Educated Preferences:
Explaining Attitudes Toward Immigration In Europe

Jens Hainmueller and Michael J. Hiscox

Last revised: March 2004

Supplement I: Data on Skills of Immigrants by Country of Origin

A. Introduction

The aim of this supplement is to provide reviewers of our paper with additional, detailed evidence supporting the assumption that immigrants from richer countries have higher skill levels on average than immigrants from poorer countries (for both European and non-European sets of countries).

The evidence we present is based on the *International File of Immigration Surveys* database compiled by sociologist Frank van Tubergen (Utrecht University, Netherlands).[1] This database combines survey data on more than 300,000 immigrants from 180 countries of origin and 18 destination countries (14 of which are European). The data have been extracted from the European Union's *Labour Force Survey*, national census data, and additional country-specific immigrant surveys. All surveys were harmonized and pooled by van Tubergen into a cross-national data set that provides comparable individual-level information on immigrants, classified by country of origin, for the period 1980-2001. To our knowledge this represents the most comprehensive data set on immigrant populations currently available.

Van Tubergen generously provided us with the data on the condition that, due to contractual agreements he signed with data archives in different nations, and given the fact that his own work with the data is yet to be published, we agreed not to present the data in any extensive detail in the paper we are submitting for publication. We have summarized the main findings from our analysis of the van Tubergen data and included them in the relevant section of the paper. In order to provide reviewers with more details about the data we present a more extensive treatment of the analysis here.

The presentation of the evidence is organized as follows. First, we briefly describe the characteristics of the data we are drawing upon. Second, we consider various possible definitions of "richer" and "poorer" countries. In part three we show comparative data on the skill distributions of immigrants, examining differences between those from richer and poorer origin countries (within and outside Europe) as defined in these various ways. We find that on average

---

[1] Van Tubergen, Frank. 2004. *International File of Immigration Surveys.* [Codebook and machine-readable data set]. Utrecht: Department of Sociology/ICS. We are deeply indebted to Frank van Tubergen for allowing us to examine this data. By agreement with him, we only present aggregated results of our analysis of the IFIS data at this time.

immigrants from richer countries are indeed considerably more skilled than immigrants from poorer countries. This skill gap is strongest among immigrants from non-European countries.

B. The Data

Since we are only concerned with immigration to European countries, we discard data on non-European destinations, leaving us with the 14 European destination nations provided in the *International File of Immigration Surveys* database. These destination countries are Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Luxembourg, Netherlands, Portugal, Spain, Sweden, and the UK. For these destinations the database provides information on immigrants from 51 origins – 26 European[2] and 25 non-European countries[3] (the definition of "Europe" we use follows the official UNSD classification). Only legal immigration is recorded, of course, and this might be a problem if survey respondents are also thinking about illegal immigrants when stating their views about "immigration" in general. Note, however, that the bias that might creep in here (in terms of expectations about skill differences among different categories of immigrants) most likely works in the direction of *understating* the differences between the skill levels of immigrants from richer and poorer countries as we can assume that, on average, illegal immigration is much stronger from poorer origin countries and that the probability of legal immigration is increasing in skill level.

Van Tubergen provided the data to us in the form of a multivariate table consisting of country of destination, country of origin (both in UNSD codes), proportions of immigrants with low, middle, and high education levels, and the total number of immigrants for each origin-destination pair. Education is coded as the highest level of education completed (using the ISCED-97 codes): *low*, *middle*, and *high* categories refer respectively to primary or first stage of basic education completed, upper secondary education completed, and tertiary education completed. These categories thus match the *educational attainment* variable we use in our paper – and our ELEMENTARY, HIGHSCHOOL, and COLLEGE dummy variables – with the exception that van Tubergen also includes individuals with PHDs in the high education category rather than coding them separately. The data refer to all males and females between the ages of 25 and 54 who are active in the labour market (either unemployed or employed).

C. Distinguishing between "Richer" and "Poorer" Countries

Our assumption is that immigrants from poorer origins are, on average, clearly less skilled than immigrants from richer origins. Following the set of questions in the *European Social Survey* that generate the dependent variables in the analysis in our paper, we intend to compare the skill distributions of immigrants from richer European to those from poorer European, as well as the skill distributions of immigrants from richer countries outside Europe to those immigrants from poorer countries outside Europe. To this end, we split the data into two sub-samples, one consisting of all European and the other of all non-European origin countries.

---

[2] Albania, Austria, Belgium, Bulgaria, Denmark, Ex-Czechoslovakia, Ex-Yugoslavia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Spain, Sweden, Switzerland, and the UK.

[3] Algeria, Argentina, Australia, Brazil, Cambodia, Canada, China, Cyprus, Egypt, Ex-Russia, India, Indonesia, Japan, Lebanon, Mexico, Morocco, New Zealand, Pakistan, Philippine, South Africa, Thailand, Tunisia, Turkey, USA, and Vietnam.

Both comparisons require an arbitrary choice by the researcher (and by survey respondents) about what constitutes a "richer" and a "poorer" country in the given context. In order to reduce arbitrariness to a minimum we rely on the following classification strategy. Whether an origin country is defined as "richer" or "poorer" depends on its relative position within the GDP per capita distribution of its respective origin sub-sample (European or non-European origins).[4] Within each sub-sample we choose various cut-off-points (in terms of the distribution of GDP per capita) to divide the countries into richer and poorer groups, including: the mean, the median, the 40/60 quantile, and the 30/70 quantile.[5] The idea here is to experiment with various distinctions that might approximate the distinctions made intuitively by survey respondents. The GDP per capita distributions and the various cut-off points are displayed in Table 1.

[Table 1]

The first two columns in each panel in Table 1 contain the descriptive statistics of the country sample and the next two columns the countries that correspond to the respective sample cut-off points. The European origin sample thus contains 26 countries, with a mean GDP per capita of $22,445. If we use the mean as the cut-off point in the European sample, all origin countries with GDP per capita above this value are assigned to the "richer" country group and all origin countries below are assigned to "poorer" country group. In this case, Italy is the richest country that ends up in the "poorer" group and the UK is the poorest country that ends up in the "richer" group. The same method is applied when the quantiles are used as a cut-off point. For example, again looking at the European origin sub-sample, the 30th quantile of the GDP per capita distribution is $11,603, and when we use this as the cut-off all countries below this value are assigned to the "poorer" country group (the richest of these being Malta). The 70th quantile is $31,843, and when this is used as the cut-off all countries above this value end up in the "richer" country group (the poorest being Iceland).

D. The Skill distributions of Immigrants from Richer vs. Poorer Countries

For each of the two sub-samples (European and non-European), and using different cut-offs between richer and poorer origins, we calculate the proportions of immigrants arriving in Europe with different levels of educational attainment. This provides a basic profile of the skill levels of immigrants coming to Europe, broken down by whether the immigrants come from richer or poorer origin countries. In the last section, we also examine data on immigrants for each individual origin. The results of our comparisons, for both the European and the non-European sub-samples, are displayed in Table 2.

[Table 2]

The findings lend strong support to the notion that immigrants from richer countries are on average more skilled than their counterparts from poorer countries. This holds true across all

[4] In order to maximize coverage, we use GDP per capita, PPP adjusted, for the year 2001 (measured in constant 1995 US dollars). The GDP data is taken from the World Bank's 2003 WDI database.

[5] Note that in the case of non-European origins, qq.normal plots revealed a substantial amount of leptokurtosis in the upper tail of the GDP distribution. Therefore, the mean cut-off point is to be interpreted with caution. The skew does not affect any of the other cut-off points, of course. We also computed additional tests using the mean of the log transformed GDP distribution and the results were almost identical. QQ.normal plots are displayed in Appendix A

cut-off points (i.e. no matter how "richer" and "poorer" are defined) or whether we compare immigrants from European or non-European origins. The estimated differences in educational attainment are (highly) significant in the great majority of cases and all go in the expected direction as immigrants from richer countries have a lower (higher) proportion of low (high) educational attainment compared to immigrants from poorer origin countries. More important, the differences are also of considerable magnitude, in particular in the case of immigration from non-European origins: If we take the median GDP per capita as a cut-off-point for defining richer and poorer origin countries, the proportion of immigrants from poorer non-European origins that have low levels of education is 0.48, compared to only 0.23 for immigrants from richer origin countries. The respective proportion of immigrants with high levels of educational attainment is only 0.24 for those from poorer compared to 0.45 for those from richer non-European countries. This skill gap is somewhat weaker for immigrants from European origins. However, depending on the cut-off point, the differences are still large in substantive terms.

Finally, we have computed the proportions of immigrants (arriving in Europe) with low and high education levels for each of the origin countries. In line with our assumption, we expect to see that on average the proportion of low (high) skilled immigrants is falling (rising) with higher levels of an origin country's GDP per capita. This is exactly what we find: In the case of immigrants from European origins, the correlation between origin GDP per capita and the proportion of low education immigrants is negative 0.22; the correlation is positive 0.16 between origin GDP per capita and the proportion of high education immigrants. This pattern is even more pronounced in the case of immigration from non-European origins, where the respective GDP correlations are negative 0.49 (0.55 for logged GDP per capita) for the proportion of low education immigrants and positive 0.72 (0.70) for the proportion of high education immigrants.

This rising skill gap is most clearly demonstrated in the "scissoring" of the lines of best linear fit (as well as the dashed lowess fitted lines) in Figures 1 and 2, where we have plotted the low/high skilled proportions against origin country GDP per capita. Note that as expected, the scissoring pattern is most dramatic in the case of immigration from non-European origins. Clearly, the richer the country from which immigration originates, the higher (lower) the proportion of high (low) skilled immigrants we observe.

[Figures 1-2]

Figure 3 simply displays the same plot for the log transformed GDP per capita of the non-European origin countries. The log transformation yielded the best results in terms of approximating a normal distribution of GDP per capita among these nations (compared to other transformations down and up the ladder of powers). Here the scissoring pattern emerges even more clearly. Note that, in the case of European origin countries, the untransformed GDP per capita distribution performs best compared to other transformations down and up the ladder of powers, so we do not need to perform the additional analysis for this sub-sample (see Appendix A for qq.normal plots of GDP per capita in both logged and unlogged forms).

[Figure 3]

E. Conclusion

Overall, the data strongly confirm the assumption that immigrants coming to Europe from richer countries are on average significantly more skilled than immigrants from poorer countries. This holds true for immigration from European and non-European countries, although the skill gap is more pronounced in the latter case. This general result is robust across all tests we compute.

To be sure, in order to gain more confidence in this general finding, further data would be desirable, including more destinations, more origins, and in particular an inclusion of illegal immigration (note that exclusion of the latter most likely biased our estimates of the skill gap towards zero). To our knowledge such data is currently unavailable. Nonetheless, we believe that the extensive empirical evidence presented here is sufficient to shift the burden of proof to those that might challenge our assumption.

Table 1: Definitions of Richer and Poorer Origin Countries

European Origin Countries*			
Descriptive Statistics		Sample Cutoff-Points (according to respective definition of Richer/Poorer)	
	GDP per capita 2001*	Richest of the "Poorer Countries"	Poorest of the "Richer Countries"
Mean	22445	Italy	UK
Median	26049	UK	Ireland
Quantiles			
0.10	1596		
0.30	11603	Malta	
0.40	17595	Greece	
0.60	31218		Belgium
0.70	31843		Iceland
0.90	38504		
SD	15575.		
N	26		

Non-European Origin Countries*			
Descriptive Statistics		Sample Cutoff-Points (according to respective definition of Richer/Poorer)	
	GDP per capita 2001*	Richest of the "Poorer Countries"	Poorest of the "Richer Countries"
Mean	7962	Argentina	Cyprus
Median	2774	Former Russia**	Thailand**
Quantiles			
0.10	495		
0.30	1276	Egypt	
0.40	2184	Algeria	
0.60	3222		Mexico
0.70	4523		Brazil
0.90	23754		
SD	11577		
N	25		

* PPP adjusted constant 1995 US dollars
** Turkey is the median and thus assigned to neither country group for this cut-off point. Russia is the richest country below and Thailand the richest country above the median.

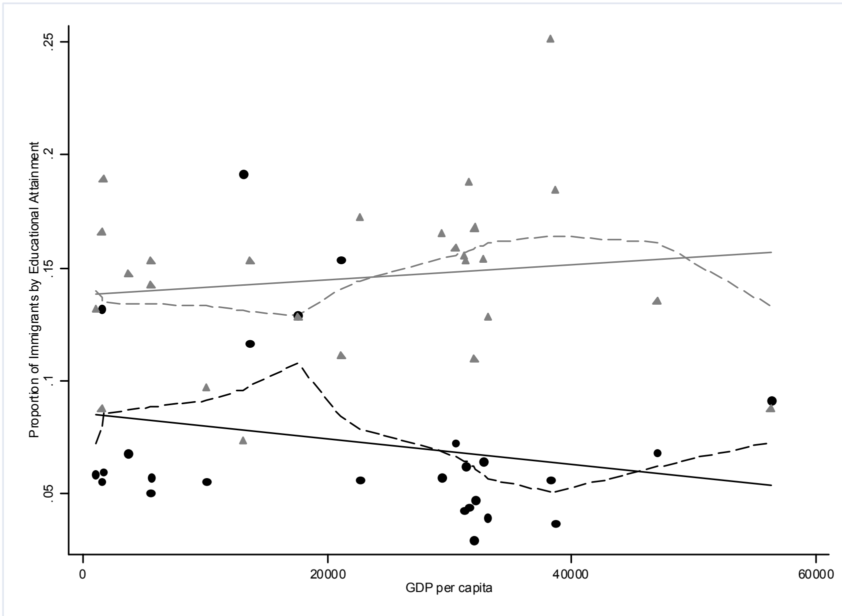
Table 2: The Skill Levels of Immigrants from Richer vs. Poorer Countries

European Origin Countries											
Sample cut-off point	"Richer Countries"				"Poorer Countries"				Difference:[1] Richer-Poorer Countries		
	Average Educational Attainment			N	Average Educational Attainment			N	Average Educational Attainment		
	Low	Middle	High		Low	Middle	High		Low	Middle	High
Sample Mean	0.286	0.384	0.330	187	0.487	0.334	0.179	133	**-0.201*** **	**0.050** **	**0.151*** **
Sample Median	0.293	0.401	0.306	174	0.423	0.333	0.244	146	**-0.130*** **	**0.068** **	**0.062** **
40th / 60th Quantile	0.284	0.424	0.292	148	0.400	0.360	0.240	106	**-0.116*** **	**0.064** **	**0.052* **
30th / 70th Quantile	0.293	0.428	0.279	110	0.367	0.400	0.233	81	**-0.074* **	**0.028**	**0.046**

Non-European Origin Countries											
Sample cut-off point	Richer Countries				Poorer Countries				Difference:[1] Richer-Poorer Countries		
	Average Educational Attainment			N	Average Educational Attainment			N	Average Educational Attainment		
	Low	Middle	High		Low	Middle	High		Low	Middle	High
Sample Mean	0.212	0.307	0.481	101	0.500	0.293	0.207	209	**-0.288*** **	**0.014**	**0.274*** **
Sample Median	0.227	0.319	0.454	160	0.481	0.280	0.239	137	**-0.254*** **	**0.039**	**0.215*** **
40th / 60th Quantile	0.219	0.317	0.464	145	0.493	0.277	0.230	112	**-0.274*** **	**0.040**	**0.234*** **
30th / 70th Quantile	0.220	0.313	0.467	123	0.414	0.300	0.286	85	**-0.194*** **	**0.013**	**0.181*** **

1. Differences computed using two-sample *t* tests (two tailed) with unequal variances. * p<0.10 ** p<0.05 *** p<0.01.

Figure 1: The Proportion of High- vs. Low-Educated European Immigrants by GDP per capita of Origin Country

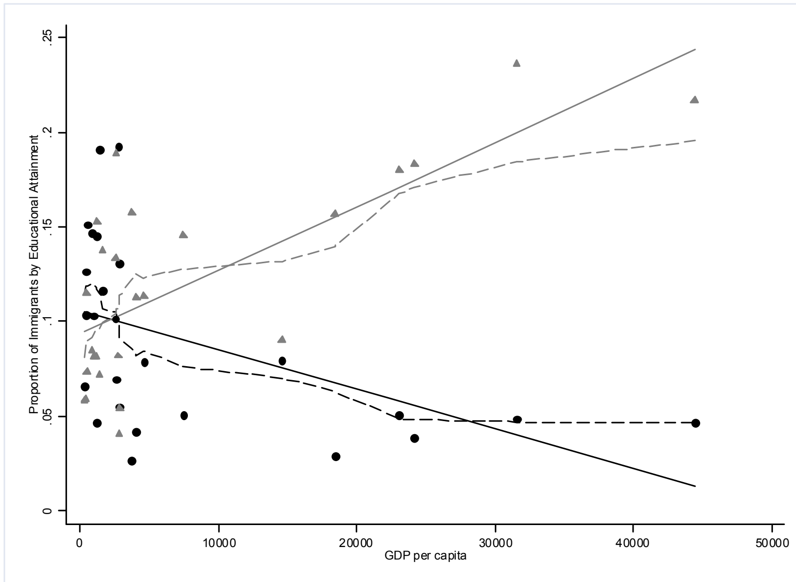


Legend:

Gray lines and (triangles) plot proportion of immigrants with **high** education. Black lines (and dots) plot proportion of immigrants with **low** education.

Solid lines are linear lines of best fit. Dashed lines are lowess fitted lines (bandwidth 0.9).

Figure 2: The Proportion of High- vs. Low- Educated Non-European Immigrants by GDP per capita of Origin Country

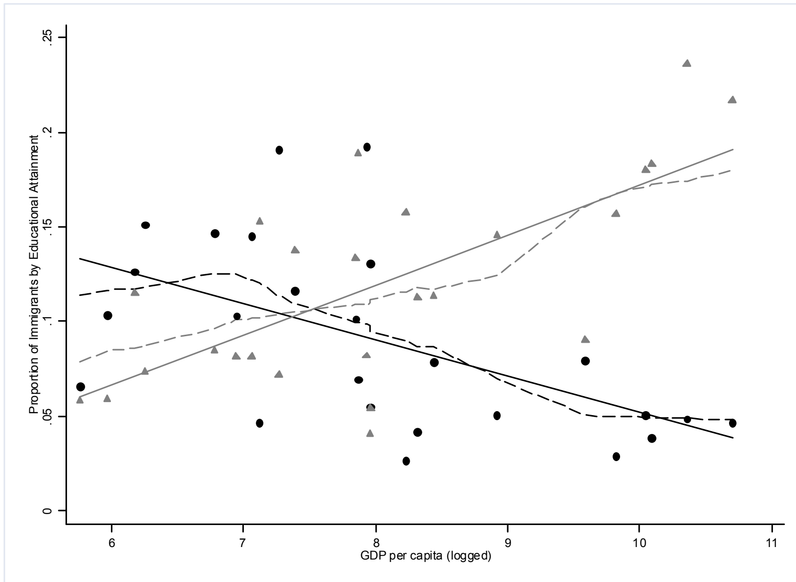


Legend:

Gray lines (and triangles) plot proportion of immigrants with **high** education. Black lines (and dots) plot proportion of immigrants with **low** education.

Solid lines are linear lines of best fit. Dashed lines are lowess fitted lines (bandwidth 0.9).

Figure 3: The Proportion of High- vs. Low- Educated Non-European Immigrants by Logged GDP per capita of Origin Country



<u>Legend:</u>

Gray lines (and triangles) plot proportion of immigrants with **high** education. Black lines (and dots) plot proportion of immigrants with **low** education.

Solid lines are linear lines of best fit. Dashed lines are lowess fitted lines (bandwidth 0.9).

Appendix A: QQ.normal Plots for the GDP Distributions of the Origin Sub-Samples
(95 % dashed confidence envelope around solid fitted line)