Understanding International Migration: Evidence from a New Dataset of Bilateral Stocks (1960-2000)

Joan Llull

September 2013

Barcelona GSE Working Paper Series

Working Paper nº 715

Understanding International Migration: Evidence from a New Dataset of Bilateral Stocks (1960-2000)

By Joan Llull[*§]
MOVE, Universitat Autònoma de Barcelona, and Barcelona GSE

This version: September 2013

In this paper I present a new database of bilateral migrant stocks, and I provide new evidence on the determinants of international migration. The new Census-based data are obtained from the National Statistical Offices of 24 OECD countries, and they cover the total stock of immigrants in each destination country for 1960-2000, including 188 countries of origin. Empirically, I find strong evidence of heterogeneous effects of income gains on migration prospects depending on distance. For example, a 1,000$ increase in U.S. income per capita increases the stock of Mexican immigrants in the country by a percentage three times larger than the percentage increase in the stock of Chinese.

Keywords: *International Migration, Determinants, Data collection*

JEL Codes: *F22, J61, O15*

I. Introduction

International migration has increased dramatically in recent decades. Understanding the determinants of the movement of workers across international borders is crucial for immigration policy design. This paper aims to enhance our knowledge about these determinants by presenting new data on bilateral migrant stocks, and new empirical evidence on the determinants of international migration.

To create the new database on international migrant stocks presented in this paper, I collected data from National Statistical Offices of the 24 richest OECD countries. This dataset includes bilateral stocks of immigrants from 188 countries of origin into these 24 destination countries for the period 1960 to 2000. The database has four main advantages. First, it covers a larger time period than

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

[§] I am indebted to Manuel Arellano for his constant encouragement and advice. I wish to thank Stéphane Bonhomme, Juan José Dolado, Manu García-Santana, Jesús Fernández-Huertas, Nezih Guner, Claudio Michelacci, Pedro Mira, Francesc Ortega, Roberto Ramos, Pedro Rey-Biel, Rob Sauer, Jim Walker, and seminar participants at CEMFI for helpful comments and discussions. Financial suport from European Research Council (ERC) through Starting Grant n. 263600, and the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centers of Excellence in R&D (SEV-2011-0075) is gratefully acknowledged. I very much appreciate the help from several statistical offices from many OECD countries in providing me with the data.

previous databases used in the literature. Second, it is based on destination country Censuses, as opposed to issues of residence and work permits, reducing, as a result, the undercounting of undocumented immigrants. Third, it collects information on stocks by country of origin, as opposed to flows, which is desirable both because equilibrium values are often expressed in terms of stocks (e.g. Grogger and Hanson, 2011), and because migration flow data are less reliable due to the impossibility of keeping track of out migration and return migration flows (Docquier and Marfouk, 2006). And, fourth, unlike previous databases, it fully covers the stock of immigrants in each of the considered destination countries.[1]

Empirically, I test for the existence of heterogeneous effects of income gains on migration prospects depending on distance. According to a static model — the approach which mostly followed by the literature— when individuals decide whether to migrate to another country, they base their decision on *net* income gains from migration, i.e. the differential in wages between the two countries net of (one time) moving costs.[2] From a dynamic point of view, however, individuals may care about moving costs (distance in particular) even after having migrated. Large moving costs will reduce their flexibility to move back and forth to their home country as a consequence of income shocks;[3] and, if individuals dislike living far away from home, they may require a compensating wage differential for living abroad that might be increasing in distance. Forward looking individuals will take these two factors into account when deciding whether to migrate in the first place. As a result, the effect of income gains on moving prospects may be heterogeneous depending on distance: individuals from farther countries would be less reactive to income fluctuations compared to individuals from closer countries. Results suggest that these heterogeneities are indeed very important. For example, a 1,000$ increase in U.S. income per capita increases the stock of Mexican immigrants in the U.S. by a percentage that is three times larger than the percentage increase in the stock of Chinese immigrants. This result has important consequences for

[1] As widely discussed below, some of these data is presented in grouped categories. Some existing databases in the literature use imputation methods to distribute these migrants across specific countries. I opted to present the data keeping track of grouped categories, letting the user to decide whether to use the data as is (as I do in the empirical exercises presented below), or to use the imputation method that better adapts to her research interests.

[2] Examples of papers using this approach include Borjas (1987), Borjas and Bratsberg (1996), Karemera, Oguledo and Davis (2000), Chiquiar and Hanson (2005), Clark, Hatton and Williamson (2007), Pedersen, Pytlikova and Smith (2008), Mayda (2010), and Grogger and Hanson (2011) among many others. Recent papers like Bertoli and Fernández-Huertas Moraga (2013), Bertoli, Fernández-Huertas Moraga and Ortega (2013), or Ortega and Peri (2013) estimate nested logit models that allow for different elasticities across destinations.

[3] Kennan and Walker (2011) and Lessem (2013) argue that migration is a dynamic decision, and that repeated and return migration are important in the data.

immigration policy design. For example, a pull-driven immigration shock (i.e. due to a positive income shock) will imply significant changes in the composition of immigrant population in terms of nationalities. Similarly, a negative shock to a developing country will have a much larger effect for neighboring countries than what it was previously estimated; this larger effect implies that destination countries may want to favor neighboring countries when planning development assistance policies if they are interested in reducing immigrant inflows.

Collecting data on bilateral migration is, in general, a difficult task. Reliability of statistics from origin countries is low because it is difficult to keep track of the people who leave the country. Data from destination countries is more accurate. The lack of comparable cross-destination country bilateral data led many papers in the literature to follow a single destination country over time (e.g. Borjas and Bratsberg, 1996; Karemera, Oguledo and Davis, 2000; Clark, Hatton and Williamson, 2007; Bertoli and Fernández-Huertas Moraga, 2013). More recently, researchers and institutions have put some effort in gathering comparable bilateral migration data across destination countries. Pedersen, Pytlikova and Smith (2008) and Mayda (2010) are the first papers using cross-destination country panel data on bilateral inflows to analyze the effect of income gains and moving costs on migration flows. Mayda (2010) uses a database from OECD on annual legal inflows of workers by country of origin; she uses these data to investigate the determinants of migration inflows into 14 OECD countries between 1980 and 1995. Pedersen, Pytlikova and Smith (2008) produce a similar database collecting data on issues of residence and work permits from National Statistical Offices from 1990 to 2000. They use these data to look at the effects of networks and welfare benefits on international migration. These two databases have recently been recently expanded by Ortega and Peri (2013) and Adsera and Pytlikova (2012) respectively.[4] The four databases contain information on inflows of immigrants and, with a lower accuracy, net flows. They are based on the number of issues of residence and work permits, which is likely to produce a severe underestimation the real numbers due to illegal migration. And, as acknowledged by the authors, the data also have an important amount of missing data and incorrect zero values (for countries with relatively small flows), thus covering, as a result, a limited fraction of total inflows (Mayda, 2010, pp.1258-59).

[4] The database in Ortega and Peri (2013) includes information for 15 destination countries and 120 countries of origin for the period 1980-2006. The data presented in Adsera and Pytlikova (2012) cover the period 1980 to 2009 for 30 destination countries and many countries of origin (with some missing data).

In a similar spirit to the database I present in this paper, Docquier and Marfouk (2006) and Docquier, Lowell and Marfouk (2009) collect Census-based data. The aim of their databases is to gather information on stocks of immigrants by educational level, and, for this reason, they only cover two census dates, 1990 and 2000. Two papers use these data to analyze the determinants of migration flows. Grogger and Hanson (2011) use them to the analyze the determinants of scale and composition of migration flows. Ortega and Peri (2009) combine those two years of data on stocks with the OECD database on annual legal inflows used in Mayda (2010) to extrapolate stocks back to 1980, and devote a portion of their paper to analyze the determinants of migration flows.[5]

The rest of the paper is organized as follows. Section II describes in detail the data collection process and the contents of the new database presented in this paper. Section III presents estimates for linear effects of income gains on moving prospects, in a similar specification to the one used in the literature. Section IV shows new evidence on the existence of non-linearities in the effect of income gains on moving prospects depending on distance. And Section V concludes.

II. Data

A. A new database on bilateral migrant stocks (1960-2000)

Observing international migration is not easy. In general, origin countries do not collect statistics on the number of people who leave the country, so the main source of data is at the destination level. The fact that different countries count immigrants in different ways requires additional effort from the researcher to work on the comparability of the different statistics. For that reason, several papers such as Borjas and Bratsberg (1996), Karemera, Oguledo and Davis (2000), and Clark, Hatton and Williamson (2007) among many others focus on explaining the determinants of immigrant inflows into a single destination country (United States).

Recently, OECD collected data on annual flows based on the Continuous Reporting System on Migration.[6] These data has several issues that harm their reliability. First, the data are based on the number of issues of residence and work permits. Therefore, they only cover legal migration flows. Second, they include an important amount of misreported zeros and missing values —and missing countries as well. As a result, for a given destination country, the sum of inflows over all countries of origin is not equal to the 100% of the total inflow. For exam-

[5] These data is no longer used in the published version of the paper (Ortega and Peri, 2013).

[6] Contemporaneously, Pedersen, Pytlikova and Smith (2008) built a similar database of official flows from 1990 to 2000 collecting information from different National Statistics Offices.

ple, Mayda (2010) states that in her sample "the percentage of the total immigrant inflow covered by the disaggregate data ranges between 45% (Belgium) to 84% (United States)". A reason for this could be that OECD does not keep track of countries with relatively small inflows that are reported by national offices into residual grouped categories. The elimination of data from countries that were lately dissolved (USSR, Yugoslavia, Czechoslovakia, Rhodesia,...) may also play an important role.[7] Third, despite its annual frequency, that database covers a relatively small time span, as it starts in 1995 (an earlier version of the dataset, used in Mayda (2010), goes back to 1980 but it only covers 14 OECD destination countries). And fourth, these are data on inflows (and, in some cases, less reliable data on outflows), which makes hard to construct reliable measures of immigrant stocks by origin country.

In this paper, I collected Census-based data from National Statistical Offices of 24 OECD countries.[8] The data contains stocks of immigrants by country of origin from 1960 to 2000. This dataset have important advantages. First, it covers 100% of stocks of immigrants in all of these destination countries. Second, for economic and statistical reasons it is more attractive to work with stocks rather than flows: from an economic point of view, equilibrium values are often expressed in terms of stocks; statistically, it has long been recognized that migration flow data are less reliable than stock data, because of the impossibility of evaluating emigration and return migration movements (Docquier and Marfouk, 2006). Third, although Censuses do not record all undocumented immigrants, they do a much better job in counting them than registers of residence and work permits issues (especially when census interviews are carried personally at dwellings). And fourth, the dataset covers a wider time period than existing databases (from 1960 to 2000).

Data is based on destination countries' Censuses.[9] From each Census, I collect data on the stock of immigrants by country of birth or country of nationality. The dataset contains information on stocks of immigrants from 188 countries of origin —sometimes in a grouped category— into each of the 24 richest OECD destination countries.[10] Although some destination countries carry a Census every five years,

[7] Other databases, like Docquier and Marfouk (2006) cover stocks completely —for a much shorter time period—, but make imputations for countries in which data are in a grouped category. This imputation may generate a correlation structure between the error term and some regressors that may produce biases. Hence, keeping track of aggregate observations instead of making imputations seems a safer option as discussed below.

[8] These countries include Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Korea (Rep.), Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom and United States.

[9] Nordic countries replaced Censuses for continuous population registers during 1980s.

[10] Source countries include all Member States of United Nations except Andorra, Liechten-

most of them do it every 10 years, so data is presented at a 10-year frequency. Hence, the database is well suited for looking at long-run effects (annual flow data from OECD would still be useful to analyze short-run adjustments of legal inflows).

Although they are unlikely to systematically affect the analysis below, there are some comparability issues that worth mentioning. First, the definition of immigrant is different across countries.[11] Some countries define immigrants on the basis of the place of birth whereas others do it based on nationality. This might affect the comparability of stocks across destination countries, but changes over time are reasonably comparable —so destination country fixed effects should account for these differences, given the log-specification of stocks in the specification estimated below. Second, census dates vary across destination countries —roughly a half of them are carried in even years (1960, 1970,...) and the other half in odd years (1961, 1971,...).[12] Dates are generally consistent, so the difference between two census dates is usually of ten years.

B. Description of the data

Although the database covers a 100% of immigrant stocks in the selected destination countries, sometimes data for some origin countries are grouped in some aggregate categories. One of the aims of this section is to show to what extent this grouping may be important, in order to let the user to decide which exercises the data are suitable for. I also discuss why they are well suited for the empirical analysis in this paper.

Data may be grouped for several reasons. One of them is that Statistical Offices decide to group several countries into one or some *residual categories* (usually labeled as "Other countries in region X"). In some other cases, they report the stock of immigrants born in a former country that later on split into several countries: USSR, Czechoslovakia, Yugoslavia, Ethiopia/Eritrea, Rhodesia, and

stein, Monaco, Myanmar, Marshall Islands, Nauru, San Marino, Timor-Leste, and Tuvalu (none of them are available in Penn World Tables). Additionally, it includes the dependent territories of Taiwan, Macao, Hong Kong, Bermuda, Netherlands Antilles, and Puerto Rico. Serbia and Montenegro are considered as a sole country.

[11] This issue is also present in existing databases in the literature.

[12] The only exception is France, whose Censuses were carried in 1954, 1962, 1968, 1975, 1982, 1990, 1999 and 2006. I interpolated them linearly to fit census dates to 1961, 1971, 1981, 1991, and 2001. There are three additional countries for which I have to extrapolate some values. Denmark and Finland did not carry a census circa 1960 and 1970, so I extrapolate using information on residence permits for Denmark and on main language used for Finland. For Germany, I do not have pre-unification censuses, so I do an extrapolation based on data on legal flows into West Germany. I do robustness analysis to the exclusion of 1960 and 1970 from the analysis, and results are not affected. Finally, data for United Kingdom includes only immigrants living in England and Wales; however, for year 2000 they represent a 95% of the total stock of immigrants in the UK, a percentage that was uniformly distributed across origin countries.

Table 1—Number of origin countries with grouped data across destinations

	1960		1970		1980		1990		2000	
	Num. of countries	% of stock	Num. of countries	% of stock	Num. of countries	% of stock	Num. of countries	% of stock	Num. of countries	% of stock
Australia	143	7.78	134	9.97	116	8.35	46	6.15	11	0.00
Austria	187	100.00	136	49.03	125	45.92	125	42.25	110	3.54
Belgium	155	4.04	153	2.80	143	2.01	143	3.03	118	1.63
Canada	172	20.82	172	25.43	175	36.26	24	5.59	22	5.97
Denmark	170	57.24	169	26.39	24	7.15	24	6.63	22	13.05
Finland	187	100.00	187	100.00	19	15.17	24	18.12	22	46.86
France	178	15.59	177	10.67	177	14.46	156	7.87	156	11.17
Germany	187	100.00	187	100.00	187	100.00	171	65.15	109	11.61
Greece	178	17.71	178	25.67	121	2.48	121	11.34	65	0.09
Iceland	187	100.00	187	100.00	133	2.41	150	6.67	131	11.48
Ireland	181	8.60	179	8.00	177	8.29	175	10.18	150	6.61
Italy	152	11.30	187	100.00	160	23.28	137	15.34	105	1.97
Japan	184	1.84	184	3.48	184	4.54	133	2.45	151	1.10
Korea (Rep.)	179	0.48	166	0.73	166	3.60	166	7.14	169	10.41
Luxembourg	178	2.31	177	3.93	175	5.40	175	8.58	175	14.98
Netherlands	180	46.73	180	33.43	180	28.83	180	25.81	22	5.20
New Zealand	186	89.60	186	89.47	186	89.34	139	3.23	132	3.09
Norway	169	7.20	163	11.19	161	14.11	161	22.04	155	18.08
Portugal	179	11.83	179	16.80	24	0.15	24	0.32	152	10.33
Spain	187	100.00	133	1.21	124	1.59	24	0.30	0	0.00
Sweden	148	4.03	164	10.82	24	10.11	24	11.20	5	11.34
Switzerland	74	0.42	47	2.63	24	6.60	24	13.42	22	24.78
United Kingdom	132	9.08	148	11.59	172	47.00	157	36.64	0	0.00
United States	126	13.94	122	10.90	24	5.20	24	3.08	75	4.52
Average	167	34.61	162	31.42	125	20.09	105	13.86	87	9.08
Excluding 100%'s	*161*	*17.40*	*161*	*21.26*	*122*	*16.60*	*105*	*13.86*	*87*	*9.08*

Note: The first column for each year represents the number of countries that are in grouped categories in that period. The total amount of possible origin countries is 187. The second column for each year is the % from the total stock of immigrants that is in grouped categories. Each destination country may have several grouped categories. The last two rows are averages across destination countries.

the West Indies Federation are good examples. Finally, in some cases all origin countries are grouped, either because I only observe the total stock of immigrants in the destination country (a single group), or because the data is presented in big aggregate categories (e.g. data by continent of origin).

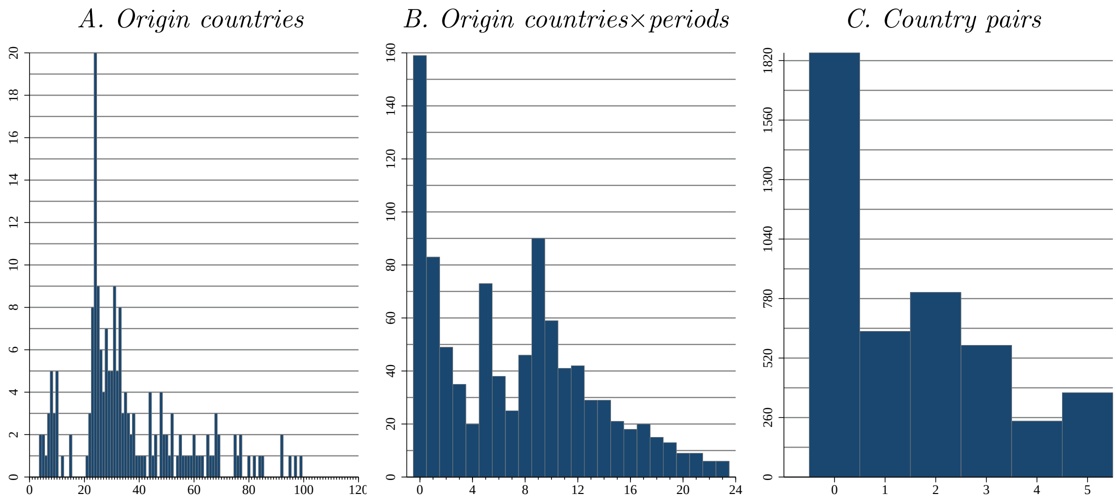
Table 1 summarizes the importance of grouped data. There are several aspects to highlight from the Table. First, data are more disaggregated in recent years: the average number of countries in grouped categories decrease from 167 to 87, and the share of the total stock that they represent decreases from more than one third in 1960 to less than 10% in year 2000. Second, even though in 1960 and 1970 the coverage of total migrant stocks by bilateral data is only of around two thirds of the stock, this coverage increases to 80% if we exclude the destination countries for which we only observe the total migrant stock. And third, even considering only disaggregated bilateral observations, the coverage of the total stock of immigrants if much larger than in the OECD database. Mayda (2010) states that the coverage of total inflows in her database ranges from from 45% (Belgium) to 84% (US); for the time period covered by her database, the average coverage by bilateral observations ranges from 80% to 91%. Regarding the number of countries with disaggregate bilateral observations, Mayda (2010) and Ortega and Peri (2013) use a sample of 79 and 120 origin countries respectively —including zero flows that "are likely to correspond to very small flows rather than zero flows" (Mayda, 2010); Pedersen, Pytlikova and Smith (2008), report a substantial portion of missing values among their sample of 129 countries of origin. The country coverage for these years is similar on average in Table 1, but it increases both if we restrict to the sample of 15 destination countries considered in Mayda (2010) and Ortega and Peri (2013), or if we consider federations of countries that were single countries at that time (e.g. USSR, Yugoslavia,...) as ungrouped countries. For instance, Yugoslavia accounted for almost a half of the stock of immigrants in Austria in years between 1970 and 1990, one quarter of the stock in Switzerland in year 2000, and around a 10% of the Swedish stock in years between 1980 and 2000, and the USSR represented between 5 and 8% of US and Canadian stocks in years 1960 and 1970, and around a 3% in other several destination countries.

A potential limitation of working with grouped data is in the identification of fixed effects in the estimation. In the baseline regression estimated below, I introduce origin and destination country fixed effects, and year dummies. Additionally, in several specifications of the model I introduce either country pair fixed effects, or country of origin×year dummies. Destination country and time fixed effects are identified in all cases, as grouping only affects origin countries. To identify a dummy for an origin country, we need to observe, at least, one bilateral observation from that country, or that the country appears in a unique combination of grouped observations.[13] To identify a country of origin×year dummy, this bilateral observation or unique combination of groups has to be observed in each year. And the identification of a country pair dummy requires the bilateral observation to be observed for each destination country. Otherwise, a single dummy for each combination of groups would be identified.

Figure 1 summarizes the availability of this variation. The left histogram shows the number of origin countries with 0,1,...,120 (=24×5) country of destination×year observations. All countries of origin have between 4 and 99 destination×year observations, which is enough to identify all origin country fixed effects; in most of the cases (105 out of 188 countries, 55% of them) we have between 20 and 40 observations. The central histogram shows the number of countries of origin×years with

[13] For example, consider that for a destination country A we have two observations that belong to the "Rest of Europe" group. If in another destination country B one of them belongs to the "Yugoslavia" category and the other does not, I would be able to identify them separately.

Figure 1. Available ungrouped bilateral observations



Note: Left histogram presents the number of origin countries with 0,1,...,120 destination×year observations; the total amount of origin countries is 188. Center histogram shows the number of origin×years with 0,1,...,24 destination country observations; there are $187 \times 5 = 935$ origin×year observations. Bottom histogram presents the number of country pairs with 0,1,...,5 yearly observations; the total amount of country pairs is $24 \times 187 = 4,488$.

bilateral data for the 0,1,...,24 destination countries. The figure shows that we cannot separately identify country of origin×year dummies in 160 out of $188 \times 5 = 935$ origin×year combinations (17%), in most of the cases due to former federations of countries —USSR, Yugoslavia,...— that were still federated. Finally, the right histogram shows the number of country pairs with bilateral data for the 0,1,...,5 periods. According to the figure, we cannot identify a country pair dummy for 1,854 out of $24 \times 188 = 4,488$ country pairs (41%).[14] This limitation does not affect consistency of the estimates below, but precision is harmed substantially.

Table 2 presents averages, standard deviations, and extreme values for each destination country, and the number of available observations. The left panel refers to the baseline sample, which includes all disaggregated bilateral observations plus one observation for each set of grouped countries.[15] The right panel restricts the sample to disaggregated bilateral observations. The baseline sample includes 6,804 bilateral observations plus 625 grouped observations for different aggregations of countries of origin. These observations are not uniformly distributed across destination countries, ranging from the 55 single bilateral observations for Luxembourg (plus 26 grouped observations) to the 744 for Switzerland (plus 28 groups).

[14] Additionally, for 83 countries of origin×years (9%) and for 637 country pairs (14%) we only have one observation. In this case, the available observation together with the grouped data for other destination countries is enough to identify the fixed effect, but such observations do not contribute to the identification of other parameters.

[15] In the computation of the different statistics from the table and in regressions below, grouped observations are weighted by the number of countries included in the group.

Table 2—Descriptive statistics for migrant stocks

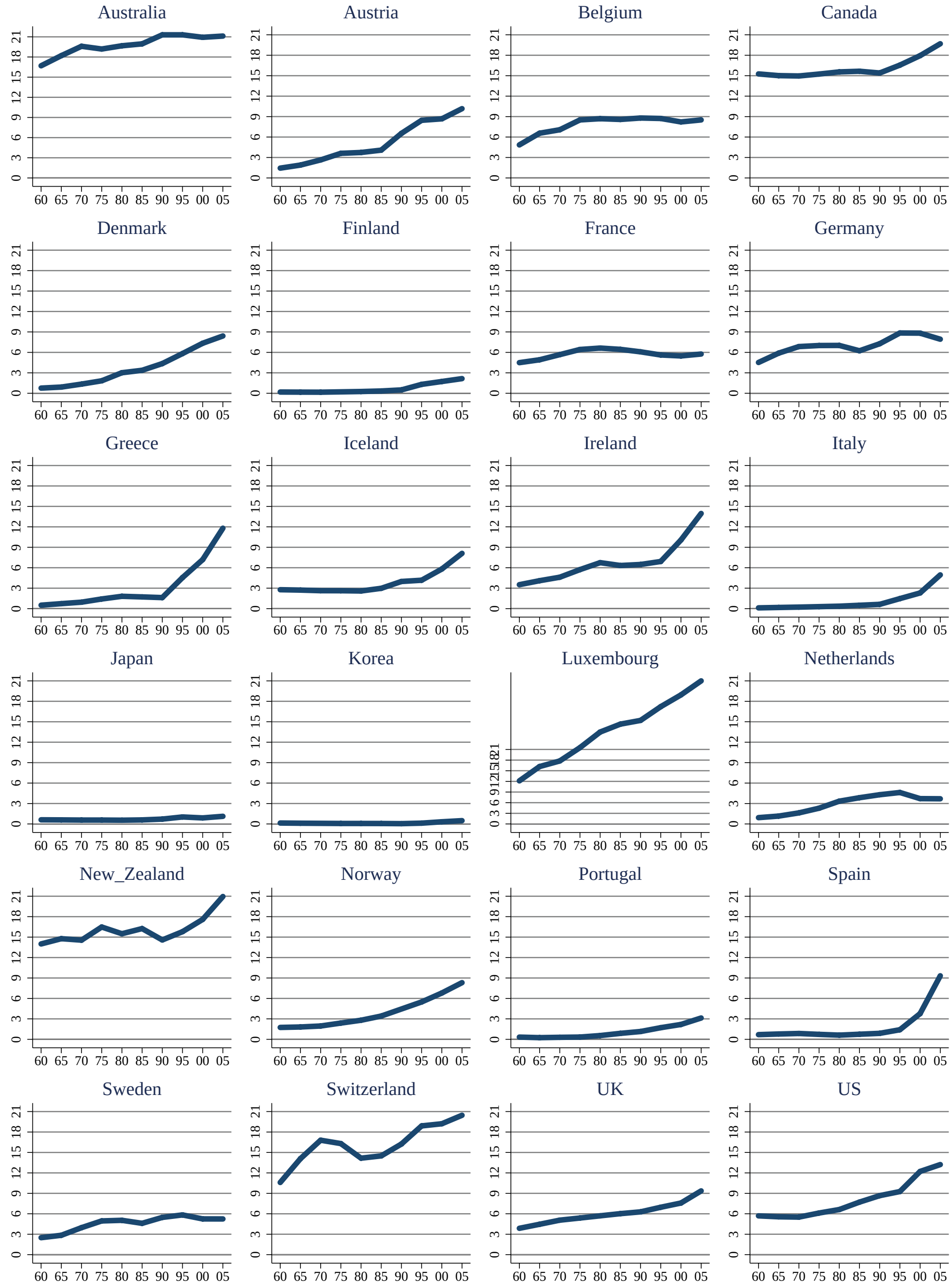
	Full sample (weighted)					Ungrouped observations only				
	Obs.	Mean	Sd.	Min.	Max.	Obs.	Mean	Sd.	Min.	Max.
Australia	531	15,982	78,882	1	1,104,594	486	28,975	107,671	1	1,104,594
Austria	280	1,924	9,938	1	132,975	253	4,859	17,424	1	132,975
Belgium	261	3,965	21,011	0	279,700	223	16,201	40,676	17	279,700
Canada	405	21,362	75,118	1	969,715	370	44,759	115,159	1	969,715
Denmark	555	935	3,322	0	50,470	526	1,444	4,261	1	50,470
Finland	367	152	527	1	7,887	354	218	690	1	7,887
France	126	16,787	77,617	107	791,627	91	152,310	203,663	5,728	791,627
Germany	108	29,159	78,535	288	1,947,938	95	95,841	227,470	366	1,947,938
Greece	301	1,315	14,653	1	438,036	275	4,279	26,790	8	438,036
Iceland	174	46	182	0	2,456	147	204	424	1	2,456
Ireland	82	1,163	12,441	20	242,155	73	13,694	42,575	20	242,155
Italy	225	2,214	9,956	3	180,103	194	9,353	20,320	15	180,103
Japan	109	4,110	40,867	26	567,598	99	37,854	120,424	102	567,598
Korea (Rep.)	106	284	2,272	0	47,474	89	2,775	6,882	1	47,474
Luxembourg	81	500	3,198	0	58,657	55	7,737	10,857	115	58,657
Netherlands	208	2,167	11,284	1	191,500	193	8,146	23,896	1	191,500
New Zealand	138	2,601	15,140	31	232,764	106	12,021	32,276	39	232,764
Norway	162	801	2,691	8	33,251	126	4,944	5,801	26	33,251
Portugal	401	482	2,619	1	37,014	377	1,111	4,038	1	37,014
Spain	487	2,811	13,692	0	244,630	467	5,157	19,080	1	244,630
Sweden	592	1,993	10,369	0	181,477	570	2,936	13,110	1	181,477
Switzerland	772	5,411	33,453	0	583,855	744	6,008	37,014	1	583,855
United Kingdom	361	17,338	56,261	3	675,870	327	39,377	90,436	3	675,870
United States	597	99,276	395,483	11	9,325,452	564	154,516	501,317	11	9,325,452
All	7,429	9,699	90,729	0	9,325,452	6,804	26,483	162,994	1	9,325,452

Note: The unit of observation is origin-destination-year. All figures (except the number of observations) are in individual counts. Left panel refers to the baseline sample, which includes disaggregate bilateral observations and grouped observations —grouped observations are weighted by the number of countries included in the group. Right panel restricts the sample to disaggregated bilateral observations.

The comparison of averages across the two samples suggests that grouped observations tend to include countries with smaller stocks of migrants, which is not surprising given that some grouping occurs due to labeling like "Other countries in region X". The difference in average stock size between the two samples, however, may be exaggerated by the fact that data are more grouped in earlier years of the sample, when immigrant stocks are smaller. The fact that grouping does not occur at random highlights the importance of including grouped observations in the analysis (as opposed to dropping them from the sample).

Table 2 shows substantial variation in average migrant stocks, ranging from 46 immigrants per origin country in Iceland to 99,276 individuals per country in the United States. There is also a large variation across origin countries, as appreciated from the size of standard deviations. The extreme case is the U.S., with a standard deviation of 395,483 individuals, and stocks of immigrants that range from the 11 immigrants from Djibouti in 1990 to the 9,325,452 Mexicans in year 2000, but it is not the only one: Canada, Germany, France and Japan have also sizeable standard deviations, and they are also quite large in Greece and Ireland compared to averages. Overall, the standard deviation in the whole

Figure 2. Share of immigrants (%) for sample OECD countries (1960-2000)



Note: Each plot presents destination country's share of immigrants (immigrants over population). Left axes have a common scale, ranging from 0 to 21% —which is compressed for Luxembourg due to its exceptionally large fraction of immigrants (36.4% in year 2000).

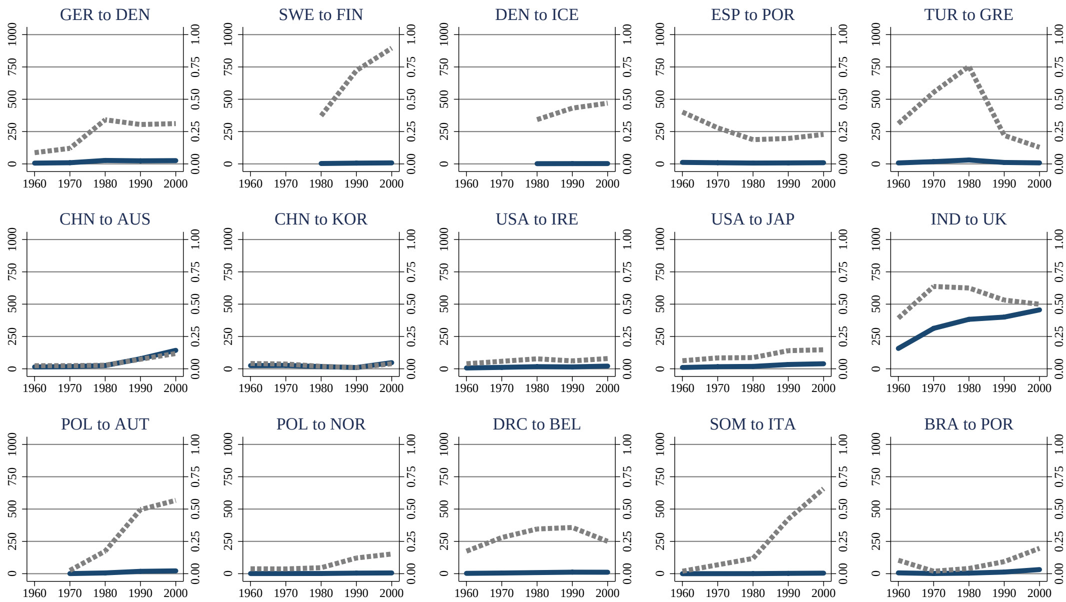
sample is 90,729 individuals, roughly ten times the sample average.[16]

Table 2 does not provide a sense of time series variation. Figure 2 draws the evolution of immigrant shares (i.e. stock of immigrants over population) across

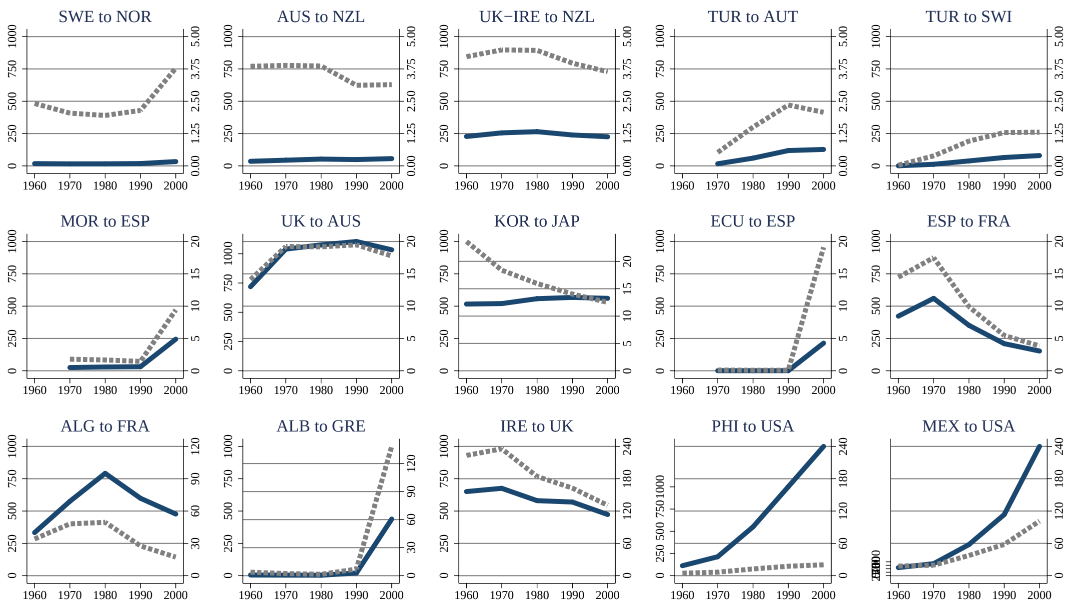
[16] These sample standard deviations are downward biased unless the stock of immigrants from all countries in each grouped observation is the same; the underestimation of the *true* standard deviations will be larger the larger the (unobserved) dispersion within each grouped observation.

Figure 3. Stocks (1000s) and share of population who migrated (‰) for selected country pairs (1960-2000)

A. Some country pairs with low migrant rates



B. Some country pairs with high migrant rates



Note: Solid lines are bilateral migrant stocks (in 1,000s, left axis) from the origin to the destination countries listed in each title. Dashed lines are migrant rates (in ‰, right axis), i.e. stock of migrants from country "X" in country "Y" over total population of country "X" (origin). Left axis scale is common to all country pairs ranging from 0 to 1,000 thousands —which is compressed for MEX and PHI to USA (9.3 and 4.4 million respectively in year 2000). Right axes from top panel have also common scale (0 to 1‰); in the bottom panel, it ranges from 0 to 5‰ in the first rows, from 0 to 20‰ in the second one, and from 0 to 120‰ and 0 to 240‰ in the last row.

destination countries over the sample period. Different patterns are observed across countries: stable low-immigration countries (Korea and Japan), stable high-immigration countries (Australia, Canada and New Zealand), old immigration countries with a strong increasing trend (U.S., Luxembourg, Switzerland, and the UK), old immigration countries with a slight decrease (Belgium and France), and new immigration countries (Spain, Italy, Austria, Greece, Portugal, and Nordic Countries). Figure 3 adds the country of origin layer. In particular, I plot the evolution of the stock of immigrants and of the bilateral migration rate for a selected group of country pairs.[17] The figure shows substantial variation across countries and over time. The top panel includes a sample of country pairs with low migration rates, which include some pairs with a constant trend (e.g. North Americans in Korea or Japan, Chinese in Korea), and others with important increases over the sample period (Somalis in Italy, Polish in Austria, Swedes in Finland). The bottom panel includes country pairs with high migrant rates, including pairs with decreasing rates (Koreans in Japan, Irish in the UK, Spanish in France), roughly constant rates (Australian and British/Irish in New Zealand, British in Australia), and sharply increasing rates (Ecuadorians in Spain, Albanians in Greece, Yugoslavian in Denmark, and, most extremely, Filipino and Mexicans in the U.S.).

C. Other variables

The remaining variables used in the regression analysis below come from different sources (descriptive statistics provided in Table 3). All variables are averages over years $t-10$ to $t-1$. GDP per capita, population, and government share of GDP come from Penn World Tables (versions 6.2 and 7.0). In order to minimize the number of missing values for GDP per capita, I use Total Economy Database (Conference Board) to extrapolate backwards discontinuous Penn World Tables series. Both origin and destination countries' series are in constant international dollars of 2005 (chain). Population in origin and destination countries are in logarithms. Government share is public sector consumption over real GDP. Age dependence ratio at destination country —individuals older than 65 years over population of working age— is from World Develoment Indicators. Unemployment rate (in %) is obtained from the OECD. Geographic variables include physical distance —great circle distance between the two capitals— and dummies for having a common language, a past colonial relationship and a common border. The distance variable is based on Rose (2004) data, extended to cover the whole

[17] Bilateral migration rate is defined as the stock of migrants for a country pair over the population of the origin country.

Table 3—Descriptive statistics of the explanatory variables

	Full sample (weighted)					Ungrouped observations only				
	Obs.	Mean	Sd.	Min.	Max.	Obs.	Mean	Sd.	Min.	Max.
GDPpc dest. (1000s)	7,429	17.85	8.33	1.59	49.94	6,804	21.99	6.87	1.59	49.94
GDPpc origin (1000s)	7,340	8.11	8.06	0.26	215.02	6,727	9.68	12.24	0.26	215.02
Log Distance	7,429	8.31	0.72	4.80	9.79	6,804	8.11	0.95	4.80	9.79
Common language	7,429	0.10	0.24	0.00	1.00	6,804	0.15	0.36	0.00	1.00
Colonial relationship	7,429	0.05	0.17	0.00	1.00	6,804	0.07	0.26	0.00	1.00
Common border	7,429	0.01	0.10	0.00	1.00	6,804	0.03	0.18	0.00	1.00
Log Pop. origin	7,429	3.52	1.40	-2.06	9.40	6,804	4.26	2.04	-2.06	9.40
Log Pop. dest.	7,429	4.73	1.62	0.45	7.88	6,804	4.96	1.35	0.77	7.88
Unemployment rate dest. (%)	7,172	4.39	2.43	0.03	11.10	6,608	5.29	2.34	0.03	11.10
Age dep. dest.(% old/w-age)	7,429	18.45	4.75	2.32	27.66	6,804	20.38	3.86	2.32	27.66
Government share dest.	7,427	9.62	2.81	2.83	19.80	6,804	9.31	2.67	2.83	19.80
War origin	7,429	0.06	0.13	0.00	1.00	6,804	0.08	0.22	0.00	1.00
Polity IV origin	7,232	0.60	5.31	-10.00	10.00	6,607	2.18	7.32	-10.00	10.00

Note: The unit of observation is origin-destination-year. Left panel includes both observations with bilateral migrant data, and observations for which migrant stocks are grouped —which are grouped equivalently, weighting by the number of countries in the group. Right panel includes only observations with available bilateral stocks.

sample. The common language dummy is constructed using data from Alesina et al. (2003) and The World Factbook from the CIA; a pair of countries is considered to have a common language if there is a particular language that is spoken by at least a 10% of the population in each of the two countries. Colonial relationship and common border dummies are also based The World Factbook. War and Polity IV autocracy-democracy index are constructed with data from the Polity IV Project. The war variable measures the fraction of months over the preceding decade that the country was in any type of war. The Polity IV index ranges from -10, indicating autocracy, to 10, which indicates democracy, through values around 0, which indicate anocracy (a situation of instability emerged from the absence of a strong power and of the rule of law).

III. The determinants of international migration flows

In the reminder of the paper, I use the new data presented above to analyze the determinants of international migration flows. In this section, I revisit the standard gravity equation that has been estimated in the literature. In particular, I regress the log of the stock of immigrants from country j in country i at time t on different variables that measure income gains and moving costs:

$$\begin{aligned}\ln M_{ijt} =& \alpha_1 GDPpc_{it} + \alpha_2 GDPpc_{jt} + \alpha_3 \ln dist_{ij} + \alpha_4\, \mathbb{1}\{CommLang_{ij}\} + \\ &+ \alpha_5\, \mathbb{1}\{Colony_{ij}\} + \alpha_6\, \mathbb{1}\{Border_{ij}\} + \alpha_7 \ln Pop_{it} + \alpha_8 \ln Pop_{jt} + \\ &+ \text{Fixed effects} + \upsilon_{ijt}, \end{aligned} \tag{1}$$

where the different variables are described in Section II.C, and where fixed effects vary across specifications, including country of origin, destination country, year, country of origin×year, and/or country pair fixed effects. Migration is expected to be positively affected by income gains (hence, α_1 is expected to be positive and α_2, negative), by having a common language, a colonial relation, and a common border, and by the population in the origin country, and negatively affected by physical distance; the expected sign of the effect of population in the destination country is ambiguous.

This regression is micro-founded in the model by Grogger and Hanson (2011), and is comparable to the previous studies in the literature (Mayda, 2010; Grogger and Hanson, 2011; Ortega and Peri, 2013). The importance of fixed effects in such specification is noted by Bertoli and Fernández-Huertas Moraga (2013), who argue that specifications without fixed effects may suffer biases due to the "Multilateral Resistance to Migration".[18] [19]

Table 4 presents the results for the estimation of different versions of equation (1). All regressions include at least origin and destination country fixed effects, and year dummies. The first column is the baseline specification. The stock of migrants is positively associated with the GDP per capita of the destination country. This result suggest that better economic opportunities in the destination country encourage migration. In particular, everything else constant, a 1,000$ increase in GDP per capita of the destination country increases the immigrant stock by a 5.2%. This magnitude is in line, for example, with Ortega and Peri (2013), who find a positive effect of a 5-6%. According to the results in Table 4, a 10% increase in GDP per capita of the average country of destination (which is 17,848$, see Table 3) would increase the immigrant stock by a 9.3%.[20] GDP per capita in OECD countries averaged 9,101$ in 1960, and 27,341$ in year 2000. According to the results in Table 4, this 200% increase would have increased the stock of immigrants in a 95% (25 millions of immigrants over the OECD), more than a half of the actual increase (45 millions).

Theoretical predictions from models like the ones in Grogger and Hanson (2011) or in Mayda (2010) suggest that α_1 and α_2 should be similar in magnitude and

[18] These authors propose a formal test to the "Multilateral Resistance to Migration". Unfortunately, such test cannot be implemented in this context because grouped data makes the estimation of the model in differences unfeasible

[19] Several papers in the literature estimated equation (1) using the Poisson ML estimator due to the presence of a substantial number of zero observations. This concern does not apply to this paper, as the current database includes very few zero migrant stocks (countries with few immigrants are typically in grouped categories).

[20] This result is qualitatively in line with Mayda (2010), who finds that a 10% increase in destination country GDP per capita increases emigration rates by a 20%. However, these numbers are hard to compare quantitatively, as her dependent variable is expressed in flows instead of stocks.

Table 4—Determinants of bilateral migration stocks—linear effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
GDPpc dest.	0.052		0.069	0.053		0.053	0.043
	(0.020)		(0.014)	(0.062)		(0.072)	(0.017)
GDPpc origin	-0.007		-0.007		-0.009	-0.013	-0.017
	(0.007)		(0.005)		(0.019)	(0.023)	(0.008)
GDPpc gap		0.022					
		(0.009)					
Log Distance	-0.901	-0.899	-1.051	-0.926	-0.907		-0.834
	(0.063)	(0.063)	(0.036)	(0.190)	(0.155)		(0.061)
Common language	0.572	0.569	0.760	0.573	0.569		0.611
	(0.099)	(0.099)	(0.064)	(0.316)	(0.240)		(0.089)
Colonial rel.	2.286	2.281	2.107	2.289	2.272		2.399
	(0.133)	(0.132)	(0.094)	(0.415)	(0.342)		(0.120)
Common border	0.038	0.040	0.036	0.006	0.032		0.225
	(0.163)	(0.162)	(0.113)	(0.501)	(0.431)		(0.148)
Log Pop. orig.	1.365	1.088	1.466		1.345	1.250	1.747
	(0.301)	(0.288)	(0.158)		(0.631)	(1.118)	(0.305)
Log Pop. dest.	1.209	1.053	-1.902	1.137		1.146	-0.196
	(1.221)	(1.201)	(0.372)	(3.799)		(4.606)	(1.014)
Grouped obs.	Yes	Yes	No	Yes	Yes	Yes	Yes
St.devs. of controls	No	No	No	No	No	No	Yes
Time dummies	Yes	Yes	Yes	No	No	Yes	Yes
Origin dummies	Yes	Yes	Yes	No	Yes	No	Yes
Origin-time dummies	No	No	No	Yes	No	No	No
Destination dummies	Yes	Yes	Yes	Yes	No	No	Yes
Dest.-time dummies	No	No	No	No	Yes	No	No
Country-pair dummies	No	No	No	No	No	Yes	No
Obs	7,339	7,339	6,727	7,428	7,339	7,339	7,339
$\bar{R}^2$	0.96	0.96	0.97	0.87	0.93	0.88	0.97

Note: Robust standard errors in parentheses. Dependent variable: log migrant stocks. Unit of observation: origin-destination-time. Regressions include the specified fixed effects, as indicated.

of opposite sign. However, Table 4 shows a much smaller effect of origin country GDP per capita. Although it is negative (consistently in all specifications), the coefficient is far from being significantly different from zero, and point estimates are one order of magnitude smaller than destination country counterparts. This result is not new; Mayda (2010) also finds a non-significant effect, although her point estimates are indeed positive. This finding could result from an additional positive effect of origin country GDP per capita on migration prospects. Borrowing constraints could be a plausible explanation: if individuals from poorer countries (lower GDP per capita) are financially constrained, then, other things equal, their chances to migrate are lower; therefore, the larger the GDP per capita, the less constrained they are, and the larger is the probability that they migrate. If that were the case, one would expect that this effect should be homogeneous across all destination countries, which is in line with findings discussed in Section IV.

Physical and cultural distance play an important role in explaining moving costs.

The elasticity of the migrant stock with respect to physical distance is about 0.9. Having a common language or a colonial relationship increases importantly the stock of immigrants. A common border, however, seems less important. These results are, again, qualitatively similar to Mayda (2010), Grogger and Hanson (2011), and Ortega and Peri (2013).

Finally, we cannot reject neither that the coefficient of log population in the origin country is equal to one, nor that the one of log population in the destination country is zero, which are the values predicted by a model like the one described in Grogger and Hanson (2011). If anything, they are larger than the expected values, which would be consistent with capacity constraints from destination countries in absorbing immigrants and from origin countries in supplying them.

The remaining columns of Table 4 check the stability of the estimates across different versions of the same equation. In order to obtain estimates which are fully comparable to Grogger and Hanson (2011), in column (2) I impose the same coefficient (of opposite sign) for origin and destination countries' GDP per capita. The coefficient of income gap is 0.022 (s.e. 0.009) very close to their estimate of 0.018 (s.e. 0.029) and much more precisely estimated, given the larger coverage by the dataset presented in this paper. Additionally, the coefficients for the variables associated with moving costs are extremely similar. As I mentioned before, data used in Grogger and Hanson (2011) is obtained from Docquier and Marfouk (2006), who collect census based data in the same spirit of the database presented in this paper. Grogger and Hanson (2011) estimate their regressions using data for year 2000. The fact that coefficients are that similar indicate three things. First, the quality of the data presented in this paper is similar to the one in Docquier and Marfouk (2006), widely used in the literature, although it covers a much wider time period. Second, parameter estimates seem to be very stable over time, even distance coefficients —which one may expect to have a smaller effect in recent decades. And third, the fact that some observations are grouped does not seem to be an issue for the estimation of the model.

In order to analyze the importance of including the 100% of migrant stocks, I drop grouped observations in column (3). Although qualitative results hold, point estimates are somewhat different. In particular, four out of the eight coefficients are statistically different from point estimates in column (1), and a Wald test of the null hypothesis that all eight coefficients are equal to their counterparts in column (1) clearly rejects it. Therefore, we can conclude that including grouped observations —so that we cover the 100% of total migrant stocks— is very important to obtain unbiased estimates.

In columns (4) to (6), I change the specification of fixed effects. On top of origin, destination, and time fixed effects that are included in columns (1) to (3), I enrich the analysis by adding destination×time, origin×time, and country pair dummies respectively. These specifications are more demanding in terms of degrees of freedom (see discussion on Figure 1). Point estimates are virtually unchanged (they are identical up to the second or third decimal), although, by construction, precision drops. This stability of the coefficients is very interesting, as each specification controls somewhat for different versions of migration policies that may affect the results. Ortega and Peri (2013) show that a specification like the one in column (4) —which includes country of origin×year dummies— emerges from a version of the model in Grogger and Hanson (2011) in which individual fixed effects are allowed for in the utility function.

A problem of having some observations aggregated in grouped categories is that, since we only observe the stock of immigrants for the group, the dependent variable is measured with error provided that the log of the average stock of the group is not equal to the average of logs of bilateral stocks. The problem with this measurement error is that it is obviously correlated with the covariates. In order to check to what extent this could be a relevant issue, in column (7) I include as controls standard deviations of the regressors within the grouped observations (zero for bilateral observations). Given that the measurement error increases as the countries in the grouped observation differ in the stock of immigrants, these standard deviations are good proxies for the measurement error.[21] Results are again robust; none of the coefficients is statistically different from their column (1) equivalents, and the test of the joint difference cannot reject the null hypothesis that all coefficients are (jointly) equal to their counterparts in column (1). The only qualitative change is that origin country GDP per capita becomes significantly different from zero (although still smaller than the coefficient of destination country GDP per capita, and not significantly different from point estimates in other columns).

A final robustness check for this specification is in Table A1 from Appendix A. We have seen in Section II that data is particularly grouped in years 1960 and 1970 (especially the former), and that data reliability is slightly lower in those years (see footnote 12). For this reason, in Table A1 I repeat all regressions from in Table 4 excluding observations for 1960 first, and then for 1960 and 1970. Results are again robust, confirming also the stability of the coefficients over time.

[21] Given that the logarithm is a concave function, by the Jensen inequality we know that the logarithm of the average of the group will be larger than the average of the logarithm unless all elements of the average are equal.

In sum, we can conclude that GDP per capita at the destination country has a large and significant effect, much larger than the effect of income from the origin country, which is negligible. The finding that the magnitude of the coefficient of origin country GDP per capita is smaller than the one of GDP per capita of destination country is somewhat against the predictions of the theory. However, this could be the result of an indirect positive effect of origin country income per capita on migration prospects: the presence of more severe borrowing constraints in poorer countries. From the comparison with the existing literature, we can extract three conclusions: first, the database presented in this paper provides comparable results to previous findings in the literature, although estimates presented above are more precise; second, the effect of income gains and moving costs on migration has been very stable over time (results are not affected by dropping 1960 and 1970, and they are in line with previous studies that cover a much shorter time period); and third, the differences between long-run and short/medium run adjustments seem to be small (given that the results of this paper, with 10-year periods are similar to others using an annual frequency).

IV. Heterogeneous effects of income gains

A. Heterogeneous effects of income gains depending on distance

Different papers in the literature estimate regressions like equation (1) to analyze the determinants of migration flows. An important implication of that model is that an increase in GDP per capita of a destination country would increase the stock of immigrants from all origin countries by the same percentage. Likewise, an increase in the GDP per capita of a given country of origin would increase the stock of migrants from that country into all destinations by the same relative amount.

It is not implausible, however, that the effect of income shocks on moving prospects is more marked for closer countries, compared countries that are farther apart. This heterogeneous effect of income gains depending on distance can be motivated by two factors. First, large moving costs (distance) reduce the flexibility of individuals to move back and forth to their home country when income changes. As a result, in the migration decision, individuals from farther countries will give more weight to long run income (as opposed to income shocks), whereas individuals from neighboring countries will be more prone to go back and forth to take advantage of income fluctuations. Second, if individuals dislike living far away from home, they might require a compensating wage differential to offset the unpleasantness of living abroad.[22] If the disutility of being far from home increases with distance, they will require an increasing wage premium to take de decision to migrate. Hence, these compensating wage differentials would also introduce an heterogeneous effect of income gains on moving prospects depending on distance that would make migration more reactive to income at closer distances.

For these reasons, I modify equation (1) to allow for these heterogeneous effects:

$$\begin{aligned}\ln M_{ijt} = \gamma_1 GDPpc_{it} + \gamma_2 GDPpc_{jt} + \gamma_3 \widetilde{GDPpc}_{kt} \widetilde{\ln dist}_{jk} + \gamma_4 \widetilde{GDPpc}_{jt} \widetilde{\ln dist}_{jk} + \\ + \gamma_5 \ln dist_{ij} + \gamma_6\, \mathbb{1}\{CommLang_{ij}\} + \gamma_7\, \mathbb{1}\{Colony_{ij}\} + \gamma_8\, \mathbb{1}\{Border_{ij}\} + \\ + \gamma_9 \ln Pop_{it} + \gamma_{10} \ln Pop_{jt} + \text{Fixed effects} + \upsilon_{ijt}, \qquad (2)\end{aligned}$$

where $\widetilde{x} \equiv x - \bar{x}$ indicates that variables are in deviations with respect to sample means. Parameters γ_3 and γ_4 allow for the heterogeneous response of migration to shocks to destination and/or origin country's income as a function of distance.

Estimates for equation 2 are presented in Table 5. As the interacted terms in equation (2) are expressed in differences from sample means, the linear terms can be interpreted as effects for the average country pair (and they are comparable to estimates in Section III). Again, all regressions include at least origin and destination country fixed effects, and year dummies.

Column (1) in Table 5 is the baseline specification. The effect of destination country GDP per capita for the average country is slightly smaller than in Table 4, although the difference is not significant. Point estimates suggest that a 1,000$ increase in the GDP per capita of a given destination country increases the stock of immigrants from a country located at the average distance between all country pairs (5,285 Km) by a 4.6% —as opposed to the 5.2% from Table 4.

As the coefficient of the interaction of destination country GDP per capita and distance suggests, this effect is not homogeneous across all origin countries. These coefficients are interpreted as follows: the effect of a 1,000$ increase in GDP per capita of the destination country is 0.23 percentage points smaller if the distance from the origin country is a 10 percent larger than the average. To give a sense to these numbers, note that the distance between Washington DC (U.S.) and Dublin (Ireland) is 5,448Km, roughly the average distance in the sample. On the other hand, the distance between Washington DC and Beijing (China) is 11,159Km, roughly twice as large. Therefore, a 1,000$ increase in GDP per capita in the U.S.

[22] For example, the farther they live, the higher the incidence of homesickness is, the lower are the chances of regular interaction with relatives and friends (e.g. different time frames, less opportunities to travel home from time to time,...), and, when we consider a broad definition of distance —e.g. cultural distance—, the more difficult they find to interact with locals in social life.

Table 5—Heterogeneous effects of income gains

	(1)	(2)	(3)	(4)	(5)	(6)
GDPpc dest.	0.046	0.045	0.071	0.046		0.044
	(0.020)	(0.019)	(0.014)	(0.060)		(0.070)
GDPpc dest.× Log Distance	-0.023	-0.027	-0.014	-0.022	-0.022	-0.033
	(0.005)	(0.006)	(0.004)	(0.021)	(0.013)	(0.027)
GDPpc dest.× Common lang.		-0.009	-0.023	-0.021	-0.009	0.021
		(0.021)	(0.008)	(0.083)	(0.034)	(0.130)
GDPpc dest.× Colonial rel.		-0.063	-0.019	-0.070	-0.026	-0.076
		(0.027)	(0.012)	(0.092)	(0.044)	(0.151)
GDPpc dest.× Common border		-0.006	-0.002	0.002	0.011	-0.047
		(0.024)	(0.016)	(0.076)	(0.060)	(0.160)
GDPpc origin	-0.004	-0.000	-0.002		-0.003	-0.006
	(0.007)	(0.008)	(0.006)		(0.021)	(0.026)
GDPpc orig.× Log Distance	0.032	0.028	0.028	0.032	0.028	0.018
	(0.004)	(0.004)	(0.003)	(0.016)	(0.010)	(0.031)
GDPpc orig.× Common lang.		-0.003	0.000	0.001	-0.002	-0.017
		(0.009)	(0.006)	(0.028)	(0.020)	(0.104)
GDPpc orig.× Colonial rel.		-0.024	-0.042	-0.032	-0.026	-0.006
		(0.009)	(0.009)	(0.032)	(0.023)	(0.117)
GDPpc orig.× Common border		-0.016	-0.010	-0.012	-0.012	-0.011
		(0.021)	(0.017)	(0.065)	(0.061)	(0.158)
Log Distance	-0.930	-0.926	-1.173	-0.994	-0.942	
	(0.067)	(0.068)	(0.041)	(0.241)	(0.166)	
Common language	0.592	0.568	0.841	0.584	0.577	
	(0.099)	(0.143)	(0.083)	(0.561)	(0.321)	
Colonial rel.	2.224	2.299	2.141	2.418	2.261	
	(0.132)	(0.151)	(0.101)	(0.520)	(0.369)	
Common border	0.388	0.551	0.387	0.443	0.464	
	(0.155)	(0.245)	(0.177)	(0.780)	(0.671)	
Log Pop. orig.	1.426	1.468	1.088		1.366	1.542
	(0.313)	(0.318)	(0.188)		(0.656)	(1.221)
Log Pop. dest.	1.481	1.536	-1.882	1.406		1.546
	(1.224)	(1.281)	(0.399)	(4.351)		(5.002)
Grouped obs.	Yes	Yes	No	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	No	No	Yes
Origin dummies	Yes	Yes	Yes	No	Yes	No
Origin-time dummies	No	No	No	Yes	No	No
Destination dummies	Yes	Yes	Yes	Yes	No	No
Dest.-time dummies	No	No	No	No	Yes	No
Country-pair dummies	No	No	No	No	No	Yes
Obs	7,339	7,339	6,727	7,339	7,339	7,339
$\bar{R}^2$	0.96	0.96	0.97	0.88	0.93	0.88

Note: Robust standard errors in parentheses. Dependent variable: log migrant stocks. Unit of observation: origin-destination-time. Regressions include the specified fixed effects, as indicated.

would increase the stock of Irish living in the U.S. by around a 4.6%, whereas the stock of Chinese-born would only be increased by approximately 2.3%. As an extreme example, a 1,000$ increase in GDP per capita in the U.S. would increase the stock of Mexicans by a 7.7%, but the stock of Taiwanese would be increased by only a 1.98%.

This is the main empirical result of this paper. Previous literature assumes that an income shock in a destination country increases the stock of immigrants from all origin countries by the same percentage. If that were the case, then income shocks would not affect the composition of the immigrant population. But the finding described above indicates that income shocks in a destination country have indeed very important compositional effects. This result is very important for shaping immigration policy. For example, if the policy maker is willing to preserve the ethnic mix (e.g. it was one of the goals of the U.S. immigration policy from 1920s to mid-1960s), countermeasures will be required to compensate market forces. Additionally, if the skill composition of immigrants from a particular country of origin was not affected by changes in the size of the flow, income shocks would affect the skill composition of the immigrant workforce by changing the weight of each origin country in the total stock.

A similar story can be told for origin countries' GDP per capita. Average effects are virtually unchanged with respect to the linear specification in Section III, but the interaction with distance is very important. Interestingly, the coefficient of this interaction is very similar —with the opposite sign— to the one for interaction of GDP per capita of the destination country and distance (indeed, we cannot reject statistically that their magnitudes are the same). This result, together with the small estimated coefficient for the linear term, are again suggestive of the presence of an additional effect of origin country GDP per capita on migration prospects. Following with the argument of borrowing constraints, imperfect access to credit markets in poorer countries would prevent migrants from these countries to afford the migration cost, although they would have gained from moving if they could have borrowed resources to afford it; if that were the case, credit market imperfections would increase the coefficient of the linear term (making it less negative), but would not affect the interaction term. Similarly, another positive direct effect of origin country GDP per capita could arise through immigration policies, if destination countries are more willing to accept immigrants from richer countries (which again would not affect the interaction term).[23]

The remaining coefficients are virtually equal to their counterparts in Table 4. The only exception is the coefficient for common border —0.388 (s.e. 0.155) vs 0.038 (s.e. 0.163). The elasticity of physical distance is still around 0.9, and sharing a common language or a past colonial relationship increase importantly the

[23] One could argue that immigration policy is softer in destination countries in "good periods". This would tend to produce a larger linear effect of destination country GDP per capita, but it would not affect the interaction term.

stock of migrants. Finally, the coefficient of log origin country population is again not statistically different from one and that of destination country population is not different from zero as predicted by the theory.

The remaining columns of Table 5 check the stability of the estimates across different versions of the same equation. In column (2) I extend equation (2) by including interactions of origin and destination country GDP per capita with all other measures of distance. Results are virtually unchanged. Only interactions with colonial relationship are significant. Surprisingly, both of them have a negative sign. This result, however, may be driven by policy issues as one would expect that (after controlling for having a common language) a past colonial relationship only affects migration through a special treatment by destination countries in terms of immigration policy. For example, a negative income shock would reduce the stock of immigrants from non-former colonies in a larger magnitude than from former colonies, which would receive a special treatment.

In column (3), I check the importance of including the 100% of migrant stocks by dropping grouped observations. As in Table 4, qualitative results hold, but point estimates are different. In particular, seven coefficients are statistically different from their counterparts in column (2), and a Wald test of the hypothesis that all coefficients are equal to their counterparts in column (2) clearly rejects.

As in Table 4, in columns (4) to (6) I change the specification of fixed effects. Again, on top of origin, destination, and time fixed effects (as in columns (1) to (3)), I introduce destination×time, origin×time, and country pair dummies respectively. As I mentioned above, these specifications are more demanding in terms of degrees of freedom. Again, point estimates are virtually unchanged but precision drops dramatically.

Finally, as in Section III, I estimate all regressions in Table 5 excluding 1960 and excluding 160 and 1970. Results, presented in Table A2 in Appendix A, do not show substantial differences.

B. *Additional results for other* push *and* pull *factors*

In Table 6, I extend regression (2) to analyze other *push* and *pull* determinants of migration in more detail. Specifically, I add unemployment rate, age dependency ratio (older than 65 over working-age population), and government consumption share of GDP (pull factors), and wars and political regimes (push factors).

Aside from income gains, individuals value their probability of finding a job in the destination country. For this reason, higher unemployment at the country of destination reduces migration. Column (1) shows this empirically by includ-

Table 6—Additional results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDPpc dest.	0.035	0.046	0.041	0.025	0.046	0.044	0.045	0.024
	(0.018)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
GDPpc dest.× Log Distance	-0.021	-0.023	-0.023	-0.023	-0.022	-0.022	-0.021	-0.022
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Unemployment rate dest. (%)	-0.106			-0.080				-0.085
	(0.029)			(0.033)				(0.033)
Age dep. dest.(% old/w-age)		0.016		0.034				0.038
		(0.035)		(0.033)				(0.034)
Government share dest.			-0.085	-0.088				-0.081
			(0.069)	(0.076)				(0.076)
GDPpc origin	-0.010	-0.004	-0.004	-0.010	-0.003	-0.003	-0.002	-0.008
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
GDPpc orig.× Log Distance	0.033	0.032	0.032	0.032	0.032	0.033	0.032	0.032
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
War origin					0.772		0.718	0.607
					(0.155)		(0.154)	(0.146)
PolityIV origin						-0.006	-0.006	-0.005
						(0.012)	(0.012)	(0.010)
PolityIV^2 origin						-0.003	-0.003	-0.003
						(0.001)	(0.001)	(0.001)
Log Distance	-0.963	-0.930	-0.930	-0.957	-0.931	-0.934	-0.935	-0.960
	(0.069)	(0.067)	(0.066)	(0.068)	(0.067)	(0.069)	(0.069)	(0.070)
Common language	0.635	0.592	0.590	0.635	0.592	0.602	0.601	0.636
	(0.094)	(0.099)	(0.098)	(0.092)	(0.099)	(0.098)	(0.098)	(0.091)
Colonial rel.	2.196	2.223	2.222	2.192	2.228	2.158	2.162	2.137
	(0.131)	(0.132)	(0.130)	(0.129)	(0.131)	(0.131)	(0.130)	(0.127)
Common border	0.320	0.388	0.384	0.317	0.382	0.414	0.410	0.353
	(0.154)	(0.155)	(0.152)	(0.151)	(0.156)	(0.155)	(0.156)	(0.151)
Log Pop. orig.	1.234	1.433	1.437	1.274	1.400	1.392	1.377	1.247
	(0.293)	(0.313)	(0.311)	(0.292)	(0.311)	(0.293)	(0.292)	(0.276)
Log Pop. dest.	0.998	1.557	0.642	0.162	1.477	1.487	1.483	0.161
	(1.348)	(1.264)	(1.253)	(1.252)	(1.223)	(1.222)	(1.222)	(1.259)
Grouped obs.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs	7,093	7,339	7,337	7,091	7,339	7,143	7,143	6,895
$\bar{R}^2$	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.97

Note: Robust standard errors in parentheses. Dependent variable: log migrant stocks. Unit of observation: origin-destination-time. Regressions include the specified fixed effects, as indicated.

ing unemployment rate in the regression. Its effect is estimated to be negative, as expected, and very significant. Columns (2) includes age dependence ratio as a regressor. Countries with older populations are more willing to admit immigrants to increase social security revenues and sustain increasingly unbalanced pay-as-you-go systems. Additionally, an older population brings in additional work opportunities for immigrants, both in terms of elderly caring services and because of a lower competition in the labor market. The coefficient of this variable has the expected positive sign, although its effect is small and statistically not different from zero. In column (3) I include the government consumption as a share of GDP. More generous welfare state governments will spend more, and will attract more immigrants. However, larger government expenditure implies higher tax rates, and this may discourage migration. If all countries were equally efficient in their spending, the sign of the effect should depend on whether immigrants are net contributors or receivers. In that case, South-North migration should be affected positively by expenditure. However, larger expenditures in some countries may be due to lower efficiency, which might imply that everyone becomes a net contributor, making the effect unambiguously negative. Results in Table 6 suggest that the effect is negative but very small and not significant. All three *pull* factors together are included together in column (4). None of the coefficients change significantly, and the same conclusions hold.

Column (5) includes a warfare measure for the origin country. This variable measures the share of months over the last decade that the country was involved in a war of any type. Armed conflicts displace a lot of people who escape from the tragedy. This fact is reflected in the estimates: a decade of war in an origin country increases the stock of immigrants from that country in a 77%. The political regime is also important for migration. People is less willing to leave a good democracy (everything else constant); moreover, in a dictatorship, they are usually not allowed to escape from the country. The situations of weak central authorities (known as anocracies) seem to be an encouraging environment for migration. In column (6), I introduce the Polity IV index, which ranges from -10 (autocracy) to 10 (democracy). Intermediate values (with small absolute values)indicate the presence of an anocracy. For this reason, I include a quadratic in the indicator. The quadratic term is negative and significantly different from zero. The linear term is negative but small, indicating that slightly more people migrate from autocracies than from democracies. In particular, the stock of migrants is a 36% lower if the origin country is a democracy than if it is an anocracy, whereas it is only a 24% lower if the origin country is an autocracy instead of an anocracy. In column (7) I introduce all *push* factors together, obtaining the same results. Finally, column (8) includes both *push* and *pull* factors all together. Again, results are unchanged.

V. Conclusions

In this paper I present a new database of bilateral migrant stocks, and I provide new evidence on the determinants of bilateral migration. The database introduced in this paper was collected from the National Statistical Offices from 24 OECD countries based on population censuses. For each destination country and census date, it covers 188 countries of origin (sometimes in a grouped category). The dataset has four important advantages compared to others in the literature. First, it covers a longer time period (1960-2000). Second, being based on censuses, it reduces the undercounting of undocumented immigrants. Third, the information is based on stocks as opposed to flows; working with stock data is more attractive both because equilibrium values are often expressed in terms of stocks, and because flow data are less reliable as a result of the impossibility of quantifying emigration and return migration flows (Docquier and Marfouk, 2006). And fourth, unlike previously used databases, it fully covers the total stock of immigrants, keeping track of the residual categories that Statistical Offices often present as a group of countries of origin.

Empirically, I test for the existence of non-linear effects of income gains on migration prospects depending on distance. The motivation for such heterogeneity can be cost-based (individuals from closer countries can move back and forth as a consequence of income fluctuations, whereas it is more costly for individuals from farther countries), or by means of a compensating wage differential (individuals dislike living far away from home, and require a compensating wage differential to move, that would increase with distance). Results suggest that this heterogeneity is indeed very marked. For example, a 1,000$ increase in U.S. income per capita would increase the stock of Mexican immigrants in the U.S. by a 7.7%, the stock of Irish immigrants by a 4.6%, and the stock of Chinese-born by only a 2.3%. This result is very robust to many different specifications.

Empirical findings in this paper suggest that income shocks have significant compositional effects, which are important for shaping immigration policy. For example, if a policy maker is willing to preserve the ethnic mix (e.g. it was one of the goals of the U.S. immigration policy from 1920s to mid-1960s), countermeasures will be required to compensate market forces. If country of origin is a good proxy for skills of immigrants, this result would also have implications for the skill composition of migrants. Additionally, destination countries should be more concerned about income shocks in neighboring countries than what is suggested in the literature, and may want to target nearby countries in their development assistance policies if they want to mitigate migration flows.

Two limitations of the current analysis worth a mention here. The first one is the grouping of the data. Although I am able to identify and precisely estimate the standard fixed effects specifications even with the presence of grouped observations, I estimate some more demanding fixed effects models (those including origin-time and country-pair dummies) for which I obtain quite large standard errors due to the data grouping. For example, there are 1,800 country pairs (out of

24×188=4,512) for which I observe data in grouped categories for all years (which only allows me to identify a fixed effect for each group). Also, for a similar reason, there are 160 origin country×time dummies that cannot be individually identified (out of 188×5=940). Finally, this data grouping also prevents me of being able to incorporate the role of networks in determining migration flows (e.g. Munshi, 2003; McKenzie and Rapoport, 2010; Pedersen, Pytlikova and Smith, 2008).

The second limitation is that the database does not include information on educational attainment by immigrants. Such information would be useful to test whether the compositional effects that I observe with respect to nationality have important implications for skill composition of immigrants. This is not essential for the results of the paper, but it would be a nice extension. To the best of my knowledge, Docquier and Marfouk (2006) and Docquier, Lowell and Marfouk (2009) are the only databases in the literature that include such information, although they only cover 1990 and 2000.

The paper also opens avenues for future research. It would be interesting to investigate how the heterogeneous effects found in this paper affect skill composition of immigrants. These heterogeneous effects might also be included in existing migration models to analyze, for instance, their implication in terms of the self-selection of migrants. And, finally, the database presented in this paper can be used for a variety of cross-country migration analysis (e.g., instrumental variables like in Llull (2011) or Ortega and Peri (2012)).

REFERENCES

Adsera, Alicia and Mariola Pytlikova, "The Role of Language in Shaping International Migration," NORFACE Migration Discussion Paper No. 2012-14, January 2012.

Alesina, Alberto F., Arnaud Devleeschauwer, William E. Easterly, Sergio Kurlat, and Romain Wacziarg, "Fractionalization," *Journal of Economic Growth*, June 2003, *8* (2), 155–194.

Bertoli, Simone and Jesús Fernández-Huertas Moraga, "Multilateral Resistance to Migration," *Journal of Development Economics*, May 2013, *102*, 79–100.

_ , _ , and Francesc Ortega, "Crossing the border: Self-selection, earnings and individual migration decisions," *Journal of Development Economics*, March 2013, *101*, 75–91.

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

_ **and Bernt Bratsberg**, "Who Leaves? The Outmigration of the Foreign-Born," *Review of Economics and Statistics*, February 1996, *78* (1), 165–176.

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.

Clark, Ximena, Timothy J. Hatton, and Jeffrey G. Williamson, "Explaining U.S. Immigration, 1971-1998," *Review of Economics and Statistics*, May 2007, *89* (2), 359–373.

Docquier, Frédéric and Abdeslam Marfouk, "International Migration by Educational Attainment, 1990-2000," in Çağlar Özden and Maurice W. Schiff, eds., *International Migration, Remittances and the Brain Drain*, New York: Palgrave Macmillan, 2006, chapter 5, pp. 151–200.

_ **, B. Lindsay Lowell, and Abdeslam Marfouk**, "A Gendered Assesment of Highly Skilled Emigration," *Population and Development Review*, June 2009, *35* (2), 297–321.

Grogger, Jeffrey T. and Gordon H. Hanson, "Income maximization and the selection and sorting of international migrants," *Journal of Development Economics*, May 2011, *95* (1), 42–57.

Karemera, David, Victor Iwuagwu Oguledo, and Bobby Davis, "A gravity model analysis of international migration to North America," *Applied Economics*, October 2000, *32* (13), 17451755.

Kennan, John and James R. Walker, "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, January 2011, *79* (1), 211–251.

Lessem, Rebecca H., "Mexico-U.S. Immigration: Effects of Wages and Border Enforcement," mimeo, Carneggie Melon University, July 2013.

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

Mayda, Anna Maria, "International migration: a panel data analysis of the determinants of bilateral flows," *Journal of Population Economics*, September 2010, *23* (4), 1249–1274.

McKenzie, David and Hillel Rapoport, "Self-Selection Paterns in Mexico-U.S. Migration: the Role of Migration Networks," *Review of Economics and*

Statistics, November 2010, *92* (4), 811–821.

Munshi, Kaivan, "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market," *Quarterly Journal of Economics*, May 2003, *118* (2), 549–597.

Ortega, Francesc and Giovanni Peri, "The Causes and Effects of International Migrations: Evidence from OECD Countries 1980-2005," NBER Working Paper n. 14833, April 2009.

_ **and** _ , "The Effect of Trade and Migration on Income," NBER Working Paper No. 18193, June 2012.

_ **and** _ , "The effect of income and immigration policies on international migration," *Migration Studies*, March 2013, *1* (1), 47–74.

Pedersen, Peder J., Mariola Pytlikova, and Nina Smith, "Selection and network effects —Migration flows into OECD countries 1990-2000," *European Economic Review*, October 2008, *52* (7), 1160–1186.

Rose, Andrew K., "Do We Really Know That the WTO Increases Trade?," *American Economic Review*, March 2004, *94* (1), 98–114.

Appendix A: Robustness: exclusion of 1960 and 1970

Table A1—Determinants of bilateral migration stocks—linear effects (excluding 1960 and 1970)

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
GDPpc dest.	0.043	0.052			0.049	0.044	0.043	0.052			0.043	0.052	0.033	0.040
	(0.018)	(0.018)			(0.015)	(0.017)	(0.048)	(0.040)			(0.053)	(0.041)	(0.017)	(0.018)
GDPpc origin	-0.015	-0.012			-0.007	-0.004			-0.014	-0.011	-0.018	-0.014	-0.022	-0.017
	(0.008)	(0.007)			(0.005)	(0.005)			(0.018)	(0.014)	(0.020)	(0.015)	(0.009)	(0.008)
GDPpc gap			0.023	0.022										
			(0.008)	(0.007)										
Log Distance	-0.961	-1.034	-0.961	-1.035	-1.105	-1.161	-0.967	-1.033	-0.963	-1.034			-0.956	-1.037
	(0.066)	(0.067)	(0.066)	(0.068)	(0.038)	(0.040)	(0.172)	(0.147)	(0.143)	(0.122)			(0.062)	(0.064)
Common language	0.663	0.664	0.662	0.663	0.785	0.784	0.657	0.669	0.663	0.664			0.723	0.732
	(0.091)	(0.093)	(0.092)	(0.093)	(0.066)	(0.071)	(0.251)	(0.210)	(0.202)	(0.178)			(0.081)	(0.085)
Colonial rel.	2.210	2.212	2.209	2.212	2.082	2.061	2.216	2.209	2.208	2.211			2.389	2.322
	(0.133)	(0.145)	(0.133)	(0.145)	(0.099)	(0.110)	(0.357)	(0.320)	(0.297)	(0.281)			(0.120)	(0.130)
Common border	-0.135	-0.271	-0.134	-0.270	-0.148	-0.261	-0.138	-0.271	-0.136	-0.271			-0.044	-0.202
	(0.166)	(0.175)	(0.166)	(0.176)	(0.120)	(0.134)	(0.446)	(0.392)	(0.389)	(0.352)			(0.155)	(0.164)
Log Pop. orig.	0.797	0.428	0.642	0.237	1.278	1.217			0.820	0.440	0.672	0.441	1.124	0.434
	(0.312)	(0.386)	(0.312)	(0.374)	(0.176)	(0.210)			(0.631)	(0.664)	(0.914)	(0.909)	(0.321)	(0.361)
Log Pop. dest.	2.490	1.478	2.418	1.496	-1.178	-2.000	2.468	1.467			2.468	1.471	0.876	0.164
	(1.735)	(1.975)	(1.720)	(1.999)	(0.513)	(0.666)	(4.679)	(4.469)			(5.685)	(5.626)	(1.453)	(1.455)
Excludes 1960	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Excludes 1970	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Grouped obs.	Yes		Yes		No		Yes		Yes		Yes		Yes	
St.devs. of controls	No		No		No		No		No		No		Yes	
Time dummies	Yes		Yes		Yes		No		No		Yes		Yes	
Origin dummies	Yes		Yes		Yes		No		Yes		No		Yes	
Origin-time dummies	No		No		No		Yes		No		No		No	
Destination dummies	Yes		Yes		Yes		Yes		No		No		Yes	
Dest.-time dummies	No		No		No		No		Yes		No		Yes	
Country-pair dummies	No		No		No		No		No		Yes		No	
Obs	6,765	6,070	6,765	6,070	6,272	5,707	6,810	6,083	6,765	6,070	6,765	6,070	6,765	6,070
$\bar{R}^2$	0.97	0.97	0.97	0.97	0.97	0.97	0.91	0.92	0.94	0.94	0.93	0.95	0.97	0.97

Note: Robust standard errors in parentheses. Dependent variable: log migrant stocks. Unit of observation: origin-destination-time. Regressions include the specified fixed effects, as indicated.

Table A2—Heterogeneous effects of income gains(excluding 1960 and 1970)

	(1)		(2)		(3)		(4)		(5)		(6)	
GDPpc dest.	0.038	0.051	0.035	0.047	0.048	0.050	0.038	0.054			0.030	0.041
	(0.018)	(0.018)	(0.018)	(0.018)	(0.015)	(0.017)	(0.049)	(0.041)			(0.055)	(0.044)
GDPpc dest.× Log Distance	-0.016	-0.005	-0.016	-0.005	-0.008	-0.005	-0.012	0.000	-0.012	-0.005	-0.033	-0.020
	(0.005)	(0.005)	(0.006)	(0.006)	(0.004)	(0.004)	(0.017)	(0.015)	(0.012)	(0.011)	(0.023)	(0.023)
GDPpc dest.× Common lang.			-0.047	-0.083	-0.035	-0.043	-0.053	-0.079	-0.038	-0.059	-0.014	-0.067
			(0.023)	(0.019)	(0.010)	(0.012)	(0.071)	(0.044)	(0.029)	(0.030)	(0.120)	(0.106)
GDPpc dest.× Colonial rel.			-0.015	-0.056	-0.010	-0.013	-0.016	-0.058	-0.017	-0.039	-0.014	-0.082
			(0.019)	(0.023)	(0.014)	(0.018)	(0.056)	(0.051)	(0.039)	(0.043)	(0.092)	(0.079)
GDPpc dest.× Common border			0.009	0.016	0.008	0.002	0.018	0.031	0.014	0.018	-0.034	-0.017
			(0.023)	(0.024)	(0.017)	(0.018)	(0.063)	(0.057)	(0.055)	(0.050)	(0.101)	(0.093)
GDPpc origin	-0.012	-0.009	-0.010	-0.006	-0.005	-0.004			-0.010	-0.006	-0.014	-0.007
	(0.008)	(0.007)	(0.008)	(0.007)	(0.006)	(0.006)			(0.018)	(0.014)	(0.020)	(0.014)
GDPpc orig.× Log Distance	0.032	0.033	0.031	0.032	0.032	0.033	0.033	0.034	0.030	0.032	0.019	0.015
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.013)	(0.011)	(0.009)	(0.007)	(0.021)	(0.017)
GDPpc orig.× Common lang.			-0.007	-0.009	-0.003	-0.001	-0.004	-0.004	-0.004	-0.004	-0.027	-0.038
			(0.009)	(0.008)	(0.006)	(0.006)	(0.024)	(0.018)	(0.017)	(0.014)	(0.076)	(0.088)
GDPpc orig.× Colonial rel.			-0.027	-0.022	-0.039	-0.035	-0.032	-0.023	-0.028	-0.026	-0.011	-0.014
			(0.010)	(0.008)	(0.009)	(0.009)	(0.029)	(0.019)	(0.020)	(0.016)	(0.101)	(0.055)
GDPpc orig.× Common border			-0.005	-0.001	-0.002	0.000	-0.002	0.002	0.000	0.004	-0.015	-0.013
			(0.022)	(0.023)	(0.017)	(0.019)	(0.057)	(0.050)	(0.054)	(0.049)	(0.109)	(0.087)
Log Distance	-1.010	-1.160	-1.007	-1.158	-1.267	-1.360	-1.053	-1.199	-1.024	-1.160		
	(0.075)	(0.080)	(0.076)	(0.082)	(0.046)	(0.052)	(0.217)	(0.189)	(0.163)	(0.146)		
Common language	0.676	0.670	0.884	1.170	0.974	1.071	0.905	1.139	0.830	1.010		
	(0.091)	(0.092)	(0.171)	(0.176)	(0.098)	(0.126)	(0.531)	(0.407)	(0.285)	(0.299)		
Colonial rel.	2.149	2.165	2.107	2.252	2.074	2.083	2.170	2.335	2.144	2.254		
	(0.131)	(0.141)	(0.162)	(0.191)	(0.114)	(0.149)	(0.476)	(0.436)	(0.329)	(0.364)		
Common border	0.251	0.177	0.332	0.187	0.165	0.124	0.229	0.048	0.237	0.085		
	(0.157)	(0.168)	(0.272)	(0.348)	(0.207)	(0.274)	(0.742)	(0.770)	(0.660)	(0.727)		
Log Pop. orig.	0.779	0.285	0.837	0.365	0.999	0.899			0.789	0.352	1.070	0.699
	(0.322)	(0.378)	(0.322)	(0.371)	(0.186)	(0.217)			(0.632)	(0.628)	(1.038)	(0.971)
Log Pop. dest.	2.712	1.493	2.916	1.818	-1.114	-1.593	2.957	2.077			3.280	2.152
	(1.725)	(1.956)	(1.799)	(1.968)	(0.530)	(0.676)	(5.037)	(4.430)			(6.142)	(5.683)
Excludes 1960	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Excludes 1970	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Grouped obs.	Yes		Yes		No		Yes		Yes		Yes	
Time dummies	Yes		Yes		Yes		No		No		Yes	
Origin dummies	Yes		Yes		Yes		No		Yes		No	
Origin-time dummies	No		No		No		Yes		No		No	
Destination dummies	Yes		Yes		Yes		Yes		No		No	
Dest.-time dummies	No		No		No		No		Yes		No	
Country-pair dummies	No		No		No		No		No		Yes	
Obs	6,765	6,070	6,765	6,070	6,272	5,707	6,765	6,070	6,765	6,070	6,765	6,070
$\bar{R}^2$	0.97	0.97	0.97	0.97	0.97	0.97	0.91	0.93	0.94	0.94	0.93	0.95

Note: Robust standard errors in parentheses. Dependent variable: log migrant stocks. Unit of observation: origin-destination-time. Regressions include the specified fixed effects, as indicated.