International Migration:
A Panel Data Analysis of Economic and Non-Economic Determinants*

Anna Maria Mayda†

March 2004

Preliminary and Incomplete
Comments Welcome

Abstract

In this paper I empirically investigate economic and non-economic determinants of migration inflows into fourteen OECD countries by country of origin, between 1980 and 1996. I use an annual panel data set, which allows me to exploit both the time-series and cross-country variation in immigrant inflows, and find results broadly consistent with the theoretical predictions of an international-migration model. In particular, I find evidence of robust and significant pull effects, that is improvements in the income opportunities in the host country, and of the negative impact on emigration rates of distance between destination and origin country.

JEL classification: F22.

Keywords: International Migration, Push and Pull Factors, Network Effects.

1 Introduction

Do flows of international migrants respond to economic incentives? Which non-economic determinants, such as political, cultural, and geographical factors, shape cross-country im-

*I would like to thank Alberto Alesina, Elhanan Helpman, and Dani Rodrik for support and many insightful comments. For helpful suggestions, I am also grateful to Marcos Chamon, Bryan Graham, Louise Grogan, Rod Ludema, Tara Watson, Jeffrey Williamson, and participants in the International Workshop at Harvard University and at the 2003 NEUDC Conference at Yale University. I would also like to thank the Center for International Development at Harvard University for making available office space. All errors remain mine.

†Department of Economics and School of Foreign Service, Georgetown University; email: amm223@georgetown.edu.

migration patterns? Are network effects at work? In this paper, I address these questions using an annual panel data set that allows me to exploit both the time-series and cross-country variation in international immigrant flows.

International migration flows vary considerably over time, and across destination and origin countries. Appendix 2, at the end of the paper, presents summary statistics on immigrant inflows by host and source country (see also Figure 2). It provides evidence of substantial cross-country and time-series variation of international migration movements. For example, according to this data (OECD 1997), the percentage change of the total yearly immigrant inflow between 1980 and 1995 ranges from negative 42% (Japan) to positive 48% (Canada). Countries characterized by a decrease in the size of the total annual immigrant inflow in this period are Australia, France, Japan, Netherlands, and the United Kingdom. On the other hand, the number of incoming immigrants in a year increases between 1980 and 1995 in several OECD countries (Belgium, Canada, Denmark, Germany, Luxembourg, Norway, Sweden, Switzerland, and the United States). In all destinations, such changes are anything but monotone. The variation in terms of origin countries is remarkable as well. Both economic and non-economic factors are likely to influence the size, origin, and destination of labor movements at each point in time. While it is clearly important to understand the driving forces behind recent international migration patterns, a limited amount of empirical research has been devoted to this topic, perhaps due to past unavailability of cross-country data.

In this paper, I empirically investigate economic and non-economic determinants of bilateral immigration flows, across destination and origin countries. I first derive testable predictions about the main factors affecting international migration, using a simple theoretical framework. I next relate bilateral immigration flows across destination and origin countries (normalized by origin country's population) to the economic, geographical and historical determinants suggested by the theory. The main explanatory variables I identify are income opportunities in both source and destination country, the distance between the two countries, their colonial links, the immigration-policy legislation in the host country, and a dummy variable for whether the two countries share a common language. Past works show the importance of network effects: since immigrants are likely to receive support from compatriots already established in the host country, they will have an incentive to choose destinations with larger communities of fellow citizens (see, for example, Clark, Hatton and Williamson 2002). Network effects imply that immigration to a specific destination from the same origin country tends to be highly correlated over time.

To analyze migration patterns across countries, I use yearly data on immigration inflows into fourteen OECD countries by country of origin, between 1980 and 1996. The source of this data is the International Migration Statistics for OECD countries (OECD 1997), based on the OECD's Continuous Reporting System on Migration (SOPEMI).[1]

[1] In future work, I will test the robustness of the results based on the OECD (1997) data using statistics on immigrant stocks collected by Eurostat in the EU Labour Force Survey, which covers a larger number of

I find that pull factors, that is improvements in the income opportunities in the destination country, significantly increase the size of emigration rates. This result is very robust to changes in the specification of the empirical model.

Positive and significant pull effects may appear, at first sight, to be inconsistent with restrictive immigration policies of several destination countries in the sample. From a theoretical point of view, the impact of pull (and push) factors depends on whether immigration is quantity-constrained. If immigration quotas are binding in the host country, pull (and push) factors should have no effect. However, my results show that pull effects matter, notwithstanding destination countries' official immigration restrictions. One interpretation of this finding is that the estimated coefficient simply captures an average effect, across country pairs characterized by different immigration-policy arrangements: this average effect should, according to the theory, be positive as long as immigration constraints are not binding in some destinations. Another explanation of my results is that even countries with binding official immigration quotas often accept unwanted immigration. Restrictive immigration policies are often characterized by loopholes, that leave room for potential migrants to take advantage of economic incentives. For example, immigration to Western European countries still took place after the late Seventies, in spite of the official closed-door policy (Joppke 1998). Family-reunification policies are thought to be one of the reasons of these continuing migration flows.[2]

The sign of the impact of push factors - declining levels of per worker GDP in the origin country - is consistent with the theoretical predictions, but the size of the effect is smaller than for pull factors and becomes at times insignificant. This is surprising given that, in the basic model, push and pull factors have similar-sized effects (with opposite signs). An explanation of my result is that the effect of income opportunities at home is likely to be affected by poverty constraints in the origin country, due to fixed costs of migrating and credit-market imperfections. Lower levels of per worker GDP in the source country both strengthen incentives to leave and make it more difficult to overcome poverty constraints (Yang 2003).

Among the variables affecting the costs of migration, distance between destination and origin country appears to be one of the most important ones. Its effect is negative, significant and quite steady across specifications. Finally, I empirically investigate the importance of network effects and find that their impact on the size of emigration rates is strong, positive and significant.

The empirical literature on the determinants of migration includes a number of works, some of which date back to the nineteenth century (Ravenstein 1885). More recently, Clark, Hatton and Williamson (2002) and Karemera, Oguledo and Davis (2000) both focus on the

receiving countries (Angrist and Kugler (2001) use the same type of data).

[2] Joppke (1998) writes about Germany's experience (p.285): "Since the recruitment stop of 1973, the chain migration of families of guest workers was (next to aylum) one of the two major avenues of continuing migration flows to Germany, in patent contradiction to the official no-immigration policy."

fundamentals explaining immigrant inflows into the United States by country of origin, in the last decades. Helliwell (1997 and 1998) sheds light on factors affecting labor movements in his investigation of the magnitude of immigration border effects, using data on Canadian interprovincial, US interstate and US-Canada cross-border immigration.

The contribution of this paper to the literature is threefold. First, my work is the first one I am aware of to use the OECD (1997) data on international migration to systematically investigate the economic and non-economic determinants of international flows of migrants. Previous works have either used country cross-sections (see, for example, Borjas 1987 and Yang 1995), or have focused on a single destination country (see, for example, Borjas and Bratsberg 1996, Clark, Hatton and Williamson 2002, and Karemera, Oguledo and Davis 2000) or a single origin country (see, for example, Yang 2003). By extending the focus of the analysis to a multitude of origin and destination countries and taking advantage of both the time-series and cross-country variation in the data, I can test the robustness and broader validity of the results found in the previous literature.

Second, this paper carefully reviews and proposes solutions to various econometric issues that arise in the empirical analysis, such as endogeneity and reverse causality.

Finally, the framework used in this work to study migration flows is reminiscent of a literature that analyzes bilateral *trade* flows across countries, the gravity-model literature of trade.[3] As a matter of fact, I use several variables that appear frequently in this type of works (*distance*, *common language*, and *colony*). There exists a gravity model of immigration, developed in the geography literature and sometimes used in economics papers. However, the empirical specification I use, suggested by economic theory, differs in part from the standard equation estimated by geographers.[4] By shedding light on the economic and non-economic determinants of international migration, this paper contributes to bridging the gap between economic and gravity explanations of immigrant flows.[5]

The investigation of the determinants of international migration leads to other interesting research questions. This analysis provides a framework within which it is possible to

[3] A number of works empirically analyzes trade flows within this setting (see, for example, Helpman (1987) and Hummels and Levinsohn (1995)). The same type of framework is used to explain bilateral cross-border equity flows across countries (see Portes and Rey (2002)) as well as FDI flows (see Brenton *et al.* (1999), Frankel and Wei (1996), and Mody, Razin and Sadka (2002)).

[4] The standard equation estimated by geographers looks as follows (Gallup (1997)): $flow_{ij} \propto \frac{P_i P_j}{dist_{ij}^2}$. Quoting from Gallup (1997): "H.C.Carey (1859-59) asserted that migration followed the laws of Newtonian physics: 'Man, the molecule of society, is the subject of Social Science....The great law of Molecular Gravitation [is] the indispensable condition of the existence of the being known as man....The greater the number collected in a given space, the greater is the attractive force that is there exerted....Gravitation is here, as everywhere, in the direct ratio of the mass, and the inverse one of distance."'

[5] As Helliwell (1997, p.79) points out, there is still a contrast between economic and gravity explanations of immigrant flows: "In the case of trade, the empirical success is now more widely accepted, because almost all trade theories take a gravity form under a wide range of conditions. In migration studies, there have been fewer attempts to ground the gravity form in explicit theories of migration, and to some extent there is still seen to be a contrast between "gravity" and "economic" models of migration."

address policy-related issues, as it has been done in the trade gravity-model literature. In addition, any study of the impact of labour movements on source and host economies - on their standards of living, for example - has to deal with the intrinsic problems of endogeneity of migration flows and reverse causality. Since this work helps isolate the exogenous determinants of immigrant flows, it is possible to use it to construct a *first stage* for this type of analyses (see, for example, Frankel and Romer 1999).

The rest of the paper is organized as follows. Section 2 presents a simple model of international migration. In Section 3 I describe the data sets used in the regression work, while in Section 4 I discuss the empirical model and some econometric issues that complicate the analysis. To conclude, Section 5 presents the main empirical results.

2 Theoretical framework

The size of immigration flows depends on both demand and supply factors. Migrants' decisions to move, according to economic and non-economic incentives, shape the supply side of labour movements. The host country's immigration policy represents the demand side, i.e. the demand for immigrants in the destination country. The latter one, in turn, can be thought of as the outcome of a political-economy model in which individual attitudes toward immigrants, policy-makers preferences and the institutional structure of government interact with each other and give rise to a final immigration-policy outcome (Mayda 2003 and Rodrik 1995).

I will first focus on the supply side of immigration, that is migrants' decision to move. I will consider a world with two economies: country 0, which is the country of origin of immigrant flows and country 1, which is the country of destination. I will look at the probability that an individual chosen randomly from the population of country 0 (in terms of skill) will migrate to country 1.

In each country, wages are a function of the individual skill level (s_i). In the origin country:

$$w_{0i} = \alpha_0 + \theta_0 \cdot s_i + v_{0i} = \mu_0(s_i) + v_{0i}, \text{ where } v_{0i} \sim N(0, \sigma_0^2), \tag{1}$$

while in the country of destination:

$$w_{1i} = \alpha_1 + \theta_1 \cdot s_i + v_{1i} = \mu_1(s_i) + v_{1i}, \text{ where } v_{1i} \sim N(0, \sigma_1^2), \tag{2}$$

with the correlation coefficient between v_{0i} and v_{1i} equal to ρ_{01}.

Let's assume that each individual has a CRRA utility over Cobb-Douglas-like preferences for the two goods produced in the world (x_1 and x_2):

$$U(x_1, x_2) = \frac{A[x_1^{1-\delta} x_2^{\delta}]^{1-\gamma}}{1-\gamma}, \, 0 < \delta < 1, \, 0 < \gamma < 1, \, A > 0, \tag{3}$$

which implies an indirect utility (function) from having an income y given by:[6]

$$v(p_1, p_2; y) = \overline{A}(p_1, p_2) \cdot \frac{y^{1-\gamma}}{1-\gamma}. \tag{4}$$

I assume that each country is a small open economy characterized by free trade with the rest of the world: therefore goods' prices p_1 and p_2 are given and equal - and $\overline{A}(p_1, p_2)$ also does not vary - across countries.[7] Let's restrict our attention to the case of risk neutrality ($\gamma = 0$).[8] An individual in country 0 will migrate to country 1 if the utility of moving is greater than the utility of staying at home i.e., given the assumptions above, if the expected income in the destination country net of migration costs is greater than the expected income in the origin country. Following the literature (see, for example, Borjas 1999a, and Clark, Hatton and Williamson 2002), I can define an index I that measures the net benefit of moving relative to staying at home for a risk-neutral individual:

$$I = \eta_{01} \cdot (w_{1i} - w_{0i} - C) + (1 - \eta_{01}) \cdot (-w_{0i} - C), \tag{5}$$

$$\Longrightarrow I = \eta_{01} \cdot w_{1i} - w_{0i} - C, \tag{6}$$

where η_{01} is the probability that the migrant from country 0 will be allowed to stay in country 1, w_{0i} and w_{1i} are respectively the wage in the origin and destination country, and $C = \mu_C + v_i^C$, with $v_i^C \sim N(0, \sigma_C^2)$, represents the level of migration costs.[9] The correlation coefficients between v_i^C and $(v_{0i}, \, v_{1i})$ are equal to $(\rho_{0C}, \, \rho_{1C})$.

This model focuses on *labor* mobility. Migration allows an individual to take advantage of differences in rates of return to *labor* across countries. Migrants may own capital, either at home or in the destination country, and their capital income opportunities are independent of their residence.[10] In addition, the implicit assumption in (5) is that, if the migrant is not allowed into the destination country, he still incurs the migration costs C and gives up the wage at home w_{0i}. In other words, the individual moves to the host country before knowing whether he will be able to stay (for a longer period of time) and gain the income w_{1i}. The

[6] In the following expression: $\overline{A}(p_1, p_2) = A[(\frac{1-\delta}{p_1})^{1-\delta}(\frac{\delta}{p_2})^{\delta}]^{1-\gamma}$.

[7] In the empirical section of the paper I adjust for international differences in goods' prices, by considering PPP-adjusted income levels.

[8] In future work, I would like to examine the case of risk aversion.

[9] I assume that each individual knows the wage levels w_{1i} and w_{0i} he would get in each location and the migration costs C.

[10] In other words, capital is internationally mobile. The migrant can own capital in the origin and destination country and receive income from it, no matter where he resides.

immigrant from country 0 may not be allowed into country 1 because of quotas due to a restrictive immigration policy, as is explained below. Notice that, while each individual takes the probability of being allowed into the destination country (η_{01}) as given, this probability is endogenously determined in the model, as a function of the host country's immigration policy.[11]

We can think of the level of migration costs C as being an increasing function of physical distance between the origin and destination country, since remote destinations imply higher monetary and time travel costs; a decreasing function of linguistic and cultural similarities like, for example, a common language and past colonial ties; and a decreasing function of past migration inflows from the same origin country, which capture network effects.

An individual chosen randomly from the population of country 0 has skill equal on average to $\overline{s}_0$, the average skill level in the population of the origin country. The wage in the origin country of this representative individual is therefore given by $\alpha_0 + \theta_0 \cdot \overline{s}_0 + v_{0i} = \mu_0 + v_{0i}$; in the destination country, that same individual is expected to earn a wage equal to $\alpha_1 + \theta_1 \cdot \overline{s}_0 + v_{1i} = \mu_1^0 + v_{1i}$. Notice that the latter expression is likely to be different from the wage in country 1 of a representative individual (in terms of skill) from that country's population: $\alpha_1 + \theta_1 \cdot \overline{s}_1 + v_{1i} = \mu_1 + v_{1i}$, where $\overline{s}_1$ represents the average skill level in the population of the destination country (Borjas 1999a, and Clark, Hatton and Williamson 2002). The probability that a representative individual (in terms of skill) of the origin country will migrate from country 0 to country 1 equals:

$$P = \Pr[I > 0] = \Pr[\eta_{01} \cdot (\mu_1^0 + v_{1i}) - (\mu_0 + v_{0i}) - (\mu_C + v_i^C) > 0], \tag{7}$$

which can be rewritten as:

$$P = \Pr[\eta_{01} \cdot v_{1i} - v_{0i} - v_i^C > -(\eta_{01} \cdot \mu_1^0 - \mu_0 - \mu_C)],$$

$$\Longrightarrow P = \Pr[\frac{\eta_{01} \cdot v_{1i} - v_{0i} - v_i^C}{\sigma_v} > -\frac{(\eta_{01} \cdot \mu_1^0 - \mu_0 - \mu_C)}{\sigma_v}]$$

$$\Longrightarrow P = 1 - \Phi(z), \tag{8}$$

where σ_v is the standard deviation of $(\eta_{01} \cdot v_{1i} - v_{0i} - v_i^C)$, $z = -\frac{(\eta_{01} \cdot \mu_1^0 - \mu_0 - \mu_C)}{\sigma_v}$ and $\Phi(\cdot)$ is the cumulative distribution function of a standard normal.[12]

An additional layer of uncertainty can be introduced in the model by considering in (5) and (6) the *expected* wage, both in the origin and destination country, with respect to the probability of finding a job in each place (this probability can be approximated with one

[11] My model differs from previous ones in the literature in the way it analyzes the impact of quantity restrictions induced by immigration policy. Clark, Hatton, and Williamson (2002) and Hatton and Williamson (2002) model immigration policy as affecting the level C of migration costs.

[12] In particular, $\sigma_v^2 = (\eta_{01}^2 \sigma_1^2 + \sigma_0^2 + \sigma_C^2 - 2\eta_{01}\rho_{01}\sigma_0\sigma_1 - 2\eta_{01}\rho_{1C}\sigma_1\sigma_C + 2\rho_{0C}\sigma_0\sigma_C)$.

minus the unemployment rate). The model can also be extended to a multi-period setting. In this set-up, the individual cares not only about current wage differentials, but also about future ones, which in turn depend on growth rates of wages at home and abroad.[13]

Consider a situation in which the destination country's immigration policy implies either explicit or implicit quantity constraints for immigrants coming from each origin country. Let I_{01}^D represent the maximum number of migrants from country 0 allowed each period into country 1. These immigration quotas may or may not be binding.

Given the OECD (1997) data, we can observe the actual emigration rate $\frac{I_{01}}{P_0}$, i.e. the number of immigrants coming into country 1 from country 0, divided by the population of country 0. The probability of emigration from country 0 to country 1 in (8) can be thought of as approximately equal to the *supply* emigration rate $\frac{I_{01}^S}{P_0}$, which in the absence of binding immigration quotas equals the ex-post emigration rate. On the other hand, the ex-post emigration rate that arises in the presence of binding quantity-constraints will be less than $\frac{I_{01}^S}{P_0}$. The ex-post emigration rate is thus equal to the minimum between $\frac{I_{01}^S}{P_0}$ and $\frac{I_{01}^D}{P_0}$:

$$\frac{I_{01}}{P_0} = \min(\frac{I_{01}^S}{P_0}, \frac{I_{01}^D}{P_0}), \tag{9}$$

where the immigration quota I_{01}^D represents the demand in country 1 for immigrants from country 0, which is a function of the destination country's immigration policy. The heavy lines in Figures 1 and 2 give the ex-post emigration rate as a function of μ_1^0 and μ_h, $h = 0$, C. In this paper I assume that I_{01}^D is exogenous, thus it is not affected by μ_1^0 neither by μ_h, $h = 0, C$.[14]

Given (8) and (9), it is possible to derive testable predictions for the impact of μ_1^0, μ_0, and μ_C on the ex-post emigration rate from country 0 to country 1:[15]

$$\frac{\partial(\frac{I_{01}}{P_0})}{\partial\mu_1^0} = \eta_{01} \cdot \frac{\phi(z)}{\sigma_v} > 0, \text{ if } \frac{I_{01}^S}{P_0} < \frac{I_{01}^D}{P_0}; \tag{10}$$

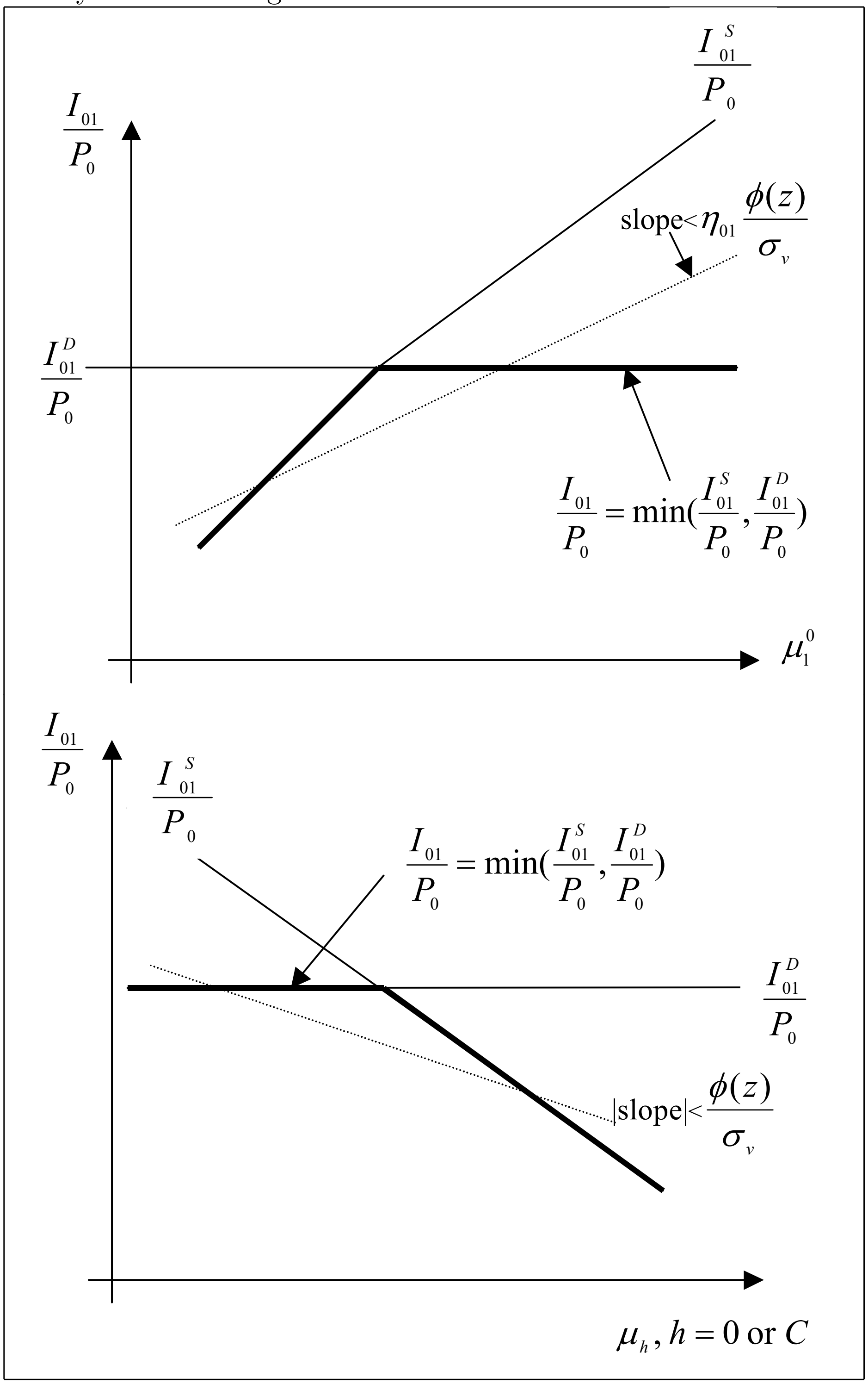
$$\frac{\Delta(\frac{I_{01}}{P_0})}{\Delta\mu_1^0} \in (0, \eta_{01} \cdot \frac{\phi(z)}{\sigma_v}), \text{ if } \frac{I_{01}^S}{P_0} < \frac{I_{01}^D}{P_0} \text{ ex-ante and } \frac{I_{01}^S}{P_0} > \frac{I_{01}^D}{P_0} \text{ ex-post, or viceversa;} \tag{11}$$

[13] In future work, I would also like to incorporate poverty constraints in the model, linked to imperfections in the credit market. Poverty constraints complicate the comparative-static result with respect to μ_0.

[14] Alternatively, I_{01}^D can be explicitly modeled within a political-economy framework. In that case, the immigration quotas are likely to depend on the capital-labor ratio of the median voter (see Benhabib 1996), on the size of past immigration flows from the same origin country (both because of family-reunification policies and because of pro-immigration votes of naturalized immigrants), and on the extent of political organization of various interest groups (Grossman and Helpman 1994 and Facchini 2004).

[15] An additional comparative-static exercise is with respect to σ_v and its single components (σ_1^2, σ_0^2, σ_C^2, ρ_{01}, ρ_{0C}, and ρ_{1C}). This type of analysis will be the focus of future work.

Figure 1: The ex-post emigration rate as a function of income opportunities in the destination and origin country and of moving costs



$$\frac{\partial(\frac{I_{01}}{P_0})}{\partial\mu_1^0} = 0, \text{ if } \frac{I_{01}^S}{P_0} \geq \frac{I_{01}^D}{P_0}, \tag{12}$$

where $\phi(\cdot)$ is the density function of a standard normal. In analogous way:

$$\frac{\partial(\frac{I_{01}}{P_0})}{\partial\mu_h} = -\frac{\phi(z)}{\sigma_v} < 0, \text{ if } \frac{I_{01}^S}{P_0} \leq \frac{I_{01}^D}{P_0}; \tag{13}$$

$$\frac{\Delta(\frac{I_{01}}{P_0})}{\Delta\mu_h} \in (-\frac{\phi(z)}{\sigma_v}, 0), \text{ if } \frac{I_{01}^S}{P_0} > \frac{I_{01}^D}{P_0} \text{ ex-ante and } \frac{I_{01}^S}{P_0} < \frac{I_{01}^D}{P_0} \text{ ex-post, or viceversa;} \tag{14}$$

$$\frac{\partial(\frac{I_{01}}{P_0})}{\partial\mu_h} = 0, \text{ if } \frac{I_{01}^S}{P_0} > \frac{I_{01}^D}{P_0}, \tag{15}$$

where $h = 0, C$. The comparative-static results in (10)-(12) show the effect of pull factors - that is, improvements in the income opportunities in the destination country - according to whether the immigration quotas are binding or not. Pull effects are positive and strongest when restrictions are not binding neither ex-ante nor ex-post (10), they are positive but smaller in size when the quota is binding ex-post but not ex-ante (11) and, finally, they are equal to zero in a quantity-constrained world (12). A parallel interpretation explains the comparative-static results in (13)-(15), which describe push effects (changes of μ_0) and the impact of average migration costs (changes of μ_C), according to the immigration-policy regime.

We can assume that the probability η_{01} equals 1 when $I_{01}^D \geq I_{01}^S$ and is smaller than 1 and an increasing function of I_{01}^D when $I_{01}^D < I_{01}^S$.[16] (If the quantity constraints are binding - $I_{01}^D < I_{01}^S$ - the higher the immigration quota in country 1 for immigrants from country 0, the higher the probability that a migrant will be allowed into the country.[17]) Therefore, the restrictiveness of the destination country's immigration policy affects both the demand and the supply emigration rates but it has an effect on the ex-post emigration rate only through the demand channel.

[16] Therefore $\eta_{01} = 1$ in (10) and (11) and $\eta_{01} < 1$ if $I_{01}^S > I_{01}^D$.

[17] We can fully endogenize η_{01}, which is equal to $\min\{1, \frac{I_{01}^D}{P_0 \cdot P}\}$ (the number of people, from country 0 to country 1, who are allowed in, divided by the number of those who try to get in). Fully endogenizing η_{01} makes $\partial(\frac{I_{01}^S}{P_0})/\partial\mu_1^0$ smaller in the portion of the supply emigration-rate curve which is not observed: $\frac{\partial(I_{01}^S/P_0)}{\partial\mu_1^0} = \frac{\phi(z)\eta_{10}}{\sigma_v}\frac{1}{(1+\frac{\mu_1^0\eta_{10}}{P}\frac{\phi(z)}{\sigma_v})} < \frac{\phi(z)\eta_{10}}{\sigma_v}$.

3 Data

In this paper, I combine an international panel on bilateral immigration flows with external macroeconomic and non-economic data on the origin and destination country of each flow. Data on immigration comes from the International Migration Statistics (IMS) data set for OECD countries (OECD (1997)), which contains information on immigrant flows by country of origin, based on the OECD's Continuous Reporting System on Migration (SOPEMI). Population registers and residence and work permits are the main sources of these statistics.[18] In particular, I use data on yearly immigrant inflows into fourteen OECD countries by country of origin, in the period 1980-1996 (see Appendix 2 for summary statistics).[19]

Appendix 2 shows that the IMS statistics on immigrant flows by country of origin don't cover 100% of the total flow into each destination. The percentage of the total average immigrant inflow, between 1980 and 1995, covered by the data by origin country goes from 69% (France) to 95% (Germany). Put differently, the data set has missing observations in correspondence of some country pairs (immigrant inflows from Italy to the United States, for example). These observations could be missing because they correspond to zero flows, or to small flows (and thus they are not recorded), or because of some other selection mechanism. In future work I would like to use either a Tobit model or a censored regression model or a selection model to deal with missing observations and test the robustness of my results.[20]

Data on macroeconomic variables comes from various sources: the 2001 World Development Indicators (World Bank (2001)), the Penn World Tables (versions 5.6 and 6.1), and the World Bank's Global Development Network Growth Database, Macro Time Series (Easterly and Sewadeh (2002)). Geographical, cultural, and historical information, such as on great-circle distance, common language, and colonial ties, come from Glick and Rose's (2001) data set on gravity-model variables. Data sources of each variable used in the empirical model are documented in Appendix 1.

I use statistics on the average number of schooling years in the total population (over age 15) from Barro and Lee's (2000) data set. Since this panel only contains data at five-year intervals (in the period I consider, the years covered are 1980, 1985, 1990, 1995), I linearly extrapolate figures for the in-between years (by assigning one fifth of the five-year change in the variable to each year).

[18] The IMS data set also includes statistics on the origin and labor market characteristics of immigrant *stocks*, based on survey and census data from Eurostat and national governments and on population registers.

[19] Good statistics on immigration are hard to find, especially for developing countries. OECD and Eurostat figures (see footnote 1) concentrate on high and middle-income economies as receiving countries of immigrant flows. In 1998 the International Labor Organization (ILO) mailed a questionnaire survey to member states to obtain basic data on stocks and flows of migrant labor worldwide. Responses to this questionnaire form the basis of the International Labor Migration Database (ILO (1998)). At this stage this data set cannot be used, due to the low degree of harmonization of data from different countries.

[20] Note that the IMS data does not include illegal immigration.

4 Empirical model and some econometric issues

The empirical specification suggested by the comparative-static analysis in (10)-(15) is characterized by the observed emigration rate as the dependent variable and, among the explanatory variables, the average wage earned by the representative individual from country 0 in, respectively, the origin and destination country. As approximations for the latter two variables, I use the (log) level of per worker GDP, PPP-adjusted (constant 1996 international dollars) in the two countries.[21] Another determinant of bilateral immigration flows implied by the model of Section 2 is the distance between the two locations. The further away the two countries are, the higher the monetary travel costs are likely to be for the initial move, as well as for visits back home. Remote destinations may also discourage migration because they require longer travel time and thus higher foregone earnings. Another explanation as to why distance may negatively affect migration is that it is more costly to acquire information ex-ante about far-away countries (Greenwood (1997) and (Lucas (2000)). Linguistic and cultural similarity are also likely to reduce the magnitude of migration costs, for example by improving the transferability of individual skill from one place to the other. Past colonial relationships should increase emigration rates, to the extent that they translate into similar institutions and stronger political ties between the two countries, thus decreasing the level of migration costs.

In a cross-country analysis, such as in this paper, unobserved country-specific effects may result in biased estimates. For example, I may estimate a positive coefficient on the destination country's wage. It is not clear whether this means that immigrants are more likely to go to a country the higher its wage or, alternatively, that a country with higher wages has other features that attract immigrants. Along the same lines, a negative coefficient on income at home leaves open the question of whether immigrants leave countries with lower wages or, alternatively, whether countries with lower wages have certain characteristics that push immigrants to leave. To (partly) get around this problem, I exploit the panel structure of the data set and I introduce dummy variables for both destination and origin countries. This allows me to control for unobserved country-specific effects which are additive and time-invariant. My preferred specification (column (5), Table (1)) has countries' fixed effects and robust standard errors clustered by country pair, to address heteroscedasticity and allow for correlation over time of country-pair observations. Notice that (destination) country fixed effects allow me to control for features of destination countries' immigration policy which don't change over time and are common across origin countries.[22]

The empirical specification thus looks as follows:

$$\frac{flow_{ijt}}{P_{it}} = const. + \beta_0 pwgdp_{it-1} + \beta_1 pwgdp_{jt-1} + \beta_2 dist_{ij} + \beta_3 comlang_{ij} + \beta_4 colony_{ij} + I_i + I_j + \varepsilon_{ijt}$$

[21] Data on per worker GDP, PPP-adjusted (constant 1996 international dollars) comes from the Penn World Tables (version 6.1).

[22] In future work, I would like to introduce indicator variables for *changes* in each destination country's immigration policy.

where i is the origin country, j the destination country and t time. $\frac{flow_{ijt}}{P_{it}}$ is the emigration rate from i to j at time t ($flow_{ijt}$ is the inflow into country j from country i at time t, P_{it} is the population of the origin country at time t). *pwgdp* is the (log) per worker GDP, PPP-adjusted (constant 1996 international dollars) and *dist* measures the (log) great-circle distance between the two countries. *comlang* and *colony* are two dummy variables equal to one, respectively, if a common language is spoken in both locations, and for pairs of countries which were, at some point in the past, in a colonial relationship. I_i and I_j are vectors of dummy variables for, respectively, the origin and the destination countries. According to the theory, I expect that $\beta_0 < 0$, $\beta_1 > 0$, $\beta_2 < 0$, $\beta_3 > 0$, and $\beta_4 > 0$.[23] Note that, as a first approximation, this empirical specification only focuses on average effects across immigration-policy regimes. In other words, it does not differentiate according to whether immigration restrictions are binding or not.[24]

Granted that per worker GDP proxies for the income opportunity of the migrant worker in each location (see below for a discussion of this point), an empirical complication is the possibility of reverse causality and, more in general, of endogeneity in the time-series dimension of the analysis. The theoretical model in Section 2 predicts that, *ceteris paribus*, higher (lower) income opportunities in the destination (origin) country increase emigration rates. However, a positive β_1 (negative β_0) may just reflect causation in the opposite direction, i.e. the impact of immigrant flows on wages (or levels of per worker GDP) in the host and source country. After all, this channel is the focus of analysis in most labour-economics papers (see Friedberg and Hunt 1995 for a survey of this literature). More broadly, other time-variant third factors may drive contemporaneous wages and immigrant flows.

As for reverse causality, notice that the bias introduced by it is likely to work against me, in the sense that it is expected to bias the estimates toward zero. The reason is that immigrant inflows are likely to decrease wages in the destination country and outflows are likely to increase wages in the origin country. While the opposite signs are a theoretical possibility (for example, in the economic-geography literature, because of economies of scale), the empirical evidence in the labor-economics literature is that immigrant inflows have a negative impact on the destination country's wages (Borjas 2003) and that immigrant outflows have a positive impact on the origin country's wages (see Mishra 2003).

[23] The empirical model can be extended by introducing additional cultural, historical, and geographical variables that are likely to have an impact on the cost C of migration (for example, measures of similarity between the two countries in terms of religious affiliation, or a common-border dummy variable).

[24] Some preliminary evidence that immigration policy affects emigration rates in the manner predicted by the model is as follows. Family-reunification policies are a very important component of the immigration policies of many destination countries in the sample. Thus, I can assume that immigration quotas are an increasing function of the immigrant inflow in the previous period, from the same origin country. The higher this flow, the less binding quotas are supposed to be (through family reunification), the more likely it is that we are in a region where the wage in the destination country has a positive (rather than zero) effect on the emigration rate. When I interact the lagged flow with the destination country's per worker GDP, I find a positive and significant coefficient.

I address reverse-causality and endogeneity issues in two ways. First of all, in the basic specification, I relate *current* emigration rates to *lagged* values of (log) per worker GDP, at home and abroad. Indeed, while it is hard to claim that average wages at home and abroad are strictly exogenous, it is plausible to assume that they are predetermined, in the sense that immigrant inflows - and third factors in the error term - only affect contemporaneous and future wages.[25]

I next use instrumental-variables estimation with countries' terms of trade as an instrument for PPP-adjusted income levels in the destination and origin country. Papers in the literature where shocks to terms of trade are used as instruments for growth rates of income are, for example, Pritchett and Summers (1996) and Easterly, Kremer, Pritchett and Summers (1993). Notice that the validity of this instrument depends on the assumption that countries are small open economies.

As pointed out above, to capture the effect of income opportunities at home and abroad, I use data on GDP per worker (PPP-adjusted) in the origin and destination country. In other words, I do not measure average wages in the two locations directly. An important issue is, therefore, whether per worker GDP is indeed proxying for the average wage. I next test the robustness of my results in this respect.

Since measures of GDP include payments to both labour and capital, I can better isolate the wage component by adjusting for differences in the level of per-worker capital ownership in each country.[26] Notice that, after isolating the wage component, a higher average wage in the destination country (μ_j) does not necessarily mean better income opportunities for the representative individual of country i (μ_j^i). As pointed out in Section 2, $\mu_j^i = \alpha_j + \theta_j \cdot \overline{s}_i$ while $\mu_j = \alpha_j + \theta_j \cdot \overline{s}_j$. I can use information on the average wage in the destination country (μ_j), together with data on the average skill level in the origin and destination countries ($\overline{s}_i$ and $\overline{s}_j$), to measure (the effect of) the average wage in country j of a representative individual of country i (in terms of skill): $\mu_j^i = \mu_j - \theta_j(\overline{s}_j - \overline{s}_i)$. In other words, controlling for the average skill level in the origin and destination countries, the comparative statics with respect to μ_j^i and μ_j are equivalent to each other (Hatton and Williamson 2003).

Past migration flows to the destination country, from the same origin country, affect the current emigration rate through both the supply and the demand channel. On the supply side, lagged emigration rates or, alternatively, the size of the immigrant stock from the same source country, proxy for network effects, which are likely to reduce the cost C of migration. On the demand side, past migration flows influence the emigration rate in two ways: through family-reunification immigration policies and through political-economy factors (see, for example, Goldin (1994), where the votes of naturalized immigrants affect immigration policy outcomes).

[25] Strict exogeneity of an explanatory variable implies $E[X_{it}\varepsilon_{is}] = 0$, for $\forall s$, t, while predeterminacy implies $E[X_{it}\varepsilon_{is}] = 0$, for $\forall s > t$.

[26] International differentials in rates of return to capital also matter but, as a first approximation, I will assume that capital is internationally mobile.

The introduction of the lagged emigration rate among the explanatory variables makes the model a dynamic one. A complication in the empirical analysis of a dynamic equation is the *incidental parameter problem*.[27] In a dynamic equation, the fixed effects (or within) estimator of the coefficient of the lagged dependent variable is consistent as $T \rightarrow \infty$, for given N, but it is not consistent for given T, as $N \rightarrow \infty$. The intuition behind this result is that, in the latter case, the number of parameters to be estimated tends to infinity, while the information used to estimate each parameter does not increase. An econometric technique used to deal with this problem is Arellano and Bond's GMM estimator. I use this estimation technique to test the robustness of my estimates, once I introduce the lagged emigration rate(s) among the explanatory variables.

5 Empirical results

Table 1, at the end of the paper, presents the results from estimation of the model exploiting both the cross-country and time-series variation. After specifying the model with a unique intercept (regression (1)), I introduce the two sets of country dummy variables sequentially. I first control for the destination countries' unobserved fixed effects (column (2)), I next add to them origin countries' dummy variables (regression (3)). In column (4) I only exploit the variation over time within country pairs, by introducing dummy variables for each combination of origin and destination countries.[28] These country-pairs fixed effects allow me to control for time-invariant features of the destination country's immigration policy which are specific for each origin country. Finally, in the last regression of the table, I go back to the specification of column (3) and I cluster standard errors by country pair, to deal with heteroscedasticity and allow for correlation over time of observations corresponding to the same combination of source and host countries.

The estimates of Table 1 show a systematic pattern, broadly consistent with the theoretical predictions of the model. The emigration rate is positively related to the destination country's (log) per worker GDP and negatively associated with the origin country's (log) per worker GDP, as predicted in Section 2. According to the estimates in regression (5), a ten percent increase in the level of GDP per worker in the destination country increases emigration by 0.1 per thousand individuals of the origin country's population (the mean of the dependent variable is, in that regression, 0.586 emigrants per thousand). In other words, a 10% increase in the host country's GDP implies a 19% increase in the emigration rate. The impact on the emigration rate of a change in the income opportunities at home is smaller in absolute value: a ten percent decrease in the level of GDP per worker of the origin

[27] In a model estimated using a panel data set (T observations for each unit $a = 1, ..., N$), the parameters specific for each unit a are called "incidental" parameters. These parameters are usually estimated introducing dummy variables, that is using a fixed-effect specification, as in my model.

[28] Regression (4) does not include the regressors (*log*) *distance*, *common language* and *colony* since they are constant within country pairs and, therefore, they would be perfectly collinear with the dummy variables.

country increases emigration by 0.02 per thousand individuals in the origin country. The interpretation of this result is that it is probably driven by the effect of poverty constraints in the origin country. A lower level of GDP per worker in the source country strengthens the incentive to migrate, but it also makes it more likely that a bigger portion of the population will be unable to move, if fixed costs are required to migrate and there are credit-market imperfections. Notice that the size of both coefficients is especially affected by the introduction of host country's fixed effects which capture, among other factors, the impact of time-invariant features of the immigration policy at destination.

According to the estimate in column (5), doubling the great-circle distance between the source and host country decreases the number of emigrants by 0.4 per thousand individuals in the origin country (significant at the 1% level). The impact of a common language, though of the right sign, decreases in size and loses significance once I control for origin countries' fixed effects. Surprisingly, past colonial relationships appear to negatively affect migration flows (the coefficient is less precisely estimated in the last regression).

In Table 2 I estimate the coefficients exploiting only the cross-country variation. I divide the period between 1981 and 1995 into three segments and I focus on each at a time. I relate average emigration rates in each subperiod to the average income opportunities at home and abroad in the previous five-year interval. In Table 3 I perform a similar exercise by estimating the model year by year. Due to the low number of observations in each regression, in Table 2 and Table 3 I don't control for country-specific fixed effects, which explains the difference in the magnitude of the effects relative to regression (5), Table 1. The coefficients are still qualitatively consistent with the panel-data results, though less precisely estimated.

I next examine each destination country at a time, in Table 4.[29] This set of results is less clear than previous ones and requires further work.

In Table 5 I run three robustness checks of the panel-data results. In the first regression, I use (within-country deviations in) the terms of trade to instrument for (within-country deviations in) the level of per worker GDP of both destination and origin country. Terms of trade affect countries' purchasing power vis a vis goods produced by the rest of the world, thus they affect the average real income in each location (in the first stage, the impact of the terms of trade on per worker GDP is positive and significant at the 1% level, for both destination and origin country). In addition, given the assumption of small open economies, terms of trade are unlikely to affect emigration rates directly or to be correlated with other country-level characteristics that have an impact on migration patterns (exclusion restriction).

In columns (2) and (3), I investigate whether per worker GDP in the two locations is a good measure of the average income opportunity of the representative individual from country 0. I first control for the average schooling level in both countries in column (2). Pull effects are still estimated to be positive and significant (at the 1% level), while the impact

[29] These regressions control for origin countries' fixed effects and have standard errors clustered by country of origin.

of push effects is greatly reduced. In line with the theoretical predictions, the average skill level in the population of the destination (origin) country has a negative (positive) impact on the emigration rate.

In Table 6 I investigate network effects by introducing the lagged emigration rate(s) among the explanatory variables. The estimates change considerably, according to the set of country dummy variables I control for. As already pointed out, fixed-effects estimation of a dynamic model with a short panel (small T) may produce biased estimates. I thus use Arellano and Bond's estimator in regression (3) and find results consistent with the theoretical predictions of the model.[30]

6 Conclusions

In this paper, I investigate economic and non-economic determinants of international migration flows. This analysis both delivers estimates consistent with the predictions of an economic model and generates empirical puzzles.

In particular, I find that pull factors, that is improvements in the income opportunities in the destination country, significantly increase the size of emigration rates. This result, which appears to be very robust to changes in the specification of the empirical model, is surprising, given restrictive immigration policies of the destination countries considered. The sign of the impact of push factors - declining levels of per worker GDP in the origin country - is consistent with the theoretical predictions of the model, but the size of the effect is smaller than for pull factors and becomes at times insignificant. Among the variables affecting the costs of migration, distance appears to be one of the most important ones. Its effect is negative, significant and quite steady across specifications.

By taking advantage of both the time-series and cross-country variation in an annual panel data set, this paper makes progress in explaining the economic and non-economic determinants of international migration flows.

References

Adams, R. H. J. (1993). The economic and demographic determinants of international migration in rural Egypt. *Journal of Development Studies*, 30(1):146–167.

Angrist, J. D. and Kugler, A. D. (2001). Protective or counter-productive? European labor

[30] In the last model, I include the emigration rate lagged by one and by two years. The reason is that, only by introducing both lags, I don't reject the null of zero autocovariance in residuals of order 2 (which is one of the requirements of the Arellano and Bond estimator). In future work, I would like to proxy network effects with the immigrant stock from the same origin country (which is likely to pass the zero second-order autocovariance test).

market institutions and the effect of immigrants on EU natives. National Bureau of Economic Research Working Paper No. 8660.

Barro, R. and Lee, J. (2000). International data on educational attainment. Data Set.

Bauer, T. K., Lofstrom, M., and Zimmermann, K. F. (2000). Immigration policy, assimilation of immigrants and natives' sentiments towards immigrants: Evidence from 12 OECD countries. IZA Discussion Paper No. 187.

Benhabib, J. (1996). On the political economy of immigration. *European Economic Review*, 40:1737–1743.

Berry, R. A. and Soligo, R. (1969). Some welfare aspects of international migration. *Journal of Political Economy*, 77:778–794.

Borjas, G. and Bratsberg, B. (1996). Who leaves? The outmigration of the foreign-born. *Review of Economics and Statistics*, 78(1):165–176.

Borjas, G. J. (1987). Self selection and the earnings of immigrants. *American Economic Review*, 77:531–553.

Borjas, G. J. (1994). The economics of immigration. *Journal of Economic Literature*, pages 1667–1717.

Borjas, G. J. (1995). The economic benefits from immigration. *Journal of Economic Perspectives*, 9(2):3–22.

Borjas, G. J. (1999a). The economic analysis of immigration. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, chapter 28, pages 1697–1760. North-Holland Elsevier Science, The Netherlands.

Borjas, G. J. (1999b). *Heaven's Door: Immigration Policy and the American Economy.* Princeton University Press, Princeton, N.J.

Borjas, G. J. (2003). The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. Harvard University.

Chiswick, B. R. and Hatton, T. J. (2002). International migration and the integration of labour markets. In *Globalization in Historical Perspective.* The University of Chicago Press, Chicago, IL. Forthcoming.

Clark, X., Hatton, T. J., and Williamson, J. G. (2002). Where do U.S. immigrants come from, and why? National Bureau of Economic Research Working Paper No. 8998.

Coppel, J., Dumont, J.-C., and Visco, I. (2001). Trends in immigration and economic consequences. OECD Economics Department Working Papers No. 284.

Davies, J. B. and Wooton, I. (1992). Income inequality and international migration. *The Economic Journal*, 102(413):789–802.

Davis, D. R. and Weinstein, D. E. (2002). Technological superiority and the losses from migration. National Bureau of Economic Research Working Paper No. 8971.

Easterly, W., Kremer, M., Pritchett, L., and Summers, L. H. (1993). Good policy or good luck? Country growth performance and temporary shocks. *Journal of Monetary Economics*, 32:459–483.

Easterly, W. and Sewadeh, M. (2002). Global development network growth database, Macro time series. World Bank. Data Set.

Faini, R. (2001). Development, trade, and migration.

Faini, R. (2002). Discussion of "International migration and the integration of labor markets" by B. Chiswick and T. Hatton. In *Globalization in Historical Perspective.* The University of Chicago Press, Chicago, IL. Forthcoming.

Faini, R. and Venturini, A. (1994). Trade, aid and migrations. Some basic policy issues. *European Economic Review*, 37:435–442.

Frankel, J. A. and Romer, D. (1999). Does trade cause growth? *The American Economic Review*, 89:379–399.

Freeman, G. (1992). Migration policy and politics in the receiving states. *International Migration Review*, 26(4):1144–1167.

Freeman, G. (1995). Modes of immigration politics in liberal democratic states. *International Migration Review*, 29(4):881–902.

Friedberg, R. and Hunt, J. (1995). The impact of immigrants on host country wages, employment, and growth. *Journal of Economic Perspectives*, 9(2):23–44.

Glick, R. and Rose, A. K. (2001). Does a currency union affect trade? The time series evidence.

Goldin, C. (1994). The political economy of immigration restriction in the United States, 1890 to 1921. In Goldin, C. and Libecap, G., editors, *The Regulated Economy: A Historical Approach to Political Economy*, pages 223–257. University of Chicago Press, Chicago, IL.

Gould, D. M. (1994). Immigrant links to the home country: Empirical implications for U.S. bilateral trade flows. *The Review of Economics and Statistics*, 76:302–316.

Greenwood, M. J. (1997). Internal migration in developed countries. In Rosenzweig, M. and Stark, O., editors, *Handbook of Population and Family Economics, Vol. 1B*. North-Holland, Amsterdam.

Hatton, T. J. and Williamson, J. G. (2001). Demographic and economic pressure on emigration out of Africa. National Bureau of Economic Research Working Paper No. 8124.

Helliwell, J. F. (1997). National borders, trade and migration. National Bureau of Economic Research Working Paper No. 6027.

Helliwell, J. F. (1998). *How Much Do National Borders Matter?*, chapter 5, pages 79–91. Brookings Institution Press.

Heston, A., Nuxoll, D. A., Summers, R., and Aten, B. (1995). Penn World Table version 5.6a. Data Set.

Heston, A., Summers, R., and Aten, B. (2002). Penn World Table version 6.1. Center for International Comparisons at the University of Pennsylvania (CICUP). Data Set.

Hoddinott, J. (1994). A model of migration and remittances applied to Western Kenya. *Oxford Economic Papers*, 46:459–476.

ILO (1998). International labour migration database. International Labor Organization.

Jenks, R. (1992). *Immigration and Nationality Policies of Leading Migration Nations.* Center for Immigration Studies, Washington, DC.

Joppke, L. (1998). Why liberal states accept unwanted immigration. *World Politics*, 50:266–293.

Karemera, D., Oguledo, V. I., and Davis, B. (2000). A gravity model analysis of international migration to North America. *Applied Economics*, 32(13):1745–1755.

Lucas, R. E. (2000). The effects of proximity and transportation on developing country population migrations. Boston University.

Mayda, A. M. (2003). Who is against immigration? A cross-country investigation of individual attitudes toward immigrants. Harvard University, dissertation chapter.

Money, J. (1997). No vacancy: The political geography of immigration control in advanced industrial countries. *International Organization*, 51:685–720.

OECD (1997). International migration statistics for OECD countries. Data set.

OECD (2001). The employment of foreigners: Outlook and issues in OECD countries. In *OECD Employment Outlook*, chapter 5. OECD, Paris.

Pritchett, L. and Summers, L. H. (1996). Wealthier is healthier. *The Journal of Human Resources*, 31(4):841–868.

Ravenstein, E. (1885). The laws of migration. *Proceedings of the Royal Statistical Society*, XLVII(2):167–235.

Rodrik, D. (1995). Political economy of trade policy. In Grossman, G. and Rogoff, K., editors, *Handbook of International Economics, Vol.3*, chapter 28, pages 1457–1494. North-Holland Elsevier Science, The Netherlands.

SOPEMI (1997). *Trends in International Migration. Annual Report 1996.* OECD, Paris.

SOPEMI (1999). *Trends in International Migration*. OECD, Paris.

SOPEMI (2000). *Trends in International Migration*. OECD, Paris.

Trefler, D. (1997). Immigrants and natives in general equilibrium trade models. National Bureau of Economic Research Working Paper No. 6209.

Yang, D. (2003). Financing constraints, economic shocks, and international labor migration: Understanding the departure and return of philippine overseas workers. Harvard University, dissertation chapter.

Yang, P. Q. (1995). *Post-1965 Immigration to the United States*. Praeger, Westport, Connecticut.

Table 1. Panel data regressions

Equation	1	2	3	4	5
	no country fixed effects	destination countries dummy variables (d.v.)	destination and origin countries d.v.	country pair d.v.	(3) plus clustered standard errors by country pair
Dependent variable	**Emigration rate**				
log per worker gdp (destination)	0.00240	0.00100	0.00110	0.00086	0.00110
	0.00021**	0.00040*	0.00028**	0.00017**	0.00032**
log per worker gdp (origin)	-0.00021	-0.00010	-0.00015	-0.00012	-0.00015
	0.00004**	0.00004*	0.00009+	0.00006*	0.00012
log distance	-0.00017	-0.00036	-0.00035		-0.00035
	0.00004**	0.00004**	0.00004**		0.00010**
common language	0.00068	0.00073	0.00012		0.00012
	0.00008**	0.00008**	0.00010		0.00016
colony	-0.00042	-0.00032	-0.00026		-0.00026
	0.00011**	0.00012**	0.00011*		0.00027
constant	-0.02172	-0.00656	-0.00681	-0.00816	-0.00681
	0.00231**	0.00425	0.00306*	0.00170**	0.00314*
number of observations	2079	2079	2079	2291	2079
R-squared	0.12	0.22	0.66	0.85	0.66

OLS estimates, with standard errors presented under each estimated coefficient. + significant at 10%; * significant at 5%; ** significant at 1%
per worker gdp is the level of per worker GDP, PPP-adjusted (constant 1996 international dollars), lagged by one year.
distance is the great-circle distance.
common language is a dummy variable equal to one if a common language is spoken in both destination and origin countries.
colony is a dummy variable for pairs of countries ever in a colonial relationship.
See Appendix 1 for data sources.

Table 2. Cross-country regressions

Equation	1	2	3
	1981-1985	1986-1990	1991-1995
Dependent variable	Emigration rate		
log per worker gdp (destination)	0.00127	0.0026	0.00302
	0.00065+	0.00087**	0.00075**
log per worker gdp (origin)	-0.00009	-0.00016	-0.00015
	0.00012	0.00015	0.00012
log distance	-0.0001	-0.0001	-0.00012
	0.00012	0.00014	0.0001
common language	0.0006	0.00069	0.0004
	0.00026*	0.00030*	0.00025
colony	-0.00033	-0.00033	-0.00024
	0.00034	0.0004	0.00033
constant	-0.01146	-0.02468	-0.02929
	0.00715	0.00953*	0.00826**
number of obs	137	154	172
R-squared	0.1	0.12	0.14

OLS estimates. The standard errors are presented under each estimated coefficient.
+ significant at 10%; * significant at 5%; ** significant at 1%
per worker gdp is the level of per worker GDP, PPP-adjusted (constant 1996 international dollars), averaged over the five years preceding the relevant period (1976-1980 for the regression in column (1), for example).
distance is the great-circle distance.
common language is a dummy variable equal to one if a common language is spoken in both destination and origin countries.
colony is a dummy variable for pairs of countries ever in a colonial relationship.
See Appendix 1 for data sources.

Table 3. Yearly regressions

Equation	1	2	3	4	5	6	7	8
	1981	1982	1983	1984	1985	1986	1987	1988
Dependent variable	Emigration rate							
log per worker gdp (destination)	0.00099	0.00209	0.00271	0.00171	0.00172	0.00187	0.00203	0.00216
	0.00139	0.00103*	0.00102**	0.00075*	0.00069*	0.00083*	0.00097*	0.00088*
log per worker gdp (origin)	-0.0001	-0.00008	-0.00008	-0.00016	-0.00016	-0.00021	-0.00024	-0.00019
	0.00032	0.00022	0.00017	0.00013	0.00012	0.00015	0.00017	0.00014
log distance	-0.00103	-0.00047	-0.00046	-0.00011	-0.00011	-0.00014	-0.00013	-0.00012
	0.00042*	0.00028	0.00022*	0.00012	0.00011	0.00014	0.00015	0.00013
common language	0.00121	0.00091	0.00068	0.00055	0.00058	0.00075	0.00097	0.00086
	0.00058*	0.00046+	0.00035+	0.00025*	0.00024*	0.00029*	0.00033**	0.00029**
colony	-0.00094	-0.00039	-0.00024	-0.00025	-0.00026	-0.00041	-0.00062	-0.00057
	0.00118	0.00056	0.00044	0.00033	0.00033	0.00039	0.00044	0.00039
constant	-0.00051	-0.01699	-0.02355	-0.01523	-0.01542	-0.01637	-0.01786	-0.01984
	0.01627	0.01137	0.01098*	0.00818+	0.00763*	0.00918+	0.01069+	0.00967*
number of obs	52	81	95	135	136	136	137	139
R-squared	0.19	0.15	0.16	0.11	0.12	0.13	0.14	0.15

OLS estimates. The standard errors are presented under each estimated coefficient. + significant at 10%; * significant at 5%; ** significant at 1%
per worker gdp is the level of per worker GDP, PPP-adjusted (constant 1996 international dollars), lagged by one year.
distance is the great-circle distance.
common language is a dummy variable equal to one if a common language is spoken in both destination and origin countries.
colony is a dummy variable for pairs of countries ever in a colonial relationship.
See Appendix 1 for data sources.

Table 3. Yearly regressions (cont.)

Equation	1	2	3	4	5	6	7
	1989	1990	1991	1992	1993	1994	1995
Dependent variable	Emigration rate						
log per worker gdp (destination)	0.00382	0.00553	0.00544	0.00331	0.00357	0.00326	0.00249
	0.00119**	0.00126**	0.00115**	0.00093**	0.00083**	0.00077**	0.00072**
log per worker gdp (origin)	-0.00018	-0.00025	-0.00036	-0.00021	-0.00015	-0.00006	-0.0001
	0.00019	0.00019	0.00018*	0.00013	0.00012	0.00011	0.0001
log distance	-0.00018	-0.00021	-0.00022	-0.00015	-0.00013	-0.00011	-0.00015
	0.00017	0.00018	0.00015	0.00011	0.0001	0.0001	0.00009+
common language	0.00072	0.00035	0.00029	0.00053	0.00047	0.00042	0.00054
	0.00038+	0.0004	0.00037	0.00028+	0.00025+	0.00023+	0.00022*
colony	-0.00051	-0.00009	-0.00015	-0.00031	-0.00021	-0.00018	-0.00042
	0.0005	0.00053	0.00049	0.00037	0.00033	0.00031	0.00029
constant	-0.03706	-0.05429	-0.05212	-0.03159	-0.03509	-0.03288	-0.02412
	0.01318**	0.01392**	0.01265**	0.01022**	0.00905**	0.00840**	0.00783**
number of obs	139	152	159	162	162	168	153
R-squared	0.14	0.16	0.18	0.14	0.16	0.14	0.15

OLS estimates. The standard errors are presented under each estimated coefficient.
+ significant at 10%; * significant at 5%; ** significant at 1%
per worker gdp is the level of per worker GDP, PPP-adjusted (constant 1996 international dollars), lagged by one year.
distance is the great-circle distance.
common language is a dummy variable equal to one if a common language is spoken in both destination and origin countries.
colony is a dummy variable for pairs of countries ever in a colonial relationship.
See Appendix 1 for data sources.

Table 4. Regressions by country of destination

Equation	1	2	3	4	5	6	7
Destination country	Australia	Belgium	Canada	Denmark	France	Germany	Japan
Dependent variable	Emigration rate						
log per worker gdp (destination)	0.0018	0.00008	0.00176	-0.00167	0.00055	-0.00297	0.00035
	0.00131	0.00009	0.00072*	0.00198	0.00029+	0.00385	0.00018+
log per worker gdp (origin)	-0.00048	0.00014	0.00019	-0.00061	-0.00049	-0.00081	-0.00017
	0.0005	0.00016	0.00032	0.00071	0.00028	0.00238	0.00012
constant	-0.01529	-0.00203	-0.02056	0.02301	-0.00053	0.04052	-0.0018
	0.01049	0.00090*	0.01036+	0.02659	0.00287	0.04469	0.00157
number of obs	202	117	256	71	76	61	147
R-squared	0.77	0.88	0.76	0.94	0.83	0.71	0.62

Equation	8	9	10	11	12	13	14
Destination country	Luxembourg	Netherlands	Norway	Sweden	Switzerland	UK	USA
Dependent variable	Emigration rate						
log per worker gdp (destination)	-0.00013	0.00061	0.00013	0.00093	0.00062	-0.00004	0.00271
	0.00006+	0.00021*	0.00015	0.00068	0.00045	0.00012	0.00111*
log per worker gdp (origin)	0.00049	-0.00033	-0.00019	-0.00037	0.00042	-0.0001	-0.00035
	0.00021+	0.00012*	0.00009+	0.00024	0.00038	0.00015	0.00017*
constant	-0.00396	-0.0032	0.00063	-0.00712	-0.01121	0.00146	-0.02507
	0.00175+	0.00120*	0.00099	0.00574	0.00875	0.00177	0.01067*
number of obs	81	94	101	163	125	401	396
R-squared	0.86	0.88	0.9	0.73	0.89	0.75	0.84

OLS estimates with dummy variables for countries of origin and standard errors clustered by country of origin. The standard errors are presented under each estimated coefficient. + significant at 10%; * significant at 5%; ** significant at 1%
per worker gdp is the level of per worker GDP, PPP-adjusted (constant 1996 international dollars), lagged by one year.
See Appendix 1 for data sources.

Table 5. Panel data regressions: Robustness Checks

Equation	1	2	3
Method	Instrumental Variables Estimation	OLS	OLS
Dependent variable	Emigration rate		
log per worker gdp (destination)	0.0016	0.00131	0.00055
	0.00076*	0.00042**	0.0008
log per worker gdp (origin)	-0.00125	-0.00053	0.00029
	0.00091	0.00039	0.00067
log distance	-0.00034	-0.00035	-0.00029
	0.00010**	0.00010**	0.00011*
common language	0.00003	0.00007	0.00017
	0.00014	0.00015	0.00016
colony	-0.00021	-0.00029	-0.00044
	0.00026	0.00028	0.00034
log yrs schooling (destination)		-0.00109	-0.00035
		0.00063+	0.00049
log yrs schooling (origin)		0.00082	0.00218
		0.00044+	0.00145
log capital per worker (destination)			0.00033
			0.00073
log capital per worker (origin)			-0.00114
			0.00057*
constant	0.00093	-0.00476	-0.00242
	0.00824	0.0029	0.00358
number of obs	1902	1905	1235
R-squared	0.53	0.67	0.58

Standard errors, clustered by country pairs, are presented under each estimated coefficient. + significant at 10%; * significant at 5%; ** significant at 1%
In regression (1), I use terms of trade (lagged by one year) as an instrument for per worker GDP (lagged by one year) in both destination and origin country.
per worker gdp is the level of per worker GDP, PPP-adjusted (constant 1996 international dollars), lagged by one year.
distance is the great-circle distance. common language is a dummy variable equal to one if a common language is spoken in both destination and origin countries. *colony* is a dummy variable for pairs of countries ever in a colonial relationship.
log yrs schooling is the log of the average schooling years in the total population over age 15, lagged by one year.
log capital per worker is non-residential capital stock per worker (1985 intl. prices), lagged by one year.
See Appendix 1 for data sources.

Table 6: Dynamic regressions: network effects

Equation	1	2	3
	destination and origin countries d.v.	country pair d.v.	Arellano and Bond estimator
Dependent variable	**Emigration rate**		
emigration rate(t-1)	0.7989	0.56037	0.63033
	0.04237**	0.05740**	0.02489**
emigration rate(t-2)			-0.26199
			0.02415**
per worker gdp (destination)	0.00018	0.00043	0.00169
	0.00016	0.00019*	0.00044**
per worker gdp (origin)	-0.00014	-0.00019	0.00000
	0.00009	0.00016	0.00025
log distance	-0.00008		
	0.00003**		
constant	0.00011	-0.00301	-0.00003
	0.00108	0.00116*	0.00001*
number of obs	2021	2232	1707
R-squared	0.88	0.9	

OLS estimates. The standard errors, clustered by country pair, are presented under each estimated coefficient.
+ significant at 10%; * significant at 5%; ** significant at 1%
per worker gdp is the level of per worker GDP, PPP-adjusted (constant 1996 international dollars), lagged by one year.
distance is the great-circle distance.
See Appendix 1 for data sources.

Appendix 1. Summary Statistics (1980-1996)

Variable	Obs	Mean	Std. Dev.	Min	Max
emigration rate	2683	0.0006167	0.0018146	2.77E-07	0.0276537
per worker gdp (destination)	2865	42505.45	7373.63	25251.65	80026.46
per worker gdp (origin)	2358	23065.63	15832.55	1027.362	57259.25
distance	2482	3782.063	2802.015	161.9276	11504.2
common language	2504	0.3178914	0.46575	0	1
colony	2504	0.1425719	0.3497056	0	1
years schooling (destination)	2804	9.960016	1.268303	6.888	11.892
years schooling (origin)	2367	6.880774	2.625928	1.897	11.892
capital per worker (destination)	2200	34317.06	11121.74	17285	76733
capital per worker (origin)	1502	18443.39	12983.49	702	48135

The emigration rate (immigrant inflow from origin to destination country, divided by origin country's population) is from the IMS data set (OECD(1997)).

Per worker GDP, PPP-adjusted (constant 1996 international dollars) is from the Penn World Tables, version 6.1.

Distance, common language, and colony (countries ever in a colonial relationship) are from Glick and Rose (2001).

Years of schooling are from Barro and Lee (2000) data set.

Capital per worker (Nonresidential Capital Stock per Worker (1985 intl. prices)) is from the Penn World Tables, version 5.6.

Appendix 2. Average inflows into each destination country, by country of origin (1980-1995)

Australia (1983-1995)

country of origin	inflow
UK	17095
New Zealand	11045
Vietnam	8048
Hong Kong	5739
Philippines	5379
Malaysia	3493
India	3069
China	2934
Former Yugoslavia	2790
South Africa	2441
Sri Lanka	2146
Lebanon	2089
USA	1724
Fiji	1682
Poland	1608
Ireland	1462
Chinese Taipei	1358
Germany	1303
Former USSR	1021
Portugal	767
total (above inflows)	77193
overall total	101492
percentage covered	76.06%
percentage change	-6.22%

Belgium

country of origin	inflow
France	6072
Netherlands	6014
USA	2930
Germany	2916
UK	2899
Morocco	2801
Italy	2495
Turkey	2239
Zaire	1966
Portugal	1435
Japan	833
Spain	833
Former Yugoslavia	829
Greece	759
Poland	655
China	589
Algeria	382
Tunisia	310
total (above inflows)	36957
overall total	44756
percentage covered	82.57%
percentage change	13.46%

Canada

country of origin	inflow
Hong Kong	19334
India	10437
Philippines	9441
UK	9034
Vietnam	8791
Poland	7550
USA	7459
China	6292
Lebanon	3917
Sri Lanka	3791
Portugal	3653
Jamaica	3543
Chinese Taipei	3255
Guyana	3108
El Salvador	2697
Haiti	2243
Iran	2193
France	2070
Former Yugoslavia	1933
South Korea	1584
Trinidad Tobago	1433
Romania	1241
Pakistan	1037
Former USSR	791
Somalia	195
total (above inflows)	117022
overall total	165588
percentage covered	70.67%
percentage change	48.29%

Denmark (1984-1994)

country of origin	inflow
Somalia	1264
UK	1068
Turkey	1042
Germany	805
Iraq	789
Norway	699
Sweden	612
USA	606
Iran	570
Vietnam	549
Former Yugoslavia	481
Iceland	479
Poland	448
Thailand	366
Pakistan	356
Lebanon	335
Netherlands	304
France	269
Morocco	215
Italy	200
Finland	181
total (above inflows)	11638
overall total	15155
percentage covered	76.80%
percentage change	75.28%

total (above inflows) is the sum of the average immigrant inflows (1980-1995) by country of origin from the table.
overall total is the total average immigrant inflow (1980-1995), as reported by OECD (1997).
percentage covered is the percentage of *overall total* covered by inflows by origin country (*total (above inflows)/overall total*)
percentage change is the percentage change of the *overall total* during the period between 1980 and 1995.

Appendix 2. Average inflows into each destination country, by country of origin (1980-1995) (cont.)

France

country of origin	inflow
Morocco	11892
Algeria	9187
Turkey	5777
Tunisia	3083
Lebanon	2818
USA	2403
Haiti	2183
Portugal	2050
Vietnam	1761
Zaire	1437
Poland	1422
Japan	1219
China	1084
Former Yugoslavia	1084
Sri Lanka	899
Romania	891
Cambodia	860
Spain	400
total (above inflows)	50450
overall total	72838
percentage covered	69.26%
percentage change	-6.23%

Germany *

country of origin	inflow
Poland	117019
Former Yugoslavia	92124
Bosnia-Herzegovina	76836
Turkey	68791
Romania	61910
Italy	39184
Croatia	24056
Former USSR	23365
Hungary	21835
Greece	20372
Bulgaria	19245
USA	17670
Former CSFR	10692
Portugal	9654
Spain	4705
Morocco	4375
Slovenia	2658
Tunisia	2249
total (above inflows)	616740
overall total	646144
percentage covered	95.45%
percentage change	24.85%

Japan

country of origin	inflow
China	35425
USA	35367
Philippines	35121
South Korea	21052
Chinese Taipei	10882
UK	9614
Brazil	6779
Hong Kong	6296
Thailand	5913
Germany	5334
Canada	3449
Peru	1008
total (above inflows)	176240
overall total	220419
percentage covered	79.96%
percentage change	-42.10%

Luxembourg

country of origin	inflow
Portugal	2170
France	1272
Belgium	897
Germany	662
Italy	441
Netherlands	281
USA	256
Spain	124
total (above inflows)	6103
overall total	7988
percentage covered	76.41%
percentage change	29.73%

total (above inflows) is the sum of the average immigrant inflows (1980-1995) by country of origin from the table.
overall total is the total average immigrant inflow (1980-1995), as reported by OECD (1997).
percentage covered is the percentage of *overall total* covered by inflows by origin country (*total (above inflows)/overall total*)
percentage change is the percentage change of the *overall total* during the period between 1980 and 1995.
* Figures for migrants from the former Yugoslavia to Germany do not include Croatia from 1992 and Bosnia-Herzegovina from 1993. Data from the former USSR to Germany does not include Russia from 1992.

Appendix 2. Average inflows into each destination country, by country of origin (1980-1995) (cont.)

Netherlands

country of origin	inflow
Turkey	8363
Former Yugoslavia	7392
Morocco	6537
Germany	5295
UK	4575
Suriname	4416
USA	2303
Belgium	2050
France	1517
Poland	1148
Italy	893
total (above inflows)	44489
overall total	62500
percentage covered	71.18%
percentage change	-16.04%

Norway *

country of origin	inflow
Bosnia-Herzegovina	3728
Denmark	2201
Sweden	1526
UK	1253
USA	987
Former Yugoslavia	868
Pakistan	682
Iran	669
Vietnam	612
Chile	537
Somalia	468
Sri Lanka	450
Germany	399
total (above inflows)	14380
overall total	16738
percentage covered	85.91%
percentage change	39.83%

Sweden (1983-1995)

country of origin	inflow
Bosnia-Herzegovina	16972
Iran	4048
Finland	3880
Norway	3118
Former Yugoslavia	2840
Iraq	2051
Denmark	1877
Somalia	1724
Chile	1631
Poland	1484
Turkey	1214
Ethiopia	947
Russian Federation	910
Lebanon	896
USA	831
Croatia	784
Germany	761
Romania	746
UK	715
Thailand	603
India	369
Greece	311
total (above inflows)	48712
overall total	
percentage covered	
percentage change	61.88%

Switzerland

country of origin	inflow
Former Yugoslavia	18716
Portugal	9085
Germany	8333
Italy	8216
France	4655
Spain	4402
Turkey	4195
USA	2530
UK	2407
Austria	1728
Netherlands	1607
Canada	687
total (above inflows)	66561
overall total	81469
percentage covered	81.70%
percentage change	24.68%

total (above inflows) is the sum of the average immigrant inflows (1980-1995) by country of origin from the table.
overall total is the total average immigrant inflow (1980-1995), as reported by OECD (1997).
percentage covered is the percentage of *overall total* covered by inflows by origin country (*total (above inflows)/overall total*)
percentage change is the percentage change of the *overall total* during the period between 1980 and 1995.
* Figures for migrants from the former Yugoslavia to Norway do not include Bosnia-Herzegovina from 1993.

Appendix 2. Average inflows into each destination country, by country of origin (1980-1995) (cont.)

United Kingdom

country of origin	inflow
Pakistan	5817
India	5047
Bangladesh	3796
USA	3776
Australia	2659
New Zealand	1964
Nigeria	1556
Iran	1501
Japan	1474
Hong Kong	1287
Ghana	1093
Canada	1035
Sri Lanka	1021
Philippines	986
South Africa	926
Turkey	822
Jamaica	775
Malaysia	701
Iraq	500
Kenya	481
Poland	481
Thailand	444
Germany	419
Cyprus	402
Morocco	380
Spain	363
Sweden	355
France	345
Italy	340
Netherlands	289
Former Yugoslavia	276
Portugal	223
total (above inflows)	41534
overall total	53831
percentage covered	77.16%
percentage change	-20.49%

United States

country of origin	inflow
Mexico	199862
Philippines	51886
Vietnam	45041
China	32824
Dominican Republic	30471
India	29754
South Korea	29197
Former USSR	23231
El Salvador	21901
Jamaica	20219
Cuba	15174
Haiti	15168
UK	14939
Iran	14596
Poland	13534
Canada	12980
Chinese Taipei	12962
Colombia	12696
Laos	12165
Ireland	12054
Guatemala	9328
Guyana	9243
Cambodia	8108
Pakistan	7725
Peru	7637
Germany	7005
Hong Kong	6994
Thailand	6270
Ecuador	6189
Nicaragua	5626
Honduras	5507
Bangladesh	2684
total (above inflows)	702970
overall total	818688
percentage covered	85.87%
percentage change	35.79%

total (above inflows) is the sum of the average immigrant inflows (1980-1995) by country of origin from the table.
overall total is the total average immigrant inflow (1980-1995), as reported by OECD (1997).
percentage covered is the percentage of *overall total* covered by inflows by origin country (*total (above inflows)/overall total*)
percentage change is the percentage change of the *overall total* during the period between 1980 and 1995.

Figure 2: Total immigrant inflow by destination country