The Effect of Immigration on Wages: Exploiting Exogenous Variation at the National Level

Joan Llull*§

MOVE, Universitat Autònoma de Barcelona, and Barcelona GSE

PRELIMINARY AND INCOMPLETE.
PLEASE ASK FOR A MORE RECENT VERSION BEFORE QUOTING.

This version: April 2013

Abstract.— This paper estimates the effect of immigration on wages at the national/cross-skill level taking into account that immigrants are not allocated randomly across skill groups. I use a twofold argument to find exogenous variation of immigrant flows: wars and changes in political regimes push relatively more migrants to physically closer countries than to countries that are further away; and this is especially true for unskilled workers, as they are financially more constrained. The analysis indicates that immigration drastically reduced wages of competing workers: a 10 percent increase in supply reduces wages by about a 10 percent. This reduction is estimated to be two to three times larger than the previous studies that assumed immigration to be exogenous.

1. Introduction

With the recent resurgence of large scale immigration into OECD countries in recent years, economists have long tried to asses whether (and by how much) immigration affects wages of native workers. This recent immigration wave has attracted so much attention in part because of its magnitude, and in part because of its composition (Card, 2009). Despite the big effort, however, there is no consensus on what are the consequences for wages of such inflows.

In order to analyze the effect of immigration on wages, the literature has searched ways to compare the evolution of wages in similar labor markets that are exposed to different immigration shocks.[1] Early studies defined these labor markets geographically,

* MOVE. Universitat Autònoma de Barcelona. Facultat d'Economia. Bellaterra Campus – Edifici B, 08193, Bellaterra, Cerdanyola del Vallès, Barcelona (Spain). URL: http://pareto.uab.cat/jllull. E-mail: joan.llull [at] movebarcelona [dot] eu.

§I wish to thank Manuel Arellano, George Borjas, Stéphane Bonhomme, Julio Cáceres-Delpiano, Nezih Guner, Tim Hatton, Jenny Hunt, Stephan Litschig, Enrique Moral-Benito, and participants at the XXXVII SAEe in Vigo, the I Barcelona GSE Winter Workshop in Barcelona, and the UCL-Norface Conference on Migration: Global Development, New Frontiers for helpful comments and discussions. Christopher Rauh provided excellent research assistance. Financial support from European Research Council (ERC) through Starting Grant n.263600, and from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075), is gratefully acknowledged.

[1] Friedberg and Hunt (1995), and Borjas (1999) provide good surveys of the literature.

mainly as metropolitan areas.[2] More recent papers, pioneered by Borjas, Freeman and Katz (1997) and Borjas (2003), identify the effects at the national level, defining labor markets in terms of skills.[3]

The main problem of this cross-labor market comparison is that immigrants are not randomly allocated across labor markets. Because labor migration is mainly an economic decision, those markets that experience positive wage shocks will attract more immigrants, thus generating a positive correlation between immigration and wages that may bias upward the estimates of the effect of immigration. In order to correct for this endogeneity bias, Altonji and Card (1991) used settlement patterns of previous immigrants as an instrument for current inflows assuming that they are not correlated with current labor demand shocks. This approach has been followed by many papers in the literature.[4] However, it has been also hardly criticized (e.g. Borjas, 1999) because any source of persistence in wage shocks breaks the exogeneity of the instrument.

At the national level, the endogenous allocation of immigrants across skill groups has not been handled. Despite the fact that many papers have studied the self-selection patterns of immigrants in terms of skills (notably Borjas, 1987b; Chiquiar and Hanson, 2005), cross-skill comparisons in the literature assume that immigrants are exogenously allocated across skill groups.

In this paper, I tackle endogeneity at the national level. In particular, I use an alternative source of exogenous variation to identify the effect of immigration on wages. In particular, this variation comes from a double finding. First, as I show in **?**, the effect of push and pull factors on migration prospects are heterogeneous depending on the distance between the two countries. For example, a war in Algeria increases flows of Algerians to France by a higher fraction than those to Canada. Hence, I use exogenous variation in push factors (wars and changes in political regimes in source countries) interacted with distance to the destination country in order to obtain variation over time and across destination countries. Second, in order to obtain variation across skill groups, I use the fact that this heterogeneity is more severe among low skilled workers (as they tend to have less resources, and distance turns out to be a more severe constraint when choosing a destination country) than among high skill workers.

And I exploit this differential effect of the exogenous variation in push factors across skill groups depending on distance to build an instrument for immigration with variation across skill groups and destination countries, and over time. To the best of my knowledge,

[2] A sample of papers using this approach include Grossman (1982), Borjas (1987a), Card (1990), Altonji and Card (1991), LaLonde and Topel (1991), Goldin (1994), Card (2001), Card and Lewis (2007), Saiz (2007), Cortés (2008), Cortés and Tessada (2011), and Dustmann, Frattini and Preston (2013) among many others.

[3] These studies include Borjas and Katz (2007), Aydemir and Borjas (2007), Borjas, Grogger and Hanson (2010), Llull (2010), Aydemir and Borjas (2011), Ottaviano and Peri (2012), and Manacorda, Manning and Wadsworth (2012) among many others.

[4] Including Card (2001), Card and Lewis (2007), Saiz (2007), Cortés (2008), Cortés and Tessada (2011), and Dustmann, Frattini and Preston (2013) among many others.

this is the first paper that tackles endogeneity of immigration at the national/cross-skill level. Moreover, I also show that this instrument can be a valid instrument to be used as an alternative to Altonji and Card (1991) for the geography based analysis if we consider large (distant) enough labor markets.

In the analysis below, I follow Borjas (2003) in exploiting the fact that U.S. immigration is distributed evenly across education and experience cells, and that this pattern is changing over time. I additionally use the fact that, as shown in Aydemir and Borjas (2007), the shape of immigrant inflows into Canada is very different from that into the U.S. Hence, I compare wages across skill cells and over time, and across the two countries using different approaches. The baseline estimates follow the fixed effects approach in Borjas (2003, Secs. II-VI), expanded to the two country setting, and, most importantly, using the previously described exogenous variation as a source of identification. Then I go deeper into the analysis by exploiting additionally the regional variation within each country. And, finally, I use the instrument to estimate the nested CES production function that has become popular in the literature[5] and simulate the effect of immigration on wages over the last decades.

In order to use this instrument, I have to face an additional difficulty. Proper identification of the first stage coefficients requires a sample of destination countries that are separated enough to have enough variation in distances to each origin country. However, because of wage microdata availability, the analysis below restricts to Canada and the US (which are not very distant countries). To circumvent this, I estimate the first stage with a larger set of countries (including many European countries, the US, and Canada), and then I use the predicted instrument in the restricted sample for the second stage. I show below that only very mild additional assumptions are needed to restrict the sample for the second stage without producing any bias in the estimates.

Preliminary results suggest that immigration reduced wages very drastically. Indeed, I consistently find that this reduction in wages is two to three times larger than the reduction found in Borjas (2003) (which itself is larger than many others in the literature). In particular, I estimate the monthly wage elasticity to immigration to be around -1, meaning that a 10 percent supply shock reduces monthly wages by about 10 percent.

The rest of the paper is organized as follows. Section 4 explains the econometric model that I estimate, handling the difficulties produced by data limitations as I have just argued. Section 3 presents the different sources of data I use, and provides a short description of some facts. Sections 5 to 7 present the results from estimation using the different approaches described above, and Section 8 concludes.

[5] Inspired in Card and Lemieux (2001), this nested CES was introduced in the immigration literature in Borjas (2003, Sec. VII), and then used in Borjas and Katz (2007), Aydemir and Borjas (2007), Borjas, Grogger and Hanson (2010), Ottaviano and Peri (2012), and Manacorda, Manning and Wadsworth (2012) among others.

2. Exogenous Variation of Immigration at the National/Cross-Skill Level

When measuring the effect of immigration on wages, an ideal approach would be to compare wages in a particular labor market, with the wages that we would have observed in that market in the absence of immigration. However, such counterfactual wages are not observed in practice. As a result, the literature has exploited the variation across different labor markets claiming that (after including the necessary controls) these markets are equal except for the fact that some received immigrants and some did not.

A typical paper defines different labor markets in terms of skills, geographic regions, and/or time. The way in which they compare wages across those labor markets can be summarized as

$$\ln w_s = \vartheta p_s + \boldsymbol{x}_s' \boldsymbol{\phi} + \upsilon_s, \tag{1}$$

where $\ln w_s$ is the log wage in labor market s; $p_s = M_s/(M_s + N_s)$ is the fraction of the workforce in that labor market that is an immigrant; $\boldsymbol{x}_s = (x_{1s}, ..., x_{Hs})'$ is a vector of control variables that may include period fixed effects, region fixed effects, skill fixed effects, a combination of them, and/or any other variables that generate differences in wage levels across labor markets; and υ_m is an i.i.d. error term with with zero mean and variance σ^2 (Aydemir and Borjas, 2011). The wage elasticity to immigration can be obtained as $\epsilon = \theta/(1+m)^2$, where $m = \bar{p}/(1+\bar{p})$ is the (average) ratio of immigrants over natives (Borjas, 2003).

The general problem is that immigrants are not allocated randomly across labor markets. Almost by construction, as immigrants are moving in search of better economic opportunities, they are going to penetrate those markets that are relatively more rewarding to them. This will build a positive correlation between υ_s and p_s that will bias the OLS estimator of ϑ upward.

This problem has been discussed extensively in the context of labor markets geographically defined (as metropolitan statistical areas). It is hard to assume that immigrants are allocated randomly across cities: they will tend to go to cites that experience positive wage shocks. Altonji and Card (1991) proposed to use settlement patterns of previous immigrants as an instrument for current inflows assuming that they are not correlated with current labor demand shocks. Since then, different versions of this approach have been used in almost every area based study in the literature (see Card (2001), Card and Lewis (2007), Saiz (2007), Cortés (2008), Cortés and Tessada (2011), and Dustmann, Frattini and Preston (2013) among many others).

Despite being widely used in the literature, this instrument has been hardly criticized as well (Borjas, 1999). Its main drawback is that any kind of persistence in labor demand shocks will create an association between current wage shocks and past immigrant inflows that will invalidate the exogeneity assumption of the instrument.

Other instruments, on the other hand, have been hard to find. Very few exceptions (including Card (1990), Hunt (1992), Friedberg (2001), and Glitz (2012)) are able to use quasi-experimental evidence to find exogenous variation in immigrant flows across regions. These exercises find very credible sources of exogenous variation, but its usage (and results as well) are limited to the context of the particular experiment at hand.

The second major strand of the literature switches the analysis at the national level. Starting with Borjas, Freeman and Katz (1997), and mostly after Borjas (2003), a growing fraction of the papers in the literature define labor markets in terms of skills (e.g. Borjas and Katz (2007), Aydemir and Borjas (2007), Card (2009), Borjas, Grogger and Hanson (2010), Llull (2010), Aydemir and Borjas (2011), Ottaviano and Peri (2012), Manacorda, Manning and Wadsworth (2012)). The endogeneity issue, however, has not been tackled in this literature. The common assumption of all these papers is that immigration into different skill cells is random.[6]

This assumption, however, is at odds with the self-selection argument introduced by Borjas (1987b). The argument states that immigrants tend to self-select into those skill groups that reward their skills the best as compared to their home countries. In the context of cross-cell wage comparisons, this implies that a demand shift in a particular skill cell will tend to change the composition of the inflow of immigrants (even for a given total inflow). Therefore, it will again create a positive correlation between the inflow of immigrants and wages in a particular labor market (skill cell), thus leading to underestimating the effect of immigration on wages.

In this paper, I propose an alternative source of exogenous variation at the national / cross-skill level. This variation comes from the combination of different sources. First, I exploit exogenous variation at the origin country: wars and political regime changes.[7] These factors alone do not provide variation neither across destination countries, nor across skill cells in a given country. To gain variation across destination countries, I interact them with distance between the two countries. For instance, a war in Algeria will tend to push more people to France than to Australia.[8] To obtain variation across skill cells for a given destination country I explore the fact that the way in which distance

[6] Dustmann, Frattini and Preston (2013) also could be included in this literature. In this paper, labor markets are defined as positions in the distribution of native wages. This variation is exploited on top of the cross-city variation. Instrumental variables (Altonji and Card (1991) type of instrument) are used, but only to correct cross-city endogeneity. The position of the distribution at which immigrants penetrate is still assumed to be exogenous.

[7] Throughout the paper, I will refer to political regime changes as changes in the level of democracy or autocracy of the country. Therefore, it does not necessarily refer to a change of the official political regime (e.g. a change in the Constitution), but, rather, a change in the "quality" of the political institutions. I describe how this is measured in Section 3.

[8] Llull (2013) presents evidence suggesting that the effect of push and pull factors on migration flows differ by distance. Llull (2011) exploits this source of exogenous variation to instrument immigrant stocks across countries in the estimation of cross-country GDP per worker regressions. Previously, Angrist and Kugler (2003) used distance to instrument changes in immigrant stocks across European countries over the different stages of the Balcans War when analyzing the effect of immigration on employment of EU natives.

Table I.—Distribution of Migrants Across Some OECD Countries After Selected Conflicts (1990-2000)

	Total	Primary	Secondary	Tertiary	Closest distance (km)	
A. Balkans War						
Australia/New Zealand	3.01	3.87	-6.71	8.38	15,689	Australia
Europe	77.99	90.16	64.68	58.99	492	Austria
U.S./Canada	19.00	5.97	42.03	32.63	7,266	United States
B. African Conflicts	Total	Primary	Secondary	Tertiary	Closest distance (km)	
Australia/New Zealand	2.07	2.40	0.78	2.69	12,636	Australia (vs Ethiopia)
Europe	66.21	74.86	75.63	55.97	707	Spain (vs Algeria)
U.S./Canada	31.72	22.73	23.59	41.33	6,472	US (vs Algeria)
C. Middle East Conflicts	Total	Primary	Secondary	Tertiary	Closest distance (km)	
Australia/New Zealand	11.29	16.70	8.67	8.97	11,433	Australia (vs Afghan.)
Europe	43.60	42.43	56.46	38.55	1,155	Greece (vs Lebanon)
U.S./Canada	45.11	40.88	34.87	52.48	9,032	US (vs Lebanon)

Note: Figures in this table represent the percentage of total net flow (changes in stocks) of migrants from the countries affected by each conflict that migrated to the corresponding group of destination countries: $m_k = \sum_q (M_{kq,2000} - M_{kq,1990}) / \sum_k \sum_q (M_{kq,2000} - M_{kq,1990})$, where k indicates a group of selected countries (Europe, Australia/New Zealand, US/Canada), q indicates origin country (among those affected by the corresponding conflict), and M_{kqt} is the stock of immigrants from country q in the group of countries k in year t. European destination countries include EU-15 (excluding Luxembourg and Ireland), Norway, and Switzerland. Balkans War affected the countries of the former Yugoslavia. African conflicts include Algeria, Angola, Burundi, former Zaïre, Ethiopia/Eritrea, Liberia, Mali/Niger (Touareg rebellion), Rwanda, Sierra Leone, and Sudan wars in the 1990s. Middle East conflicts comprise war episodes in Afghanistan, Iraq, Lebanon, Tajikistan, and Yemen. *Data sources*: immigrant stocks by educational attainment are obtained from Docquier and Marfouk (2006) and conflicts are identified from Gleditsch, Wallensteen, Eriksson, Sollenberg and Strand (2002).

maps push factors into actual flows differ across skill cells. For instance, more skilled individuals tend to be financially less constraint, and, hence, when they are pushed out from their country by a war or a political regime change, distance tends to matter less for them when selecting their destination country. Or middle-aged individuals tend to carry a larger number of dependent family members with them, which makes distance (and any other individual-specific moving cost) more binding.

Tables I and II illustrate these facts with some casual evidence. Table I analyzes migration patterns after a selected set of conflicts during 1990s. The conflicts that are analyzed are Balkans War, several conflicts in Africa, and some conflicts in the Middle East. The table presents the distribution of migrants from the countries affected by each of these sets of conflicts across nineteen OECD destination countries (grouped by continent: North America, Europe, and Oceania). This distribution is computed unconditionally, and also conditional on different educational levels.

The countries involved in Balkans War and in some of the African conflicts being considered are close to Europe and far from Oceania and North America. Middle East countries, on the other hand, are far from any of the three groups of destination countries.[9] European countries received more than 75% of all the immigrants that left the former Yugoslavia during 1990s and around 66% of the immigrants from the African countries involved in conflicts. This fraction seems pretty large both when compared to the 43% of the immigrants from countries with conflicts in the Middle East, and to the 34% they received from the total net flow from all countries in the world into these nineteen destination countries over this period. This result suggests that distance plays a role in the determination the distribution of migrants across destination countries when they are pushed out from their countries of origin by some conflict.

This distribution, however, is not homogeneous across educational levels. From the flows potentially generated by Balkans War and the African conflicts, primary educated are overrepresented in Europe. The most extreme case is Balkans War, from which 90% of the displaced primary educated went to European countries, whereas they only received 59% of the tertiary educated. Likewise, these European countries received 75% of low educated migrants from the selected African countries, as opposed to the 56% of highest educated. From the Middle East countries, on the other side, the received fractions were 42% and 39%, much more similar. And the fractions received of the flow from countries all over the world was 28% of primary educated, and 29% of tertiary educated. These differences constitute suggestive evidence of a differential effect of distance across educational groups when a conflict in a country forces its inhabitants to move abroad.

To further motivate the analysis, in Table II, I explore the correlation between migration flows pushed by wars/political regime changes and distance, for different education and experience cells. Given data availability, I consider the U.S. as the destination country. The coefficients presented in the table are the estimated slopes of a set of regressions of the decade change in the fraction of individuals of a given sub-population that is from country of origin q in a given Census, on the distance between country q and the United States. These regressions are estimated for different sets of countries of origin and for different sub-populations. Different columns in Table II consider different countries of origin: left column includes countries that experienced a conflict in the preceding decade; center column includes countries for which the level of democracy or of autocracy increased by more than three points (see data definitions in Section 3); and the right column excludes Mexico. In the top panel, sub-populations are defined by education, and in the bottom panel, by (potential) experience.

Results from the top panel of the Table II point in the same direction as Table I: distance seems to matter more for low educated than for highly educated. This would

[9] The distance between Lebanon and Greece is only slightly over a 1,100 kilometers. However, the distance between Lebanon and Italy, the second closest European country, is around 2,200 kilometers. Similarly, Iraq, the second closest origin country, is around 2,000 kilometers away from Greece, and almost 3,000 from Austria, its second closest destination country. Since distance will be included in the analysis below in logs, going from 500 kilometers to 2,500 is a "larger" increase than from 2,500 kilometers to 10,000.

Table II.—Correlation Between Distance and Migration to the U.S. After Selected Push Factors

A. By Education

	Conflicts		Political regimes		Political regimes (excluding Mexico)	
Primary	-0.146	(0.061)	-1.514	(0.738)	-0.231	(0.079)
Secondary	-0.017	(0.007)	-0.210	(0.101)	-0.027	(0.008)
Tertiary	0.019	(0.014)	-0.036	(0.023)	0.003	(0.007)

B. By Experience

	Conflicts		Political regimes		Political regimes (excluding Mexico)	
0-7 years	-0.019	(0.014)	-0.177	(0.097)	-0.036	(0.015)
8-15 years	-0.029	(0.015)	-0.420	(0.208)	-0.051	(0.020)
16-23 years	-0.025	(0.012)	-0.366	(0.185)	-0.041	(0.015)
24-31 years	-0.015	(0.010)	-0.203	(0.097)	-0.031	(0.011)
31+ years	-0.010	(0.009)	-0.151	(0.075)	-0.016	(0.010)

Note: The table reports the estimated coefficients of distance from the following regression fitted on different samples:

$$\Delta m_{qt} = \beta_0 + \beta_1 dist_{qt} + u_{qt},$$

where q indicates country of origin, t indicates Census year, $dist_q$ is the distance between country q and the U.S., and m_{qt} is the period t fraction of the workforce (with the given educational or experience level) that is from country q. Each coefficient from the table is obtained running this regression on different samples. Left column coefficients are estimated with a sample of origin countries that experienced some conflict in the preceding decade. Center column is estimated with a sample of origin countries that experienced an increase of democracy or autocracy indexes of more than three points (data description in Section 3). Right column excludes Mexico from the previous sample. By rows, top panel coefficients are estimated with the sub-samples of primary or less, secondary, and tertiary educated respectively, and those from the bottom panel are estimated with samples of individuals with the corresponding experience. Robust standard errors in parenthesis.

be consistent with low skilled being financially constraint, and, hence, when pushed out from the country by a conflict or a political regime change, not able to afford moving to a country which is further away (even when that country pays higher wages). The bottom panel shows that distance matters more for the intermediate levels of experience (i.e. for middle-aged individuals). In these intermediate ages, individuals tend to have a larger number of dependent family members to carry with them. The larger moving costs generated by these dependent family members would make distance more relevant for them when choosing a destination after being pushed out from their origin country.

3. Data

The empirical analysis of this paper uses data drawn from country-specific censuses for different years. Data are provided by IPUMS-International (Minnesota Population Center, 2011). A more detailed exposition on data construction and variable definitions is presented in Appendix A below.

The pool of destination countries used in the first stage of the baseline estimation is a balanced panel that include Austria, Canada, France, Greece, Ireland, Switzerland, and the U.S., for years around 1970, 1980, 1990, and 2000. Additionally, for some specifications, I include additional countries (the Netherlands, Italy, Portugal, and Spain) and additional dates (around 1960 and 2005), forming an unbalanced panel. Table III summarizes available years and sample sizes for each of the countries. Although sample sizes are quite large, they are, except for the U.S., too small to accurately compute the fraction of immigrants in each education-experience cell for each country of origin. This issue, as discussed in Section 4, is addressed by estimating the first stage regressions at a more aggregate level.

Table III.—Sample Sizes from Different Censuses

	1960	**1970**	**1980**	**1990**	**2000**
Austria	—	**291,466**	**320,133**	**355,905**	**387,056**
Canada	—	**80,948**	**231,020**	**412,484**	**407,716**
France	892,533	**952,903**	**1,127,204**	**1,039,352**	**1,319,045**
Greece	—	**280,261**	**315,399**	**349,839**	**424,772**
Ireland	—	**76,435**	**112,943**	**127,962**	**169,481**
Italy	—	—	—	—	1,258,993
Netherlands	—	72,106	—	—	94,793
Portugal	—	—	183,285	196,666	226,818
Spain	—	—	691,692	752,515	855,531
Switzerland	—	**130,136**	**138,680**	**172,980**	**178,030**
United States	645,004	**758,760**	**4,971,227**	**5,855,980**	**6,435,700**

Note: The table reports the number of observations from each Census microdata file used to compute immigrant shares. The balanced panel used in the baseline specification is in bold. These data include working male and female. The fraction of female among them range from 24 to 48% depending on the sample.

These samples include active individuals (employed and unemployed) aged 18 to 64. For each sample, I compute the share of immigrants by education-experience cell. A person is defined to be an immigrant either if she was born abroad or if she is a foreign citizen, depending on data availability. Education is categorized into three internationally comparable categories: primary or less, secondary, and tertiary. To compute experience, I assume that the age of entry into the labor market is 16 for primary educated, 18 for secondary educated, and 23 for tertiary educated; then, I measure work experience as the current age of an individual minus the (assumed) age at which she entered in the labor market. Experience is categorized into five groups: 0-7 years of experience, 8-15, 16-23, 24-31, and 32 or more years. For the baseline estimates, I focus on male workers, who account for 52% to 76% of the observations listed in Table III (depending on countries and years).

Wage data is only available in Canadian and U.S. Censuses. Therefore, as discussed in Section 4 below, I use a two-sample IV approach. In particular, the first stage regression is estimated with the whole sample of countries included in Table III, and the second stage regression is estimated with data from these two countries only. Average wages by education-experience-country-year cell are computed using the sample of persons who worked a positive number of weeks during the preceding calendar year, reported positive annual earnings, are not enrolled in school, and are employed in the wage and salary sector. Monthly and annual average wages are computed.

To estimate the first stage regressions, I consider 188 countries of origin. Three *push* variables are used: conflicts, democracy level, and autocracy level. The data for the instruments come from several sources. The conflict variable measures the number months during the decade preceding the Census date that the country was involved in any type of war or internal conflict (normalized to be interpreted as a fraction). This variable is constructed using data from Gleditsch, Wallensteen, Eriksson, Sollenberg and Strand (2002), which list all armed conflicts in the world since 1946.[10]

Figure I.—Years of Conflict 1950-2000

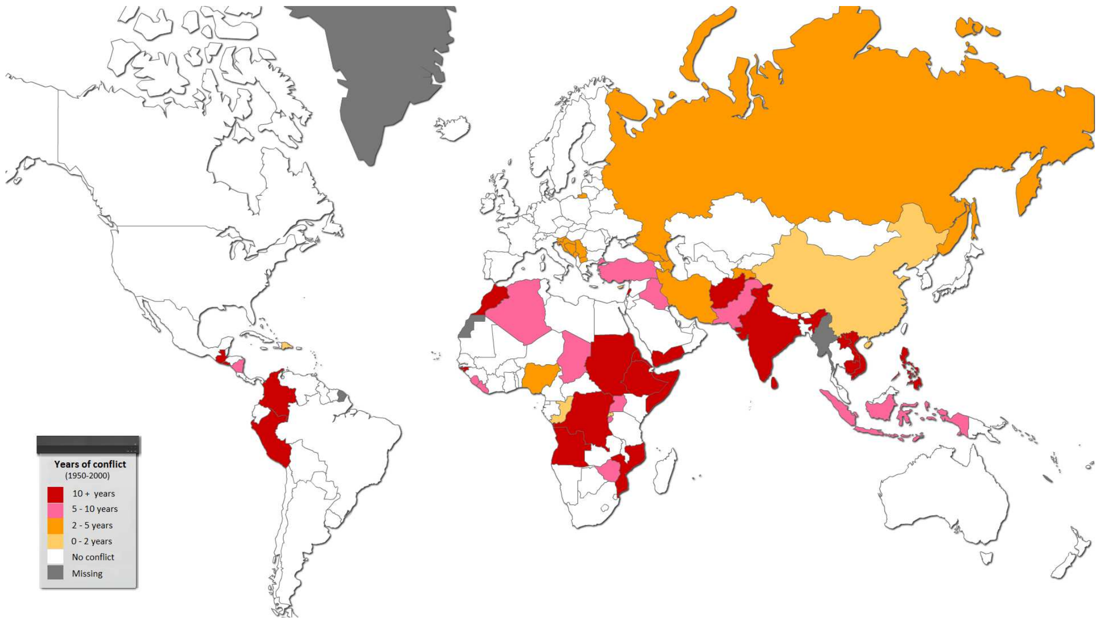


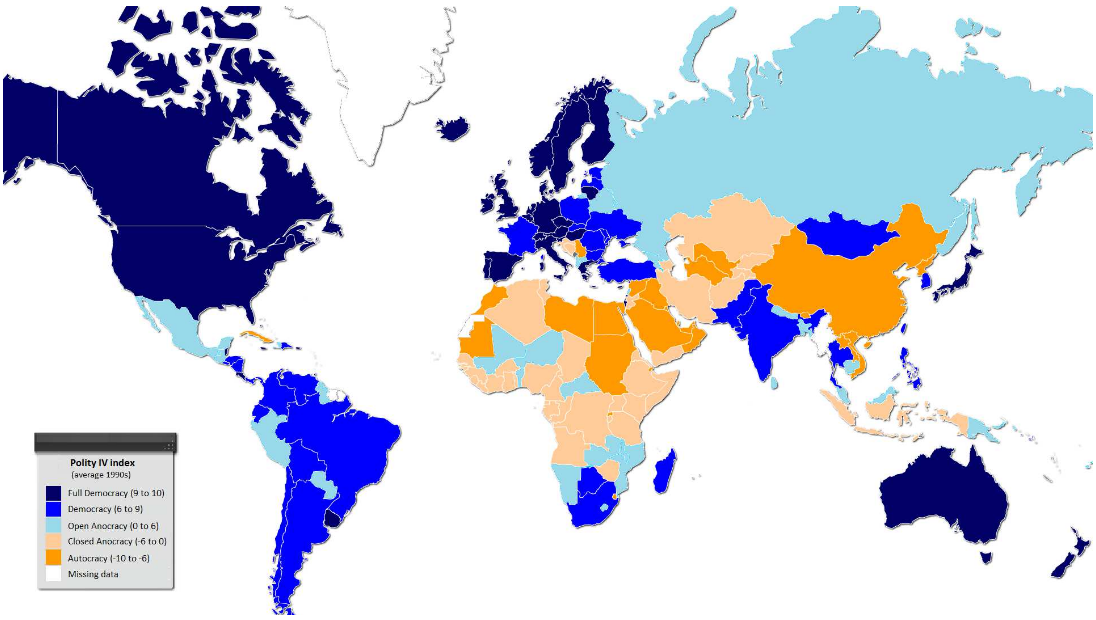
Figure I summarizes this variable. In particular, the figure plots the number of months of conflict (piled in years, and grouped into different categories) suffered by each of the countries from the sample of countries of origin. As expected, Africa and Southern Asia are the regions that were hit more severely. Yet, there is plenty of within region variation; additionally, given the average duration of a conflict, we can also see plenty of within country variation over time.

Democracy and autocracy variables are obtained from the Polity IV index (Marshall, Jaggers and Gurr, 2010). The Polity IV index ranges from -10 (autocracy) to 10 (democracy). Regimes with values around 0 are called anocracies, i.e. countries where power is not vested in public institutions (as in a normal democracy) but spread amongst elite groups who are constantly competing with each other for power.

[10] An updated version is available at the Peace Research Institute Oslo web page: `http://www.prio.no/cwp/armedconflict/`

Figure II plots average Polity IV indexes for all countries in the sample during 1990s. Most anocratic countries are in central Africa, whereas the Middle East and North Africa account for most of the autocratic regimes. All developed countries had fully democratic regimes. Latin American countries had a lower level of democracy, which turned into open anocracies in some cases. Asian countries display a wide variety of political regimes, covering all the spectrum.

Figure II.—Average Polity IV Scores 1990-2000



Anocracy is not surprisingly the least stable system, as it is very vulnerable to disruption and armed violence. In this situation, individuals are more likely to move out from the country. A strong democracy offers its citizens stability, freedom and good economic prospects. Therefore, individuals tend to be more willing to stay in the country. And under an autocratic rule, individuals might be more willing to leave the country, but cross-border movements are usually limited by the autocratic power (e.g. North Korea).

As a result of all this, I construct two variables from the Polity IV index: democracy and autocracy levels. They are a spline of the index. Democracy level is the positive side (and equal to zero if the index is negative), and autocracy level is the (absolute value of the) negative side (and equal to zero if the index is positive). Hence, either of the two variables should negatively correlate with immigration.

Distance between two countries is defined as the physical distance between the center of the most populated cities of the two countries. Physical distance is computed using the Vincenty method (Vincenty, 1975).[11]

[11] This method is based on the assumption that the figure of the Earth is an oblate spheroid, and hence are more accurate than methods such as great-circle distance which assume a spherical Earth.

4. A Two-Sample Approach with Aggregate Data

In order to estimate the effect of immigration on wages using the general approach described in Section 2, I will define labor markets by educational level i, experience level j, (destination) country k, and Census year t. As a result, the specific version of equation (1) that is used as the benchmark (second stage) equation below is:

$$\ln w_{ijkt} = \theta p_{ijkt} + \eta_i + \pi_j + \delta_k + \tau_t + \xi_{ik} + \zeta_{it} + \gamma_{jt} + \iota_{kt} + \varepsilon_{ijkt}, \tag{2}$$

where $\ln w_{ijkt}$ is the average wage of workers in educational level i, with experience j, in country k at Census year t; $p_{ijkt} = M_{ijkt}/(M_{ijkt} + N_{ijkt})$ is the share of immigrants in that group; η_i, π_j, δ_k, τ_t, ξ_{ik}, ζ_{it}, γ_{jt}, and ι_{kt} are fixed effects; and ε_{ijkt} is the zero mean error term with $\mathbb{E}[p_{ijkt}\varepsilon_{ijkt}| \, \pi_j, \delta_k, \tau_t, \xi_{ik}, \zeta_{it}, \gamma_{jt}, \iota_{kt}] \neq 0$ in general, as argued in Section 2. In particular, this endogeneity is expected to be such that $\mathbb{E}[\hat{\theta}_{OLS}] > \theta$ provided by immigrants being prone to penetrate those cells with better economic opportunities.

Given the set of fixed effects included in equation (2), a valid instrument for p_{ijkt} needs to have variation across skill cells, across (destination) countries, and over time. Distance between countries cannot be a valid instrument in this context, as it is constant over time and across skill cells. Similarly, wars or political regimes in origin countries do not vary across destinations and skill cells. Interactions of distance and these push factors and distance would provide variation across destination countries and over time, but they are still constant across skill cells.

Suggestive evidence presented in Section 2 seem to imply that the effect of distance after a push shock is more severe for lower educated and middle-aged. For instance, a war in Algeria pushes more workers to France than to Australia; however, uneducated and middle-aged workers have a larger propensity to go to France, whereas, even though other workers may also go more often to France than to Australia, distance will be less important, and, if Australia is more attractive than France, they will have a larger propensity to go there. This heterogeneity allows us to add variation across skill cells on top of the variation across destination countries and over time generated by the interactions between push factors and distance.

The first stage regression, estimated at the bilateral level, would be as follows:

$$p_{ijqkt} = (\boldsymbol{z}_{qt} \ln d_{qk})' \boldsymbol{\alpha}_{ij} + \mu_j + \lambda_k + \varrho_t + \psi_{ik} + \varsigma_{it} + \varphi_{jt} + \kappa_{kt} + \nu_{ijqkt}, \tag{3}$$

where p_{ijqkt} is the stock of immigrants with education i and experience j from country q living in country k at Census year t; $\ln d_{qk}$ is the log of the physical distance between (origin) country q and (destination) country k; $\boldsymbol{z}_{qt} = (z^1_{qt}, z^2_{qt}, z^3_{qt})'$ is a vector of exogenous push factors in which z^1_{qt} is the number of months that country q was involved in any sort of conflict during the decade preceding Census year t, and z^2_{qt} and z^3_{qt} are the democracy and autocracy variables for country q constructed as described in Section 3 from average Polity IV indexes over the decade preceding Census year t; $\boldsymbol{\alpha}_{ij}$ is the 3×1 vector of coefficients associated to $\boldsymbol{z}_{qt} \ln d_{qk}$ for education-experience cell ij; μ_j, λ_k, ϱ_t, ψ_{ik}, ς_{it}, φ_{jt}, and κ_{kt} are fixed effects; and ν_{ijqkt} is a mean zero error term.

The two-stage estimator $\hat{\theta}_{2SLS}$ is provided by the least squares estimation of equation (2) replacing p_{ijkt} by $\hat{p}_{ijkt}$, where

$$\hat{p}_{ijkt} = \sum_{q=1}^{Q} (\boldsymbol{z}_{qt} \ln d_{qk})' \hat{\boldsymbol{\alpha}}_{ij} + Q\left(\hat{\mu}_j + \hat{\lambda}_k + \hat{\varrho}_t + \hat{\psi}_{ik} + \hat{\varsigma}_{it} + \hat{\varphi}_{jt} + \hat{\kappa}_{kt}\right). \tag{4}$$

The consistence of the estimator $\hat{\theta}_{2SLS}$ comes from the following assumption:

$$\mathbb{E}\left[\varepsilon_{ijkt} \sum_{q=1}^{Q} (\boldsymbol{z}_{qt} \ln d_{qk})' \boldsymbol{\alpha}_{ij} \;\middle|\; \pi_j, \delta_k, \tau_t, \xi_{ik}, \zeta_{it}, \gamma_{jt}, \iota_{kt}\right] = 0. \tag{5}$$

In other words, what we need is that the destination country–skill cell–period i.i.d. wage shock, ε_{ijkt} is uncorrelated, given fixed effects, with the the integrated bilateral instrument (integrated over countries of origin). The richness of fixed effects included in the model makes this assumption mild. For instance, this assumption does not imply that a war in Central America is uncorrelated with aggregate shocks to wages in the U.S. (as this correlation is absorbed by the destination country–period fixed effect); it neither implies that it is uncorrelated with shocks that are common to all experience groups within an education group or to all education groups within an experience group (as this is absorbed by experience–time and education–time fixed effects respectively); and so on.

Equations (2) to (5) fully characterize the model to be estimated in this paper. However, available data have two important limitations (described in Section 3) that require additional assumptions: the (un)availability of wages in several Censuses, and the relatively small sample sizes (too small to compute the share of immigrants by education–experience–period–destination country–origin country cells).

The availability of wage data is very limited in Census records. Only a few countries gather this information when collecting microdata for the Census. In particular, among the set of developed countries considered in this paper (Austria, Canada, France, Greece, Netherlands, Ireland, Italy, Portugal, Spain, Switzerland, and the U.S.), only the U.S. and Canada report them. The fact that U.S. and Canada are physically very close to each other, may complicate the identification (with enough precision) of the first stage coefficients.

The strategy I follow in this paper is a Two-Sample Two-Stage Least Squares approach. This approach estimates the first stage regression and the second stage with two different samples. Seminal work on two-sample IV methods is in **?** and in **?**. In the first stage sample we have data on the instrument and on the regressors, but not on the dependent variable. In this paper, this is the sample that includes U.S. and Canadian Censuses, and all the Censuses from the European countries listed above. The second stage sample includes information at least on the instrument and on the dependent variable. In the case of the present paper, this sample includes U.S. and Canadian Censuses only. With the first stage sample, the coefficients from equation (3) are estimated. Using the estimated coefficients, the predicted (exogenous part of) migration shares is computed for the second stage sample.

It is important to remark that the motivation for using two samples in the estimation is somewhat different from the original motivation in **?** and in **?**. In these papers, it is the unavailability of data for the regressor in the second stage sample what motivates the authors to do the estimation of the first stage with a different sample. In the present paper, the regressor is available in the second stage sample. Theoretically, θ could be consistently estimated using the second stage sample alone to estimate equations (3) and (2). However, in this case, the variation in distance would not be exploited itself to identify the first stage coefficients, and it would be only used to weight the effect of the exogenous push factors. Having a richer first stage sample, which includes countries that are far from each other, allows us to exploit the variation in distance on top of the variation in the exogenous push factors to precisely estimate the coefficients.

In order for this to produce consistent estimates of θ, the assumption in equation (5) needs to be narrowed. In particular, it would be replaced by:

$$\mathbb{E}\left[\varepsilon_{ijkt}\sum_{q=1}^{Q}(\boldsymbol{z}_{qt}\ln d_{qk})'\boldsymbol{\alpha}_{ij}\;\middle|\;\pi_j,\delta_k,\tau_t,\xi_{ik},\zeta_{it},\gamma_{jt},\iota_{kt},k\in\{US,CAN\}\right]=0. \tag{6}$$

In words, the exogeneity assumption in equation (5) should be valid conditional when restricting the sample to U.S. and Canada. This additional assumption does not seem to be a very restrictive, given the inclusion of destination country–period fixed effects. Nonetheless, to informally check for the validity of this extra assumption, I test whether the relationship between predicted and actual migration shares (net of fixed effects) is statistically different for U.S. and Canada when compared to the other countries included in the first stage only.

The second data limitation is regarding sample sizes. Although Census samples are quite large (see Table III), they are relatively small to compute the fraction of immigrants from each country of origin in a given education–experience–period cell. Even when this is possible for top countries of origin, these shares would be very imprecisely estimated for quantitatively minor sources of immigrants.

To circumvent this problem, I estimate the first stage regression with aggregate data (aggregated over origin countries). In other words, instead of estimating the first stage at the bilateral data, and then predicting the aggregate shares as in equation (4), regress the aggregated share of immigrants on an aggregation of the instrument over origin countries:

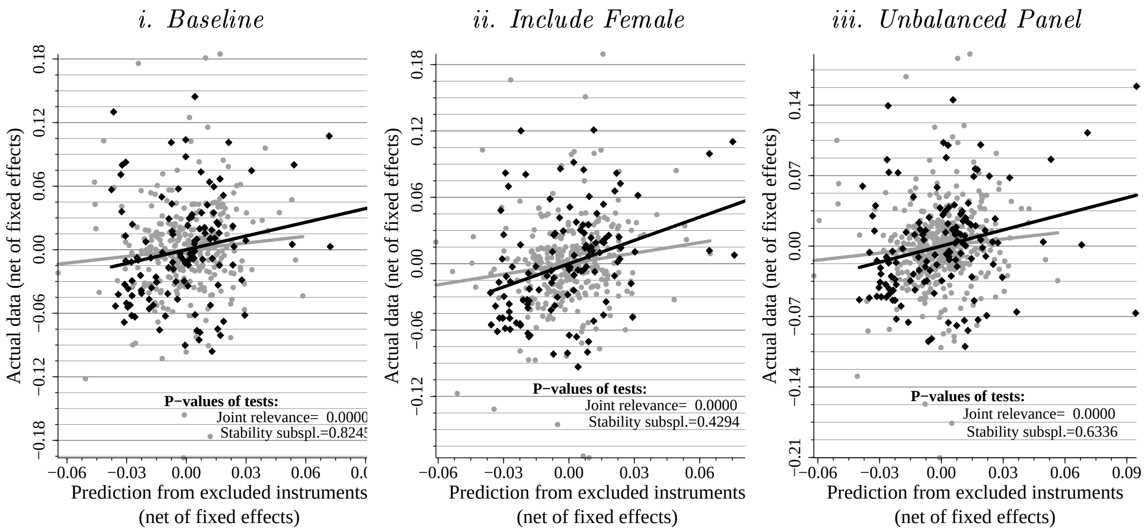
$$p_{ijkt}=\sum_{q=1}^{Q}(\boldsymbol{z}_{qt}\ln d_{qk})'\,\boldsymbol{\alpha}_{ij}+\tilde{\mu}_j+\tilde{\lambda}_k+\tilde{\varrho}_t+\tilde{\psi}_{ik}+\tilde{\varsigma}_{it}+\tilde{\varphi}_{jt}+\tilde{\kappa}_{kt}, \tag{7}$$

where the tilde denotes that the original fixed effects from equation (3) are multiplied by

Q. This estimation does not require further additional assumptions, as it is asymptotically equivalent to the estimation of the first stage using bilateral data.[12] The difference between the two approaches would be the precision with which $\boldsymbol{\alpha}_{ij}$ is estimated. On the one hand, estimating the first stage at the bilateral level would provide additional degrees of freedom, increasing the sample size by a factor of 188. However, immigrant shares at the bilateral level would be computed with important measurement error, which would reduce the precision of the estimates. Given sample sizes, it seems plausible that the second effect prevails. As a robustness check for this method, several specifications estimating the first stage aggregated only at the continent of origin level are estimated, obtaining very similar results.

5. Basic Results at the National Level

Figure III.—Actual and Predicted Immigrant Shares for Some First Stage Specifications



Note: Black: U.S. and Canada. Gray: Austria, France, Greece, Ireland, Switzerland, and, in the unbalanced panel, additionally, Netherlands, Italy, Portugal, and Spain. Left figure: baseline; center: includes females in labor force counts to compute the shares; right: unbalanced panel, which adds the four additional destination countries listed above for some periods, and year 1960 for some countries. The scatter diagrams relate the share of immigrants in each education-experience-period-country cell with the corresponding prediction using the instruments described in the text. Both actual and predicted shares are net of education, experience, country-period, education-period, experience-period, and education-country fixed effects. The lines represent the corresponding fitted regression for each group. Each observation is weighted by the number of observations used to compute the shares in each destination country/period. P-values of two specification tests are included in the bottom-right corner: an F test of the joint relevance of the excluded instruments, and a (Chow-type) F test of structural change between the two sub-samples. Sample selection and further details described in the main text.

[12] To see this, note that the moment conditions given by equation (5) or equation (6) are the same under the two approaches.

Table IV.—First Stage Coefficients for the Baseline Estimation

	Push Factor:					
	War (deaths)		Democracy		Autocracy	
Primary Education						
[0-8] years	—		—		—	
[9-16] years	0.240	(0.161)	0.018	(0.015)	0.006	(0.005)
[17-24] years	0.186	(0.120)	0.009	(0.010)	0.005	(0.004)
[25-31] years	0.132	(0.090)	0.003	(0.007)	0.003	(0.003)
32+ years	0.005	(0.183)	-0.007	(0.017)	0.001	(0.007)
Secondary education						
[0-8] years	-1.435	(0.420)	-0.096	(0.023)	0.064	(0.018)
[9-16] years	-1.243	(0.406)	-0.081	(0.024)	0.069	(0.020)
[17-24] years	-1.296	(0.426)	-0.089	(0.025)	0.068	(0.018)
[25-31] years	-1.339	(0.444)	-0.095	(0.026)	0.067	(0.017)
32+ years	-1.435	(0.468)	-0.104	(0.031)	0.065	(0.019)
Tertiary education						
[0-8] years	-1.705	(0.409)	-0.103	(0.024)	0.067	(0.018)
[9-16] years	-1.505	(0.392)	-0.087	(0.025)	0.072	(0.020)
[17-24] years	-1.558	(0.417)	-0.095	(0.026)	0.071	(0.018)
[25-31] years	-1.604	(0.437)	-0.101	(0.027)	0.070	(0.017)
32+ years	-1.698	(0.462)	-0.110	(0.032)	0.068	(0.020)

Note: The table reports the first stage coefficients of the excluded instruments for the baseline first stage regression. The regression relates the share of immigrants in each education-experience-period-country cell with the interactions of the push factors listed in the top row with distance and with education-experience cell specific coefficients, as described in the text (see equation (6)). The regression includes education, experience, country-period, education-period, experience-period, and education-country fixed effects, which are not reported. The sample includes a balanced panel of active (employed and unemployed) male aged 18-64 for census dates from 1970 to 2000. Destination countries considered in the sample include the U.S., Canada, Austria, France, Greece, Ireland, and Switzerland. The regression is fitted to 420 observations. Standard errors, clustered at the education-experience cell level, are in parentheses. Further details in the main text.

TABLE V.—THE EFFECT OF IMMIGRATION ON NATIVE WAGES OF EDUCATION-EXPERIENCE GROUPS (U.S. AND CANADA)

	Log monthly wages		Log annual wages	
	OLS	IV	OLS	IV
Baseline	-0.556	-1.331	-0.639	-1.505
	(0.151)	(0.450)	(0.333)	(1.113)
Unbalanced panel	-0.557	-1.667	-0.642	-2.146
	(0.152)	(0.555)	(0.336)	(1.293)
Includes female in LF counts	-0.620	-1.405	-0.630	-1.506
	(0.161)	(0.427)	(0.387)	(1.124)
Includes native LF as control	-0.484	-1.130	-0.407	-0.842
	(0.141)	(0.411)	(0.263)	(0.930)
Unweighted regression	-0.399	-0.974	-0.401	-0.815
	(0.215)	(0.732)	(0.351)	(1.330)

NOTE: The table reports the coefficient of the immigrant share from regressions where the dependent variable is the average log wage for native males aged 18-64 in each education-experience-period-country cell (monthly or annual as indicated). See the text for a detailed description of the instrument used in IV estimates. Standard errors are reported in parenthesis and are clustered by education-experience-country cells. All regressions have 120 observations in the second stage, except those labeled as "Unbalanced panel" (135). All regressions are weighted by the sample size used to compute wages in each cell except otherwise indicated. All regressions include education, experience, country-period, education-period, experience-period, and education-country fixed effects.

TABLE VI.—THE EFFECT OF IMMIGRATION ON U.S. NATIVE WAGES

	Log monthly wages		Log annual wages	
	OLS	IV	OLS	IV
Baseline	-0.694	-1.471	-0.910	-1.639
	(0.257)	(0.510)	(0.592)	(1.408)
Unbalanced panel	-0.697	-1.789	-0.916	-2.252
	(0.254)	(0.621)	(0.587)	(1.583)
Includes female in LF counts	-0.762	-1.545	-0.858	-1.650
	(0.260)	(0.501)	(0.721)	(1.450)
Includes native LF as control	-0.550	-1.218	-0.416	-0.874
	(0.262)	(0.482)	(0.560)	(1.174)
Unweighted regression	-0.800	-1.594	-0.853	-1.355
	(0.388)	(0.770)	(0.888)	(1.853)

NOTE: The table reports the coefficient of the immigrant share from regressions where the dependent variable is the average log wage for native males aged 18-64 in the U.S. in each education-experience-period cell (monthly or annual as indicated). See the text for a detailed description of the instrument used in IV estimates. Standard errors are reported in parenthesis and are clustered by education-experience cells. All regressions have 60 observations in the second stage, except those labeled as "Unbalanced panel" (75). All regressions are weighted by the sample size used to compute wages in each cell except otherwise indicated. All regressions include education, experience, country-period, education-period, experience-period, and education-country fixed effects.

Table VII.—Robustness to Different Combinations of Instruments

	OLS	Baseline	Deaths only	Democ./ autocr. only	Democr. only	Autocr. only	Months of war	Dissagr. instr. by continent
Log monthly wages	-0.556	-1.331	-1.746	-1.584	-1.684	-1.949	-1.699	-1.520
	(0.151)	(0.450)	(0.542)	(0.549)	(0.574)	(0.616)	(0.531)	(0.597)
Annual Wages	-0.639	-1.505	-1.803	-1.685	-1.684	-2.104	-1.724	-2.061
	(0.333)	(1.113)	(1.570)	(1.357)	(1.574)	(1.752)	(1.520)	(1.085)
Joint relevance		0.000	0.000	0.000	0.014	0.001	0.000	0.000
Stability subspl.		0.824	0.490	0.612	0.761	0.672	0.629	0.940

Note: The table reports the coefficient of the immigrant share from regressions where the dependent variable is the average log wage for males in each education-experience-period-country cell (monthly or annual as indicated). The two first columns are baseline OLS and IV estimates from Table V. The following columns use alternative combinations of instruments; see the text for further details on the instruments used in each case. The last column uses all instruments, but the first stage is estimated at a more dissagregated level, by continents; hence, in this case, the instrument is predicted at the continent of origin level, and then aggregated to be used in the second stage. Standard errors are reported in parenthesis and are adjusted for clustering within education-experience-country cells. All regressions have 120 observations in the second stage. All regressions are weighted by the sample size of the cell. All regressions include education, experience country-period, education-period, experience-period, and education-country fixed effects.

6. Reexamining the spatial correlations

Table VIII.—The Effect of Immigration on Native Wages at the Regional Level

	Log monthly wages		Log annual wages	
	OLS	IV	OLS	IV
Baseline	-0.262	-0.926	-0.217	-0.619
	(0.085)	(0.287)	(0.117)	(0.435)
Unbalanced panel	-0.274	-1.026	-0.240	-0.944
	(0.089)	(0.353)	(0.116)	(0.499)
U.S. only	-0.248	-1.043	-0.209	-0.740
	(0.102)	(0.297)	(0.141)	(0.463)
Includes female in LF counts	-0.285	-1.142	-0.210	-0.897
	(0.088)	(0.277)	(0.128)	(0.432)
Includes native LF as control	-0.243	-0.798	-0.175	-0.319
	(0.084)	(0.304)	(0.114)	(0.425)
Unweighted regression	-0.299	-0.632	-0.226	0.002
	(0.081)	(0.302)	(0.119)	(0.460)

Note: The table reports the coefficient of the immigrant share from regressions where the dependent variable is the average log wage for native males aged 18-64 in each education-experience-period-region cell (monthly or annual as indicated). The following regions have been considered: Atlantic region, Quebec, Ontario, the Prairies, and British Columbia for Canada, and New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain and Pacific divisions for the U.S. See the text for a detailed description of the instrument used in IV estimates. Standard errors are reported in parenthesis and are clustered by education-experience-region cells. All regressions are estimated with 840 observations in the second stage, except those labeled as "Unbalanced panel" (975), and "U.S. only" (540). All regressions are weighted by the sample size used to compute wages in each cell except otherwise indicated. All regressions include education, experience, country-period, education-period, experience-period, and education-country fixed effects.

Table IX.—Robustness of Results at the Regional Level to Different Combinations of Instruments

	OLS	Baseline	Deaths only	Democ./ autocr. only	Democr. only	Autocr. only	Months of war
Log monthly wages	-0.262	-0.926	-1.369	-1.053	-1.086	-1.363	-1.168
	(0.085)	(0.287)	(0.372)	(0.315)	(0.326)	(0.393)	(0.350)
Annual Wages	-0.217	-0.619	-0.743	-0.747	-0.667	-0.960	-0.603
	(0.117)	(0.435)	(0.690)	(0.495)	(0.547)	(0.705)	(0.608)
Joint relevance		0.000	0.002	0.003	0.007	0.014	0.007
Stability subspl.		0.357	0.653	0.484	0.573	0.888	0.713

Note: The table reports the coefficient of the immigrant share from regressions where the dependent variable is the average log wage for males in each education-experience-period-country cell (monthly or annual as indicated). The two first columns are baseline OLS and IV estimates at the regional level from Table VIII. The following columns use alternative combinations of instruments; see the text for further details on the instruments used in each case. Standard errors are reported in parenthesis and are adjusted for clustering within education-experience-region cells. All regressions have 840 observations in the second stage. All regressions are weighted by the sample size used to compute average wages of the cell. Education, experience, country-period, education-period, experience-period, and education-country fixed effects are included in all regressions.

7. Structural approach

8. Conclusions

References

Altonji, Joseph G. and David E. Card, *The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives*, Chicago: University of Chicago Press,

Angrist, Joshua and Adriana Kugler, "Protective or Counter-Productive? Labour Market Institutions and the Effect of Immigration on EU Natives," *Economic Journal*, June 2003, *113* (448), F302–F331.

Aydemir, Abdurrahman and George J. Borjas, "Cross-Country Variation on the Impact of International Migration: Canada, Mexico, and the United States," *Journal of the European Economic Association*, June 2007, *5* (4), 663–708.

_ **and** _ , "Attenuation Bias in Measuring the Wage Impact of Immigration," *Journal of Labor Economics*, January 2011, *29* (1), 69–112.

Borjas, George J., "Immigrants, Minorities, and Labor Market Competition," *Industrial and Labor Relations Review*, April 1987, *40* (3), 382–392.

_ , "Self-Selection and the Earnings of Immigrants," *American Economic Review*, September 1987, *77* (4), 531–553.

_ , "The Economic Analysis of Immigration," in Orley C. Ashenfelter and David E. Card, eds., *Handbook of Labor Economics*, Vol. 3A, Amsterdam: North-Holland, 1999, chapter 28, pp. 1697–1760.

_ , "The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *Quarterly Journal of Economics*, November 2003, *118* (4), 1335–1374.

_ **and Lawrence F. Katz**, *The Evolution of the Mexican-Born Workforce in the United States*, Cambridge: National Bureau of Economic Research Conference Report,

_ **, Jeffrey T. Grogger, and Gordon H. Hanson**, "Immigration and the Economic Status of African-American Men," *Economica*, April 2010, *77* (306), 255–282.

_ **, Richard B. Freeman, and Lawrence F. Katz**, "How Much Do Immigration and Trade Affect Labor Market Outcomes?," *Brookings Papers on Economic Activity*, Spring 1997, *1997* (1), 1–67.

Card, David E., "The Impact of the Mariel Boatlift on the Miami Labor Market," *Industrial and Labor Relations Review*, January 1990, *43* (2), 245–257.

_ , "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, January 2001, *19* (1), 22–64.

— , "Immigration and Inequality," *American Economic Review: Papers and Proceedings*, May 2009, *99* (2), 1–21.

— **and Ethan G. Lewis**, *The Diffusion of Mexican Immigrants during the 1990s: Explanations and Impacts*, Chicago: University of Chicago Press,

— **and Thomas Lemieux**, "Can Falling Supply Explain The Rising Return To College For Younger Men? A Cohort-Based Analysis," *Quarterly Journal of Economics*, May 2001, *116* (2), 705–745.

Chiquiar, Daniel and Gordon H. Hanson, "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States," *Journal of Political Economy*, April 2005, *113* (2), 239–281.

Cortés, Patricia, "The Effect of Low-Skilled Immigration on U.S. Prices: Evidence from CPI Data," *Journal of Political Economy*, June 2008, *116* (3), 381–422.

— **and José Tessada**, "Low-Skilled Immigration and the Labor Supply of Highly Skilled Women," *American Economic Journal: Applied Economics*, July 2011, *3* (3), 88–123.

Docquier, Frédéric and Abdeslam Marfouk, *International Migration by Educational Attainment, 1990-2000*, New York: Palgrave Macmillan,

Dustmann, Christian, Tommaso Frattini, and Ian Preston, "The Effect of Immigration along the Distribution of Wages," *Review of Economic Studies*, January 2013, *80* (1), 145–173.

Friedberg, Rachel M., "The Impact of Mass Migration on the Israeli Labor Market," *Quarterly Journal of Economics*, November 2001, *116* (4), 1373–1408.

— **and Jennifer Hunt**, "The Impact of Immigrants on Host Country Wages, Employment and Growth," *Journal of Economic Perspectives*, Spring 1995, *9* (2), 23–44.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Hårvard Strand, "Armed Conflict 1946-2001: A New Dataset," *Journal of Peace Research*, September 2002, *39* (5), 615–637.

Glitz, Albrecht, "The Labor Market Impact of Immigration: A Quasi-Experiment Exploiting Immigrant Location Rules in Germany," *Journal of Labor Economics*, January 2012, *30* (1), 175–213.

Goldin, Claudia, *The Political Economy of Immigration Restriction in the United States, 1890 to 1921*, Chicago: University of Chicago Press,

Grossman, Jean B., "The Substitutability of Natives and Immigrants in Production," *Review of Economics and Statistics*, November 1982, *64* (4), 596–603.

Hunt, Jennifer, "The Impact of the 1962 Repatriates from Algeria on the French Labor Market," *Industrial and Labor Relations Review*, April 1992, *45* (3), 556–572.

LaLonde, Robert J. and Robert H. Topel, *Labor Market Adjustments to Increased Immigration*, Chicago: University of Chicago Press,

Llull, Joan, "Immigration, Wages, and Education: A Labor Market Equilibrium Structural Model," mimeo, CEMFI, November 2010.

_ , "Reconciling Spatial Correlations and Factor Proportions: A Cross-Country Analysis of the Economic Consequences of Immigration," mimeo, CEMFI, June 2011.

_ , "Understanding International Migration: Evidence from a New Dataset of Bilateral Stocks (1960-2000)," mimeo, Universitat Autònoma de Barcelona, January 2013.

Manacorda, Marco, Alan Manning, and Jonathan Wadsworth, "The Impact of Immigration on the Structure of Wages: Theory and Evidence from Britain," *Journal of the European Economic Association*, February 2012, *10* (1), 120–151.

Marshall, Monty G., Keith Jaggers, and Ted Robert Gurr, "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2010," Version 2010. College Park, MD: Center for International Development and Conflict Management, University of Maryland. 2010.

Minnesota Population Center, "Integrated Public Use Microdata Series, International: Version 6.1 [Machine-readable database]," Minneapolis: University of Minnesota 2011.

Ottaviano, Gianmarco I. P. and Giovanni Peri, "Rethinking the Effect of Immigration on Wages," *Journal of the European Economic Association*, February 2012, *10* (1), 152–197.

Saiz, Albert, "Immigration and Housing Rents in American Cities," *Journal of Urban Economics*, March 2007, *61* (2), 345–371.

Vincenty, Thaddeus, "Direct and Inverse Solutions of Geodesics on the Ellipsoid with Application of Nested Equations," *Survey Review*, April 1975, *23* (176), 88Ũ93.

Appendix A: Variable definitions and sources