish registry data below) or drawn from other standard data sets. United Nations reports provide information on average education within the origin country, and population statistics are taken from World Bank data. The term I_i in equation (1) is defined as the actual number of immigrants from country i, measured in millions as estimated by scaling up the ACS. Similarly, N_i in equation (1) is defined as the

Table 1
Data Description

Variable	Mean	SD	Min	Max
United States (based on 129 origin countries):				
Education among immigrants from country i	13.409	1.563	8.864	16.003
Hourly wage among immigrants from country i	31.959	9.602	14.024	66.026
Annual earnings among immigrants from country i	35,833	11,154	17,630	69,497
Log of average wage	3.421	.297	2.641	4.19
Log of average earnings	10.441	.3	9.777	11.149
Number of immigrants from country in millions (I_i)	.239	.767	.002	8.187
Population of origin countries (N_i)	.545	1.737	.001	14.042
Representation ratio (R_i)	5.64	11.65	.04	69.74
Log(I_i)	−2.624	1.416	−6.379	2.103
Log(N_i)	−2.147	1.922	−7.475	2.642
Average education in origin country (μ_i)	8.611	2.796	1.6	12.9
Log(R_i)	.37	1.641	−3.33	4.245
Following based on 69 observations:				
GDP per capita in 1,000s of PPP dollars	6.9766	6.589	.051	30.828
Log of GDP per capita	1.434	1.201	−2.972	3.428
Five-year average GDP growth rate (%)	3.774	3.437	−4.943	10.446
Percentage of GDP from agriculture	6.97	6.58	.051	30.82
90/10 wage ratio in origin country (based on 31 OECD countries)	3.516	.938	2.280	5.71
Sweden (based on 79 origin countries):				
Proportion of immigrants from country i with advanced degree	.316	.129	.059	.597
Number of immigrants from country i in 100,000 (I_i)	.1307	.1904	.0050	1.0119
ln(I_i)	−2.7909	1.2188	−5.2887	.0119
Representation ratio (R_i)	8.27	18.64	.05	109.03
ln(R_i)	.606	1.759	−2.934	4.692

SOURCE.—American Community Survey for I_i, R_i, N_i, education, wages, and earnings of US immigrants; United Nations Development Program Human Development Reports for information on average education within origin country (μ_i); World Bank database 2015 for origin country population, N_i, and share of gross domestic product (GDP) from agriculture; Organization for Economic Cooperation and Development (OECD) "Income Distribution," OECD Social and Welfare Statistics (database) for 90/10 wage ratio; Heston, Summers, and Aten (2012) for GDP per capita and GDP growth; Statistics Sweden for all variables specific to Sweden.

NOTE.—All means are computed by averaging the countries' averages on an unweighted basis. PPP = purchasing power parity.

population of country i in 100 millions. Sources for gross domestic product (GDP) per capita, share in agriculture, growth rates, proportion with tertiary education, the 90/10 wage ratio, and distance of the origin country from the United States are listed in the note to table 1, which reports the summary

statistics for variables used. Data from Sweden are used to corroborate findings and to distinguish the hypothesis from other explanations. Swedish data were obtained from the Swedish registry administrative data (Statistics Sweden), Registret över befolkningens utbildning.[5]

IV. Results

The main results for the United States are presented in table 2. There are 129 countries of origin with information necessary for the analysis. The 129 observations reflect the immigrants from most of the countries that are sources of immigrants in the United States. An observation consists of averages for each variable across all immigrants from a particular origin country.

Columns 1–3 of table 2 test the predictions that are derived from equations (1) and (2) using three different dependent variables, namely, average educational attainment, average wages, and earnings.

All predictions are supported by the results for each of the dependent variables. The average educational attainment of immigrants from country i decreases in the number of immigrants from the origin country, $\log(I_i)$; increases in the population of the origin country, $\log(N_i)$; and increases in mean education of the origin country, μ_i. The same is true for dependent variables log of average wage and log of average annual earnings, as shown in columns 2 and 3. Furthermore, the three-variable model explains 53% of the variation in immigrant educational attainment and 68% of the variation when weighted by the size of the immigrant group in column 7. The three variables also explain 53% of the variation in wages by group and 29% of group earnings.

A 1 standard deviation increase in the log of a country's number of immigrants decreases the predicted level of educational attainment in the United States by about .9 years on a mean level of education of 13.4 years. The predicted difference in education among immigrants from the highest immigrant provider, Mexico, and the lowest immigrant provider, Estonia, is 5.5 years, compared with an actual difference of 5.6 years.

A 1 standard deviation increase in the population of the origin country implies about a full year increase in the education levels of the immigrants

[5] The Swedish data set contains statistics for 98 countries (including Sweden) between the years 2011 and 2015. The number of observations included for each year are as follows: 5,472,582 (2011); 5,499,465 (2012); 5,528,680 (2013); 5,568,916 (2014); and 5,605,685 (2015). Immigrants account for about 19% of the sample. The earnings and transfer values refer to individuals between the ages of 20 and 64, while the education values refer to individuals between the ages of 25 and 64. All variables come from register data (administrative sources) by Statistics Sweden and therefore cover all individuals who were registered in Sweden each particular year. All variables associated with education level come from Statistics Sweden's Registret över befolkningens utbildning, or "Register of Population Education."

Table 2
Educational Attainment, Wages, and Earnings

	Immigrants from Country *i*							
Variable	Average Education (1)	Log of Average Hourly Wage (2)	Log of Average Annual Earnings (3)	Average Education (4)	Log of Average Hourly Wage (5)	Log of Average Annual Earnings (6)	Average Education (7)	Average Education (8)
Log of number of immigrants from country in millions ($\ln(I_i)$)	−.643***	−.0677***	−.0687***				−1.374***	−.628***
	(.0811)	(.0155)	(.0193)				(.0869)	(.0825)
Log of population of origin country in 100 millions ($\ln(N_i)$)	.545***	.0892***	.0842***				.946***	.519***
	(.0609)	(.0116)	(.0145)				(.0888)	(.0586)
Average educational attainment in origin country (μ_i)	.304***	.0696***	.0430***	.308***	.0688***	.0424***	.275***	.314***
	(.0349)	(.00666)	(.00829)	(.0349)	(.00669)	(.00826)	(.0550)	(.0332)
Log of representation ratio ($\ln(R_i)$)				−.565***	−.0850***	−.0811***		
				(.0595)	(.0114)	(.0141)		
Constant	10.27***	2.835***	10.07***	10.97***	2.860***	10.11***	10.14***	10.16***
	(.355)	(.0678)	(.0843)	(.312)	(.0598)	(.0738)	(.468)	(.357)
Observations	129	129	129	129	129	129	129	125
R^2	.534	.530	.289	.527	.520	.284	.680	.560
Notes							Weighted by number of immigrants from origin country	Excludes Mexico, Philippines, China, India

NOTE.—The unit of analysis is an origin country. Variables reflect means of the immigrants from a particular origin country, taken from the American Community Survey (ACS) 2011–15. There are 134 countries reflected in the ACS, but only 129 have the complete data necessary for the analysis. Standard errors are in parentheses.

*** $p < .01$.

from that country. Compare a tiny country like Cape Verde, with a half million people, to a large country like Nigeria, with almost 200 million people. Both have similar average levels of education, but in accordance with the prediction on the effect of population on educational attainment, the average level of education among immigrants in the United States from Cape Verde is 9.8 years, compared with more than 15 years for those from Nigeria.

The more obvious prediction of a positive effect of education level in the origin country on attained level of education among immigrants is also borne out for all three dependent variables.

Equations (1) and (2) imply a particular homogeneity. Specifically, the coefficients on $\ln(I_i)$ and $\ln(N_i)$ should be equal but of opposite signs. This is true for the coefficients using all three dependent variables, as shown in columns 1–3. For example, in column 1 the coefficient on $\ln(I_i)$ is $-.643$ and that on $\ln(N_i)$ is .545. Each coefficient is estimated rather precisely, but the difference between their absolute values is not statistically significant.

Recall that the predictions can be expressed more parsimoniously and perhaps more intuitively in terms of the representation ratio, R_i. Columns 4–6 of table 2 test those predictions, which are verified for all three dependent variables. The more overrepresented is an origin country, the lower are education attainment, wages, and earnings of immigrants from that country. Little explanatory power is lost by using $\ln(R_i)$ in lieu of $\ln(I_i)$ and $\ln(N_i)$ together, another reflection of the coefficients on $\ln(I_i)$ and $\ln(N_i)$ being equal of opposite signs.

Consistent with the negative coefficient on $\ln(R_i)$ is the important example of India. India is significantly underrepresented among US immigrants relative to its population, even though it is the third largest supplier of immigrants. Immigrants from India have the second-highest level of educational attainment among all origin countries despite having one of the lowest levels of educational attainment in the home country. The group with the highest level of attainment are those from the former USSR, who came in the 1980s primarily because the USSR let them out and the United States let them in, mostly under special circumstances. They are also among the most underrepresented group.

Additional specifications are tested with education as the dependent variable in columns 7 and 8.[6] Column 7, as mentioned above, weights each observation by the number of immigrant individuals used to compute the means of the dependent variables for each country. The results are qualitatively similar, but the magnitudes of the coefficients on $\ln(I_i)$ and $\ln(N_i)$ rise. Additionally, the R-squared rises substantially to .68. The unweighted analysis in column 1 does not give disproportionate weight to any one country. Still, it is useful to see how the results change when the largest origin

[6] The NBER working paper version of this study includes results for the other two dependent variables. None of the conclusions is altered.

countries in the analysis—namely, Mexico, Philippines, China, and India—are excluded. Mexico and India are outliers in that Mexico has the third-lowest level of immigrant educational attainment while India has the second-highest level and Mexico is overrepresented while India is underrepresented among US immigrants. Not surprisingly, the results do not change much from the unweighted ones presented in column 1. Dropping Mexico alone gives results very similar to those in column 8. The results imply a general pattern among immigrants that is not confined to immigrants from a few major countries, whether the weighted or unweighted specifications are used.

V. Supply-Based Explanations

There are a variety of supply-based models of migration, with the earliest being that of Sjaastad (1962). The Sjaastad approach postulates that an individual migrates when the returns to migrating from higher wages exceed the costs. The well-known work of Borjas (1987) also predicts migration on the basis of costs and returns, but the Borjas model formalizes the Roy (1951) model of comparative advantage, arguing that it can be used to explain differential within-origin country migration patterns.[7]

Table 3 examines implications of supply-based models and provides results that suggest they are insufficient to explain attainment of immigrants in the United States or even the pattern of migration. Column 1 attempts to relate the representation of origin countries among immigrants in the United States, that is, to explain R_i using factors that might reasonably be interpreted reflecting primarily immigrant supply considerations. Variables like GDP per capita, share in agriculture, and economic growth rates should be closely related to wage differences between the United States and origin countries and therefore returns to migrating to the United States. Only distance from the United States, consistent with gravity models of migration,[8] enters as a significant determination of representation of particular origin countries. More detailed information on within-origin-country educational attainment can be obtained using International Public Use Microdata Series (IPUMS) data (Minnesota Population Center 2018). Perhaps having more variation in education for a given mean implies that there are some with high

[7] Grogger and Hanson (2011) find some support for the wage differential model driving selection of immigrants into a country. In particular, because there is a large difference between wages of high- and low-skilled immigrants in the United States, skilled immigrants tend to prefer the United States as a destination country. What seems most relevant is the difference between the wages of the skilled in the destination country and the origin country, but this is likely to be correlated with the destination country's skill premium.

[8] Gravity models, used in trade, can be applied to immigration. See, e.g., Bergstrand (1985), Karemera, Oguledo, and Davis (2000), and Lewer and Van den Berg (2008), the latter two of which are direct applications to immigration.

Table 3
Determining Immigrant Flows and Attainment Using Origin Country Variables

	Representation Ratio R_i		Average Education of Immigrants from Country i			
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Log of number of immigrants from country in millions ($\ln(I_i)$)					−.640***	
					(.128)	
Log of population of origin country in 100 millions ($\ln(N_i)$)					.643***	
					(.118)	
Average educational attainment in origin country (μ_i)	.0753	−.145	.378***	.656***	.442***	.442***
	(.572)	(.422)	(.101)	(.188)	(.0858)	(.0828)
Log of GDP per capita, origin country	−1.263	2.321	−.176	−1.600	−.334*	−.333*
	(1.271)	(2.184)	(.224)	(.972)	(.186)	(.184)
Percentage of GDP from agriculture, origin country	.264	−.820	.0802*	.700**	.1000**	.0998**
	(.271)	(.656)	(.0478)	(.292)	(.0399)	(.0390)
Five-year GDP growth rate, origin country	−.321	−.0383	.0211	.0561	−.00171	−.00151
	(.288)	(.172)	(.0509)	(.0765)	(.0426)	(.0415)
Distance from the United States	−1.010**	−.529*	.221***	.0712	.0407	.0406
	(.444)	(.296)	(.0783)	(.132)	(.0714)	(.0708)
90/10 wage ratio in origin country		.814		−.166		
		(.599)		(.267)		
$\ln(R_i)$						−.642***
Log of representation ratio						(.111)
Constant	9.225	3.936	8.610***	6.054**	8.737***	9.280***
	(6.817)	(5.252)	(1.202)	(2.337)	(.985)	(.984)
Observations	69	31	69	31	69	69
R^2	.129	.282	.332	.501	.566	.566
Notes		OECD countries		OECD countries		

SOURCE.—American Community Survey for I_i, R_i, N_i, education, wages, and earnings of US immigrants; United Nations Development Program Human Development Reports for information on average education within origin country (μ_i); World Bank database 2015 for origin country population (N_i) and share of gross domestic product (GDP) from agriculture; Heston, Summers, and Aten (2012) for GDP per capita and GDP growth; Organization for Economic Cooperation and Development (OECD) "Income Distribution," OECD Social and Welfare Statistics (database) for 90/10 wage ratio.

NOTE.—The unit of analysis is an origin country. Variables reflect means of the immigrants from a particular origin country. Standard errors are in parentheses.

* $p < .10$.

** $p < .05$.

*** $p < .01$.

levels of education who might find it worthwhile to migrate to the United States.[9] This finds no support. The variation in education within origin country (not shown) does not enter significantly when added to the specification in column 1.

Column 2 and especially column 4 provide more direct tests of the comparative advantage approach to migration. The OECD gathers and harmonizes wage data on a subset of the countries in this study (see OECD 2019a, 2019b). Skill differentials in the United States are large relative to those in other countries, particularly those in the OECD data. It is well known that there are high returns to education in the United States. It is also well known that the United States has high inequality, as measured by the 90th to 50th or 90th to 10th percentile wage ratios compared with other OECD countries.[10] Comparative advantage implies that when the skill differential is low in a country, the highly educated should be the ones who find it relatively most attractive to move to the United States because educated, high-wage workers do relatively better in the United States. "Brain drain" is the commonly used term to describe the phenomenon of the most educated emigrating from a country where education is underrewarded to countries where the education premium is high.

Column 2 replicates column 1 but uses the subset of countries for which the distribution of wages, here measured by the 90/10 wage ratio, is available. Inclusion of the 90/10 ratio does not change the earlier conclusion. None of the supply-side variables predicts the representation ratio in the subset of 31 countries. Specifically, the 90/10 wage ratio does not enter significantly.

Column 3 examines whether supply factors present in the origin country affect the educational attainment of US immigrants. One might expect that the most talented would leave countries with poor prospects for the highly educated, which would imply a negative relation of educational attainment in the United States to GDP per capita, growth, and a positive relation with the share in agriculture. None shows up with statistical significance at conventional levels, although the share in agriculture does enter significantly in alternative specifications. Distance of the origin country from the United States matters. The more distant, the higher the attainment of immigrants, perhaps reflecting a cost-benefit trade-off as postulated in the original Sjasstad model. It is also true that the education level in the origin country is a significant and positive predictor of immigrant attainment, but this variable is common to the simple supply-based explanation and to the policy-selection-from-the-top model posited here.

[9] These measures are obtained from International Public Use Microdata Series, International: version 7.0 [data set], Minneapolis, MN, IPUMS 2018.

[10] See Lazear (forthcoming) for recent evidence on the comparison between the United States and other OECD countries.

The most direct test of the comparative advantage argument is reported in column 4. The 90/10 ratio measures dispersion in wages in origin countries. The comparative advantage mechanism would suggest that when the 90/10 ratio is low, the most skilled leave the country because their opportunities are relatively better elsewhere. Therefore, the immigrants in the United States from a country with low 90/10 ratios should be more skilled, which implies that educational attainment of immigrants from that country should be high when the 90/10 at home is low. If the most talented cannot use their talents at home, they have incentives to move to those countries in which education and skill receive larger premiums. The implication is that coefficient on the 90/10 ratio should be negative in column 4. The 90/10 wage ratio at home enters negatively but not significantly.

Column 5 adds the policy selection variables—namely, $\ln(I_i)$, $\ln(N_i)$, and μ_i—to the specification in column 3. As before, the policy selection variables continue to be the most important predictors of attainment, even with the inclusion of supply variables. The coefficients have the appropriate signs and statistical significance, and they follow the earlier-noted pattern that the coefficients on $\ln(I_i)$ and $\ln(N_i)$ are equal of opposite sign. In this specification, distance of the origin country from the United States drops out, and share in agriculture is now statistically significant. Column 6 replaces $\ln(I_i)$ and $\ln(N_i)$ with simply the representation ratio, $\ln(R_i)$. The results are virtually identical, and the coefficient on the representation ratio is negative and significant, as in table 2.

VI. Sweden

Swedish data provide additional evidence that policy-based selection rules are important determinants of immigrant flows and attainment. Sweden's immigration policy places strong emphasis on immigrants needing asylum and therefore accepts a large number of refugees from war-torn countries. Bosnia, Lebanon, Somalia, Eritrea, and Iraq are overrepresented as origin countries among Swedish immigrants. A comparison of Sweden to the United States is useful because Sweden uses a different explicit policy than the United States, with the United States emphasizing family reunification relatively more and refugee status relatively less. Swedish data allow the policy selection model of attainment to be tested in a different environment where different origin countries might drive the results.

The underlying data used for Sweden are similar to the ACS for the United States and covers the years 2011–15, as in the United States. The Swedish variables on educational attainment are somewhat different because education is reported as being within brackets rather than the more continuous measure of years of education completed. Thus, two attainment variables are used. The first is the proportion of immigrants from a given origin country who have at least 3 years of postsecondary education. The

second is a constructed average educational level for the origin country, created by assuming an average level of education within a bracket and aggregating to get a mean number of years completed for each immigrant group. Table 4 reports the results, all of which support the predictions of equations (1) and (2).

The results in column 1 uses the proportion with advanced education as the dependent variable. Both policy selection variables, $\ln(I_i)$ and $\ln(N_i)$, enter as predicted, and as was the case using the US data, the coefficients are approximately equal of opposite signs. The three variables explain 60% of the variation. Column 2 repeats the analysis but substitutes the representation ratio for the number of immigrants from and population of an origin country taken separately. The effect is negative and statistically significant, just as it was using US data. Columns 3 and 4 perform the same estimation, but the dependent variable is the estimated average level of education for immigrants from each origin country. The results are qualitatively similar to those in columns 1 and 2, and in column 3, as in column 1, coefficients on $\ln(I_i)$ and $\ln(N_i)$ are statistically equal of opposite sign.

Table 4
Educational Attainment of Immigrants in Sweden

	Proportion of Immigrants from Country *i* with at Least 3 Years of Postsecondary Education		Average Education of Immigrants from Country *i*	
Variable	(1)	(2)	(3)	(4)
Number of immigrants from country *i* in 100,000 ($\ln(I_i)$)	−.0343***		−.239***	
	(.00670)		(.0666)	
Population of origin country in 100 millions ($\ln(N_i)$)	.0446***		.306***	
	(.00674)		(.0561)	
Average educational attainment in origin country (μ_i)	.0307***	.0294***	.136***	.128***
	(.00365)	(.00346)	(.0305)	(.0285)
Log of representation ratio ($\ln(R_i)$)				
		−.0394***		−.279***
		(.00497)		(.0453)
Constant	.00878	.0766**	10.94***	11.41***
	(.0359)	(.0323)	(.305)	(.267)
Observations	78	78	77	77
R^2	.595	.588	.395	.390

NOTE.—Shown are unweighted regressions on data from Statistics Sweden. The unit of analysis is an origin country. Variables reflect means of the immigrants from a particular origin country to Sweden, taken from Swedish registry data. The last two columns have one fewer observation because of incomplete data required to compute mean education. Standard errors are in parentheses.

** $p < .05$.

*** $p < .01$.

The conclusion is that the predictions are borne out and the results for Sweden mirror those for the United States, despite Sweden's explicit immigration policy differing from that of the United States. Both countries have in common that slots are rationed and that there is an excess supply of potential immigrants, which results in attainment of immigrants being negatively related to representation in the pool of immigrants.

VII. Additional Considerations

A. Policy Deviates from That Assumed

In a typical year, more than 60% of those issued permanent resident status in the United States are family sponsored. This seems at odds with the assumption that selection is from the top of the origin country's distributions. The model is admittedly stylized, but evidence on the validity of the assumption is useful.

To begin, in 129 of 129 cases the difference between the average educational attainment among immigrants and the average attainment at home is positive, equaling 4.8 years on average. Still, those who migrate to the United States might have higher education than those who remain in the origin country because the United States has higher average education than the world as a whole and education might be cheaper to acquire in the United States.[11] The raw ACS data allow an examination of individuals who came to the United States as adults and were therefore likely to have received most of their education elsewhere. The individual data reveal that 86% of immigrants who are more than 25 years old and arrived within the last 6 years have attained levels of education that exceed the mean in their origin country. Those who come to the United States, even as adults, appear to have higher educational achievement than those left behind.

B. Undocumented Immigrants

Because undocumented entrants use unconventional immigration channels, the model is less appropriate for them. Supply considerations would seem more important and immigration policy less important for those who enter without a permanent visa or who overstay their temporary ones. Passel and Cohn (2017) estimate the number of "unauthorized" immigrants by region of the world from which they migrated. Those estimates can be used as an additional independent variable to determine whether immigrants from countries in regions with a larger number of people in the United States without authorization have lower attainment.

To check this, all the regressions shown in table 2 were repeated with the addition of the Passel and Cohn number of unauthorized immigrants from

[11] Chiswick and Miller (2011) argue that there is some negative assimilation that occurs after migration, and Chiswick and Miller (2012) investigate both negative and positive assimilation.

the region as a variable. The variable enters negatively and significantly in the wage and earnings specifications but not in the educational attainment regressions. Even when the variable is significant, the main predictions of the policy selection model are supported, with the coefficients on $\ln(I_i)$, $\ln(N_i)$, μ_i, and $\ln(R_i)$ changing only slightly and not qualitatively.

VIII. Conclusion

The larger the number of immigrants from a given country, the lower are educational attainment, wage rates, and earnings of that group. The pattern is a result of a selective immigration process. Because of variations in the way the various origin countries are treated by the immigration system, a particular distribution of immigrants results, which gives rise to differences in educational attainment and earnings by country of origin. Those countries that are awarded the largest number of slots tend to have immigrants with lower attainment levels.

For the same reason, the larger the population of the origin country, the higher the attainment of immigrants from that country. It is easier to select 1 million highly educated people from India, which has 1.3 billion people, than it is from Laos, which has 7 million people. Consequently, immigrants of Indian origin have higher levels of educational attainment than do immigrants of Laotian origin.

A model of selection is constructed that yields specific empirical implications, all of which are borne out by data from the ACS 2011–15. The larger the number of immigrants from an origin country, the lower the level of educational attainment, wages, and earnings in the United States. The larger the population of the origin country, the higher the educational attainment, the higher the wages, and the higher the earnings of those immigrants in the United States. A more parsimonious approach expresses predictions in terms of a representation ratio, which is a measure of how under- or overrepresented an origin country is in the US immigrant stock. Immigrants from countries that are overrepresented are predicted and found to have lower attainment in education, wages, and earnings.

Data from Sweden support the hypotheses and yield results consistent with those from the United States. This is an important source of corroboration because the explicit immigration policy of Sweden differs from that of the United States and because the overrepresented origin countries among Swedish immigrants differ from the overrepresented origin countries among US immigrants.

The parsimonious model that uses only two variables explains the majority of the variation of both country-mean educational attainment and country-mean wages of immigrants in the United States. It also does well in accounting for educational attainment variation among Swedish immigrants. Overall,

the model that postulates selection from the top of origin countries' ability distribution describes the actual data well.

References

Bergstrand, Jeffry H. 1985. The gravity equation in international trade: Some microeconomic foundations and empirical evidence. *Review of Economics and Statistics* 67, no. 3:474–81.

Borjas, George J. 1987. Self-selection and the earnings of immigrants. *American Economic Review* 77, no. 4:531–53.

Borjas, George J., and Rachel M. Friedberg. 2009. Recent trends in the earnings of new immigrants to the United States. NBER Working Paper no. 15406, National Bureau of Economic Research, Cambridge, MA.

Connor, Phillip. 2018. US diversity visa lottery saw near-record number of applicants in 2017. Pew Research Center. www.pewresearch.org/fact-tank/2018/08/23/applications-for-u-s-visa-lottery-more-than-doubled-since-2007/.

Chiswick, Barry R., and Paul Miller. 2011. The "negative" assimilation of immigrants: A special case. *Industrial and Labor Relations Review* 64, no. 3:502–25.

———. 2012. Negative and positive assimilation, skill transferability, and linguistic distance. *Journal of Human Capital* 6, no. 1:35–55.

Clark, Jimena, Timothy J. Hatton, and Jeffrey G. Williamson. 2007. Explaining US immigration 1971–1998. *Review of Economics and Statistics* 89, no. 2:359–73.

Grogger, Jeffrey, and Gordon Hanson. 2011. Income maximization and the selection and sorting of international migrants. *Journal of Development Economics* 95, no. 1:42–57.

Heston, Alan, Robert Summers, and Bettina Aten. 2012. Penn world table version 7.1. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Hunt, Jennifer. 2011. Which immigrants are most innovative and entrepreneurial? *Journal of Labor Economics* 29, no. 3:417–57.

———. 2015. Are immigrants the most skilled US computer and engineering workers? *Journal of Labor Economics* 33, no. S1:S39–S77.

Karemera, David, Victor Iwuagwu Oguledo, and Bobby Davis. 2000. A gravity model analysis of international migration to North America. *Journal of Applied Economics* 32, no. 13:1745–55.

Kerr, William R., Sari Pekkala Kerr, and William F. Lincoln. 2015. Skilled immigration and the employment structure of US firms. *Journal of Labor Economics* 33, no. S1:S147–S186.

Kerr, William R., and William F. Lincoln. 2010. The supply side of innovation: H1-B visa reforms and U.S. ethnic invention. *Journal of Labor Economics* 28, no. 3:473–508.

Lazear, Edward P. 1999. Culture and language. *Journal of Political Economy* 107, no. S6:S95–S126.
———. 2017. Why are some immigrant groups more successful than others? NBER Working Paper no. 23548, National Bureau of Economic Research, Cambridge, MA.
———. Forthcoming. Uneven productivity growth as a source of inequality. National Bureau of Economic Research, Cambridge, MA.
Lewer, Joshua L., and Hendrik Van den Berg. 2008. A gravity model of immigration. *Economic Letters* 99, no. 1:164–67.
Lewis, Ethan. 2013. Immigrant-native substitutability and the role of language. In *Immigration, policy, and socioeconomic inequality*, ed. David Card and Stephen Raphael, 60–97. New York: Russell Sage Foundation.
Mayda, Anna Maria. 2010. International migration: A panel analysis of the determinants of international flows. *Journal of Population Economics* 23, no. 4:1249–74.
Minnesota Population Center. 2018. Integrated public use microdata series, international: Version 7.0. https://doi.org/10.18128/D020.V7.0.
OECD (Organization for Economic Cooperation and Development). 2019a. Gross earnings: Decile ratios. OECD Employment and Labour Market Statistics. https://doi.org/10.1787/data-00302-en.
———. 2019b. Income distribution. OECD Social and Welfare Statistics. https://doi.org/10.1787/data-00654-en.
Ortega, Francesc, and Giovanni Peri. 2013. The effect of income and immigration policy on international migration. *Migration Studies* 1, no. 1:47–74.
Passel, Jeffrey S., and D'Vera Cohn. 2017. As Mexican share declined, U.S. unauthorized immigrant population fell in 2015 below recession level. Pew Research Center. http://www.pewresearch.org/fact-tank/2017/04/25/as-mexican-share-declined-u-s-undocumented-immigrant-population-fell-in-2015-below-recession-level/.
Peri, Giovanni, Kevin Shih, and Chad Sparber. 2015. STEM workers, H-1B visas, and productivity in US cities. *Journal of Labor Economics* 33, no. S1:S225–S255.
Roy, Arthur. 1951. Some thoughts on the distribution of earnings. *Oxford Economic Papers* 3, no. 2:135–46.
Sjaastad, Larry A. 1962. The costs and returns of human migration. *Journal of Political Economy* 70, no. 5:80–93.
Statistics Sweden. Various years. Registret över befolkningens utbildning (Register of Population Education).
United Nations. Various years. United Nations Development Program Human Development Reports, education statistics. http://hdr.undp.org/en/indicators/103006.
US Census Bureau. 2011–2015. American Community Survey. https://www.census.gov/programs-surveys/acs.

US Department of Homeland Security. 2016. Yearbook of immigration statistics: 2014. US Department of Homeland Security, Office of Immigration Statistics. https://www.dhs.gov/sites/default/files/publications/DHS%202014%20Yearbook.pdf

US Department of State, Bureau of Consular Affairs. 2015. Annual immigrant visa waiting list report as of November 1, 2015.

———. 2018. Annual report of immigrant visa applicants in the family-sponsored and employment-based preferences registered at the national visa center as of November 1, 2018.

World Bank. Various years. World Bank Statistical Data Base. World Bank, Washington, DC. https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS?view=chart, https://data.worldbank.org/indicator/SP.POP.TOTL?view=chart.