

# Self-Selection and the Earnings of Immigrants

George Borjas (1987)

Omid Ghaderi & Ali Yadegari

April 7, 2018

# Abstract

- The age-earnings profile of immigrants is steeper than the age-earnings profile of the native population.
- Human capital framework: stronger investment incentives.
- The age-earnings profile of immigrants crosses the age-earnings profile of natives about 10 to 15 years after immigration.
- Unobserved characteristics: immigrants may be more able and more highly motivated.
- But, how cohort quality and immigrant self-selection are related?
- Individuals compare the potential incomes in the U.S. with the incomes in the home countries, and then make the migration decision.
- Variations in political and economic conditions in the countries of origin can explain differences in the earnings of immigrants.

# Questions

- What is the role of self-selection and income maximization?
- Are immigrants selected from the upper or lower tail of the income distribution in the sending countries?
- If immigrants are drawn from the upper tail of the income distribution in the home country, does that ensure they end up in the upper tail of the U.S. income distribution?
- If cohort quality has experienced a secular decline in the postwar period, what factors are responsible for this change?

## The Model: Assumptions

There are two countries: country 0 (home) and country 1 (destination)

Earning distribution (home country):  $\ln w_0 = \mu_0 + \varepsilon_0$

Earning distribution (destination country):  $\ln w_1 = \mu_1 + \varepsilon_1$

Unobserved characteristics (skill):  $\varepsilon_0 \sim N(0, \sigma_0^2)$  &  $\varepsilon_1 \sim N(0, \sigma_1^2)$

Time equivalent cost of migrating:  $\pi = \frac{C}{w_0}$

The correlation between earnings:  $\rho = \frac{\sigma_{01}}{\sigma_0\sigma_1}$

Each worker knows  $C, \mu_0, \mu_1$  and his individual epsilons ( $\varepsilon_0, \varepsilon_1$ )

We only observe  $\varepsilon_0$  or  $\varepsilon_1$  for any individual

## The Model: Equations

$$I = (\mu_1 - \mu_0 - \pi) + (\varepsilon_1 - \varepsilon_0)$$

$$\nu = \varepsilon_1 - \varepsilon_0$$

$$z = \frac{(\mu_0 - \mu_1 + \pi)}{\sigma_\nu}$$

$$P = Pr[\nu > (\mu_0 - \mu_1 + \pi)]$$

$$= 1 - \Phi(z)$$

$\Phi(\cdot)$  is the CDF of the standard normal

## The Model: Equations (continued)

$$\begin{aligned} E(\ln w_0 | I > 0) &= \mu_0 + E(\varepsilon_0 | \frac{\nu}{\sigma_\nu} > z) \\ &= \mu_0 + \frac{\sigma_0 \sigma_1}{\sigma_\nu} (\rho - \frac{\sigma_0}{\sigma_1}) \lambda \end{aligned}$$

$$\begin{aligned} E(\ln w_1 | I > 0) &= \mu_1 + E(\varepsilon_1 | \frac{\nu}{\sigma_\nu} > z) \\ &= \mu_1 + \frac{\sigma_0 \sigma_1}{\sigma_\nu} (\frac{\sigma_1}{\sigma_0} - \rho) \lambda \end{aligned}$$

$$\lambda = \frac{\phi(z)}{P} = \frac{\phi(z)}{1 - \Phi(z)}$$

$\phi(\cdot)$  is the PDF of the standard normal  
 $\lambda$  is the Inverse Mills Ratio (IMR)

## Theoretical Cases: Positive Selection (case 1)

- $Q_0 > 0$  and  $Q_1 > 0$
- $\rho > \frac{\sigma_0}{\sigma_1}$  and  $\frac{\sigma_1}{\sigma_0} > 1$
- Correlation between the skills valued in the destination and home country is sufficiently high.
- Destination country has a higher "return to skill" than the home country.
- "The best and the brightest" leave their home countries for greater opportunity.

## Theoretical Cases: Negative Selection (case 2)

- $Q_0 < 0$  and  $Q_1 < 0$
- $\rho > \frac{\sigma_1}{\sigma_0}$  and  $\frac{\sigma_0}{\sigma_1} > 1$
- Home country is unattractive to low earnings workers because of high wage dispersion.
- These immigrants do not perform well in the destination country's labor market.
- A compressed wage structure "subsidizes" low skill workers, thus attracting low skill workers from abroad.



## Theoretical Cases: Refugee Sorting (case 3)

- $Q_0 < 0$  and  $Q_1 > 0$
- $\rho < \min\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right)$
- Correlation between earnings in the two countries is sufficiently low (could be negative).
- This might occur, for a minority group whose opportunities in the home country are depressed by prejudice.

## Theoretical Cases: Fourth Case

- $Q_0 > 0$  and  $Q_1 < 0$
- $\rho > \max\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right)$
- Fourth case is theoretically impossible, since it requires  $\rho > 1$ .
- People leave the upper tail of the home country income distribution to join the lower tail of the destination country distribution.
- We may have this type of migration in Iran!

# Composition & Scale Effects

$$k = \frac{\sigma_1}{\sigma_0}$$

$$\gamma = \left(\frac{\sigma_0\sigma_1}{\sigma_\nu}\right)(k - \rho)$$

$$Q_1 = \gamma\lambda$$

$$\frac{\partial Q_1}{\partial \mu_0} = \frac{\sigma_0\sigma_1}{\sigma_\nu^2}(k - \rho)\frac{\partial \lambda}{\partial z}$$

$$\frac{\partial Q_1}{\partial \sigma_0} = \frac{\sigma_0\sigma_1^2}{\sigma_\nu^3}(\rho^2 - 1)\lambda - \frac{\sigma_0^2\sigma_1}{\sigma_\nu^3}(k - \rho)(1 - \rho k)\frac{\partial \lambda}{\partial z}z$$

$$\frac{\partial Q_1}{\partial \rho} = -\frac{\sigma_0^3\sigma_1}{\sigma_\nu^3}(1 - \rho k)\lambda + \frac{\sigma_0^2\sigma_1^2}{\sigma_\nu^3}(k - \rho)\frac{\partial \lambda}{\partial z}z$$

# Empirical Framework

$$\begin{aligned}\ln w_i(T) = & X_i\theta_T + (\delta + \beta_1 T + \beta_2 T^2)l_i \\ & + (\alpha_1 - \beta_2 - 2\beta_2 T)l_i y_i \\ & + (\alpha_2 + \beta_2)l_i y_i^2 + \nu_i\end{aligned}$$

- The predicted wage differential in 1979 between the most recently arrived immigrant cohort and the native base.
- The rate of wage growth (relative to natives) for an immigrant cohort that has resided in the U.S. for 10 years.
- The predicted wage differential immediately after immigration between the 1979 cohort and the 1955 cohort.

# Data Structure

- The data are drawn from the 1970 and 1980 US censuses.
- The complete samples are used in the creation of the immigrant extracts.
- Random samples are drawn for the native "baseline" population.
- Analysis is restricted to men aged 25-64 who:
  - was employed in the calendar year prior to the census.
  - was not self-employed or working without pay.
  - was not in the Armed Forces.
  - did not reside in group quarters.

## Data Structure (continued)

- All immigrants groups will be compared to a single native base:
  - White
  - Non-Hispanic
  - Non-Asian
- 41 countries were chosen for analysis at least 80 observation of immigrants.
- The 41 countries under analysis account for 90.4 percent of US immigrants.

# Regression Results

- Percent ranges from the trivially small (0.04 percent for Brazil and USSR) to the large (10percent for Jamaica).
- Migration flow isnt constant.
- Declining importance of west Europe as a source.
- Increasing importance of Asia and Latin America as a source.
- Changing characteristics of sending countries changed the type of selection that distinguish the immigrant population from the native born.

# Regression Results (continued)

Country of Birth	1951–80 Immigration		1951–60	1971–80
	Total Number (in 1000s)	As Percent of 1980 Population <sup>a</sup>	Immigrants as Percent of 1950 Population <sup>a</sup>	Immigrants as Percent of 1970 Population <sup>a</sup>
<b>Europe:</b>				
Austria	48.1	.6	.4	.1
Czechoslovakia	60.4	.4	.2	.1
Denmark	30.0	.6	.3	.1
France	90.1	.2	.1	.04
Germany	611.5	1.0	.7	.1
Greece	232.3	2.4	.6	1.1
USSR	105.4	.04	.02	.02
Argentina	81.5	.3	.1	.3
Brazil	43.1	.04	.02	.01
Canada	676.4	2.8	2.0	.5
Colombia	165.5	.6	.4	.6
Cuba	611.9	6.3	1.5	3.2
Dominican Republic	251.9	4.3	.5	3.4
Ecuador	96.7	1.2	.3	.8
Guatemala	45.1	.7	.1	.5
Haiti	100.2	1.8	.1	1.3
Jamaica	221.7	10.3	.6	7.3



# Socioeconomic Characteristics

- Year of schooling
- Age
- Age-squared
- Whether health limit work
- Whether married
- Spouse present
- Whether resident of an SMSA
- Income in the year preceding the census as the dependent variable

# Model Estimates

Country of Birth	1970		1980			Rate of Assimilation at $y = 10$	1955–79 Change in Cohort Quality
	$I$	$I \cdot y$	$I$	$I \cdot y$	$I \cdot y^2$		
<b>Europe:</b>							
Austria	.0189 (.26)	.0036 (.75)	.0321 (.52)	.0034 (.82)	-.00003 (-.45)	.0040 (.66)	.0287 (.20)
Czechoslovakia	-.1525 (-2.48)	.0147 (3.34)	-.1441 (-2.79)	.0127 (3.23)	-.00019 (-2.74)	.0088 (1.64)	-.0143 (-.10)
Denmark	.0838 (.82)	-.0033 (-.44)	.2018 (2.14)	-.0056 (-.81)	.00009 (.72)	.0068 (.78)	.2441 (1.21)
France	-.0785 (-1.28)	.0020 (.47)	.0999 (2.48)	-.0046 (-1.33)	.00005 (.79)	.0111 (2.05)	.3183 (2.74)
Germany	.0999 (3.82)	-.0025 (-1.37)	.1409 (5.40)	-.0047 (-2.62)	.00007 (2.38)	-.0002 (-.10)	.0618 (1.17)
Egypt	-.4466 (-7.00)	.0421 (5.67)	-.4586 (-10.84)	.0396 (7.57)	-.00056 (-4.34)	.0260 (4.76)	-.0706 (-.57)
India	-.2847 (-7.09)	.0453 (9.71)	-.4340 (-21.41)	.0497 (16.75)	-.00096 (-11.03)	.0179 (5.33)	-.2845 (-3.84)
Iran	-.4078 (-4.71)	.0229 (3.03)	-.3101 (-10.19)	.0249 (5.45)	-.00031 (-2.47)	.0294 (4.13)	.2690 (1.88)
Israel	-.2998 (-4.19)	.0282 (4.54)	-.3397 (-8.44)	.0260 (5.74)	-.00041 (-3.84)	.0128 (2.11)	-.1314 (-1.00)
Canada	.0645 (2.86)	.0003 (.17)	.1165 (6.06)	-.0013 (-.91)	-.00000 (-.21)	.0030 (1.50)	.0988 (2.17)
Colombia	-.2247 (-4.33)	.0169 (2.74)	-.4030 (-12.67)	.0219 (5.78)	-.00036 (-3.71)	-.0007 (-.17)	-.3444 (-3.82)
Cuba	-.4612 (-22.20)	.0214 (8.89)	-.4517 (-18.26)	.0208 (9.24)	-.00025 (-5.20)	.0164 (9.74)	.0129 (.28)

# Country Specific Variables

Variable	Definition and Source	Mean	Minimum	Maximum	U.S. Value
Politically Competitive System	= 1 if the country had a competitive party system during the entire 1950–73 period; 0 otherwise. <i>Source: Cross-National Time-Series Archive (CNTSA)</i>	.41	–	–	1
Recent Loss of Freedom	= 1 if the country had a competitive party system at the beginning of the period but had a non-competitive party system at the end of the period; 0 otherwise. <i>Source: CNTSA.</i>	.20	–	–	0
Number of Assassinations	Number of politically motivated murders or attempted murders of high government officials or politicians in 1950–73. <i>Source: CNTSA.</i>	3.27	0	22	12
Income Inequality	Ratio of household income of the top 10 percent of the households to the income of the bottom 20 percent of the households. <i>Source: World Bank (various issues) and United Nations (1977).</i>	7.50	1.42	30.0	5.91
Distance from U.S.	Number of air miles (in thousands) between the country's capital and the nearest U.S. gateway (Los Angeles, Miami, or New York). <i>Source: Airline offices contacted by author.</i>	3.37	.18	7.49	–
English Proficiency	Fraction of 1975–80 cohort of immigrants who speak English well or very well. <i>Source: 5/100 A Sample of the 1980 U.S. Census.</i>	.74	.24	1.00	–
Age at Migration	Mean age at migration. <i>Source: 5/100 A Sample of the 1980 U.S. Census.</i>	24.56	12.40	32.40	–

# Regression Results (continued)

Country of Origin Characteristics	Regression			
	1	2	3	4
Intercept	-.2214 (-3.88)	.1838 (1.06)	-.9934 (-3.41)	-.9469 (-3.30)
Politically Competitive System	.2743 (4.49)	.1306 (2.01)	.1101 (2.16)	.1264 (2.39)
Recent Loss of Freedom	-.0010 (-.01)	-.0511 (-.75)	-.0062 (-.12)	.0136 (.25)
Number of Assassinations	-.0072 (-1.20)	-.0028 (-.54)	.0021 (.51)	.0044 (.92)
Income Inequality	-.0084 (-1.78)	-.0038 (-.89)	.0039 (1.02)	.0046 (1.13)
Distance from U.S.	-	-.0114 (-.89)	-.0031 (-.31)	.0018 (.09)
English Proficiency	-	.2596 (2.20)	.1980 (2.12)	.2030 (2.21)
Mean Age at Migration	-	-.0217 (-3.55)	-.0149 (-2.99)	-.0119 (2.28)
ln (per capita <i>GNP</i> )	-	-	.1164 (4.57)	.1015 (3.77)
Country in Asia or Africa	-	-	-	-.1145 (-1.58)
Country in North or South America	-	-	-	-.0640 (-.73)
$R^2$	.504	.681	.808	.826

# Regression Results (continued)

Country of Origin Characteristics	Regression			
	1	2	3	4
Intercept	.0076 (2.96)	-.0240 (-3.88)	-.0237 (-1.50)	-.0280 (-2.32)
Politically Competitive System	-.0029 (-1.06)	-.0068 (-2.66)	-.0068 (-2.60)	-.0091 (-4.28)
Recent Loss of Freedom	.0063 (1.81)	.0029 (1.21)	.0030 (1.15)	.0021 (1.06)
Number of Assassinations	.0008 (2.68)	.0006 (2.36)	.0006 (2.14)	.0008 (3.07)
Income Inequality	-.0001 (-.50)	-.00002 (-.11)	-.00002 (-.10)	.0002 (.90)
Distance from U.S.	-	.0003 (.74)	.0003 (.70)	-.0027 (-2.89)
English Proficiency	-	.0138 (3.27)	.0138 (3.20)	.0122 (3.70)
Mean Age at Migration	-	.0009 (4.28)	.0009 (3.95)	.0009 (4.72)
ln (per capita <i>GNP</i> )	-	-	-.00002 (-.01)	.0021 (1.83)
Country in Asia or Africa	-	-	-	.0151 (5.11)
Country in North or South America	-	-	-	-.0080 (-2.08)
R <sup>2</sup>	.302	.704	.704	.842

# Regression Results (continued)

Country of Origin Characteristics	Regression			
	1	2	3	4
Intercept	-.3194 (-3.19)	-.9951 (-3.97)	-1.1779 (-4.08)	-2.2202 (-4.69)
Politically Competitive System	.1760 (2.54)	.1075 (1.60)	.0712 (.97)	.0630 (.70)
Recent Loss of Freedom	.1256 (1.67)	.1468 (2.16)	.1272 (1.81)	.1310 (1.33)
Number of Assassinations	.0077 (1.19)	.0156 (2.32)	.0122 (1.69)	.0256 (2.00)
Rate of Change in Central Government Expenditures	.0698 (1.60)	.0699 (1.75)	.0641 (1.60)	-.0099 (-.21)
Rate of Change in Per Capita GNP	4.7010 (2.27)	3.0956 (1.60)	1.1567 (.46)	-1.5321 (-.50)
ln (per capita <i>GNP</i> )	-	.0889 (1.93)	.1186 (3.22)	.2443 (4.15)
Country in Asia or Africa	-	-	.1374 (1.42)	-
Country in North or South America	-	-	.0274 (.41)	-
Change in Quota	-	-	-	.0034 (2.26)
$R^2$	.284	.418	.453	.581

## Regression Results (continued)

Country of Origin Characteristics	Regression	
	1	2
Intercept	-.6060 (-1.30)	-1.1614 (-2.46)
Politically Competitive System	.1206 (1.13)	.0801 (.81)
Recent Loss of Freedom	.1096 (.95)	-.0365 (-.32)
Number of Assassinations	-.0245 (-2.65)	-.0337 (-3.65)
Income Inequality	-.0113 (-1.51)	-.0145 (-2.00)
Distance from U.S.	-.1332 (-6.11)	-.1271 (-2.68)
English Proficiency	.1661 (.94)	.0488 (.30)
ln (per capita <i>GNP</i> )	-.1130 (-2.14)	-.0441 (-.83)
Country in Asia or Africa	-	.3386 (2.19)
Country in North or South America	-	.2923 (1.52)
$\chi^2$	98.45	108.82

Thanks !