Visa Policies, Networks and the Cliff at the Border*

Simone Bertoli[a] and Jesús Fernández-Huertas Moraga[b]

[a] *CERDI*†*, University of Auvergne and CNRS*
[b] *FEDEA*‡ *and IAE, CSIC*

September 27, 2012

Abstract

The scale of international migration flows depends on moving costs that are, in turn, influenced by host-country policies and by the size of migrant networks at destination. This paper estimates the influence of visa policies and networks upon bilateral migration flows to multiple destinations. We rely on a Poisson pseudo-maximum likelihood estimator to derive estimates that are consistent under more general distributional assumptions on the underlying RUM model than the ones commonly adopted in the literature. We derive bounds for the estimated direct and indirect effects of visa policies and networks that reflect the uncertainty connected to the use of aggregate data, and we show that bilateral migration flows can be highly sensitive to the immigration policies set by other destination countries, an externality that we are able to quantify. We also show that both the direct and the indirect effects are larger for low-skill migration flows.

Keywords: international migration, networks, visa policies, multiple destinations, externalities.

JEL classification codes: F22, O15, J61.

---

*The authors are grateful to Pedro Albarrán, Michel Beine, Michael Clemens, Francesc Ortega, Çaglar Özden, Chris Parsons and Hillel Rapoport for their comments, and to the participants in the Eleventh Journées Louis-André Gérard-Varet, in the Fifth Migration and Development Conference organized by the Agence Française de Développement, the Center for Global Development and the World Bank and in seminar presentations at CERDI and at the University of Alicante; they are also grateful to Eric Neumayer for making his dataset on bilateral visa policies publicly available; the usual disclaimers apply.

†Bd. F. Mitterrand, 65, F-63000, Clermont-Ferrand; email: simone.bertoli@udamail.fr.

‡Jorge Juan, 46, E-28001, Madrid; email: jfernandezhuertas@fedea.es (corresponding author).

"Not only is the world not flat, it is not a curb nor a barrier. Rather, the world has a massive cliff at the U.S. border (and, one suspects, most other rich industrial countries have similarly sized cliffs)."
(Pritchett, 2009, p. 274)

1 Introduction

The share of the world population currently residing outside its country of birth is estimated at around 3 percent (UN Population Division, 2008). It is generally argued that the legal restrictions on cross-border human mobility play a key role in keeping this figure low, as "policy barriers in the destination countries surely play a major role in constraining emigration" (Clemens, 2011, p. 83), and "labor mobility is likely lower than it could be by a factor of between two and five, because it is constrained by host-country policies" (Pritchett, 2006, p. 69).

The policies that exert an influence on the size of migration flows are not only the regulations that shape the legal framework for immigrant admission such as quotas or point-based systems, but they encompass any policy intervention that influences the costs and expected benefits from migration. Policy-induced migration costs create a "cliff at the border" (Pritchett, 2009) that hinders the flow of people across countries. Mayda (2010) and Ortega and Peri (2009, 2012) provide empirical evidence that an aggregate measure of restrictiveness of immigration policies reduces incoming flows from all origin countries. Still, some relevant host-country policies are bilateral in nature, so that would-be migrants from different origins can face differently sized cliffs along the same border.

Visa policies are one part of the legal framework regulating non-immigrant temporary admission at destination and represent a factor that can shape the height of the cliff at the border. The requirement of a visa to enter a country can impose substantial costs on travelers, as it forces them to submit an application to the consular offices of their intended destination, which can ask for processing fees, impose long waiting times, and possibly deny the visa (Neumayer, 2006). A visa waiver allows travelers to move across borders at a substantially lower cost, and with a greatly reduced uncertainty with respect to their admission at destination.[1] This, in turn, suggests that the bilateral visa regime can also influence the scale of migration flows, as they determine the cost of entering legally into

[1] Neumayer (2010) provides econometric evidence on the negative impact of visa requirements on the number of travelers between countries.

the country of destination, and then overstaying there beyond the period for which non-immigrant admission was granted. The US General Accounting Office (2004) reports that overstayers amounted to 2.3 million in 2000,[2] accounting for at least 27 percent of illegal immigrants in the US (US General Accounting Office, 2004, p. 10).

Still, the evidence on the influence of the visa regime, and of bilateral immigration policies more in general, upon the scale of international migration flows is limited. Bertoli, Fernández-Huertas Moraga, and Ortega (2011) present descriptive evidence on the role of the visa waivers that Spain used to grant to some of its former colonies in Latin America in determining the size of immigration flows from Ecuador,[3] and Bertoli and Fernández-Huertas Moraga (2011) provide econometric evidence of the influence of changes in visa policies in shaping the size of bilateral flows during the surge of Spanish immigration that began in the late 1990s. Visa waivers have only a marginally significant effect on migration rates in Grogger and Hanson (2011).

How can we reconcile the perception that destination country policies represent a binding constraint on international migration with the limited empirical evidence on the effects of bilateral policies? A part of the answer is probably contained in our initial quote from Pritchett (2009), where he suggests that "most other rich industrial countries have similarly sized cliffs."

The similar size of the cliffs across destinations is related to a key feature of bilateral immigration policies toward a given origin country: they are closely correlated across different destinations. An article published by *The Economist* in 2010 reveals that the visa regimes that citizens from different origin countries face is highly polarized: holding a passport of a developed country grants visa-free admission (almost) everywhere, while citizens from developing countries need to apply for a visa to be admitted in most destinations around the world.[4] Such a similarity in the bilateral visa policies toward the citizen of any given origin country can come from a policy coordination at the supranational level, as it occurs at the EU level,[5] or from a shared perception on the potential for illegal immigration from

[2]This figure does not include Mexican, Canadian or short-term overstays from other countries.

[3]Their analysis exploits the fact that the removal of the Spanish visa waiver for Ecuadorians in August 2003 occurred at a time when the conditions of entry to other destinations remained unchanged.

[4]"No visa required: Who has more freedom to travel?", *The Economist*, August 25, 2010.

[5]The European Council establishes a list of countries whose citizens must be requested a visa to be admitted to any country within the Schengen area; if a country is not on this list, then member states are free to decide whether to grant a visa waiver or impose a visa requirement (Council Regulation (EC) No

a given country. By the same token, the admission of a country within the Schengen area simultaneously changes the terms at which its citizens are admitted in *all* the other signatory countries.

This feature of bilateral immigration policies creates a key analytical challenge for the identification of their effects whenever the scale of the bilateral migration *rate* toward a destination depends on the attractiveness of alternative destinations, that is, when there is multilateral resistance to migration (Bertoli and Fernández-Huertas Moraga, 2011). Such a dependency implies that the identification of the effect of the visa regime can be confounded by the visa policies adopted by other countries toward the same country of origin: would-be migrants' destination choice depends on the *relative* size of the cliffs that characterize different borders.

These arguments entail that the limited evidence on the effectiveness of bilateral immigration policies might be related to the confounding influence of the policies adopted in other countries of destination. The contribution of this paper is to propose an econometric approach that is able to identify the effect of bilateral variables on bilateral migration rates while controlling for such a confounding effect in a cross-sectional setting.[6] In addition, the source of these third-country effects is related to the existence of externalities in bilateral migration policies and other causes of international migration that we are able to quantify.

We employ a Poisson pseudo-maximum likelihood, PPML, estimator that allows us (i) to be consistent with underlying random utility maximization models with different patterns of dependency of the bilateral flows on the attractiveness of other destinations (Guimaraes, Figueiredo, and Woodward, 2004; Schmidheiny and Brulhart, 2011), (ii) to deal with the presence of zeros (Santos Silva and Tenreyro, 2006), and (iii) to tackle the endogeneity of migration networks (Terza, Basu, and Rathouz, 2008; Tenreyro, 2007), which can also influence the scale of bilateral migration costs (McKenzie and Rapoport, 2010; Beine, Docquier, and Ozden, 2011).

The consistency of the PPML estimator with an underlying random utility maximization model was first established by Guimaraes, Figueiredo, and Woodward (2003), and then extended by Guimaraes, Figueiredo, and Woodward (2004) and Schmidheiny and Brulhart

539/2001 of 15 March 2001).

[6]Bertoli and Fernández-Huertas Moraga (2011) adopt the Common Correlated Effects, CCE, estimator proposed by Pesaran (2006), which requires a longitudinal dimension of the data that is often unavailable with international migration data.

(2011) under more general specifications of the stochastic component of location-specific utility. The RUM-consistency of PPML under different specifications of the error term creates, as discussed in Schmidheiny and Brulhart (2011), an uncertainty about the size of the estimated elasticities of bilateral flows with respect to the regressors that had not been considered yet by the international migration literature. Our paper extends Schmidheiny and Brulhart (2011), proposing bounds for the estimated elasticity under a more general specification of the stochastic properties of the underlying theoretical model describing the location-decision problem that would-be migrants face.

This paper is related to three different strands of literature. First, the literature on the determinants of international migration flows that we reviewed above.[7] Second, the literature on discrete choice models (McFadden, 1974, 1978; Cardell, 1997; Train, 2003; Wen and Koppelman, 2001; Papola, 2004); third, the papers establishing the consistency of aggregate count data models with individual-level utility maximizing behavior (Guimaraes, Figueiredo, and Woodward, 2003, 2004; Schmidheiny and Brulhart, 2011).

Our econometric analysis draws on the international migration data assembled by Docquier, Lowell, and Marfouk (2009), which we combine with the dataset by Ozden, Parsons, Schiff, and Walmsley (2011) to obtain information on the size of the migration networks in 1960, and with the dataset on bilateral visa policies by Neumayer (2006).

The choice of the various specifications of the model to be estimated with PPML are derived from a simple RUM model. The estimates confirm the significant influence of migration networks evidenced by Beine, Docquier, and Ozden (2011), and they also reveal that visa policies play a significant role in shaping the height of the cliffs at the border: when the attractiveness of other destinations is properly controlled for, a visa requirement is estimated to reduce the scale of bilateral migration flows between 40 and 47 percent on average. Such an effect is not significant in specifications that are only consistent with more restrictive assumptions on the underlying RUM model, and whose validity is questioned by the tests that we conduct on the residuals. Our results confirm the pressing need to properly control for the confounding influence of the attractiveness of alternative destinations, that is, multilateral resistance to migration.

As far as migration policy externalities are concerned, we estimate that a visa require-

[7] Other relevant empirical papers include Clark, Hatton, and Williamson (2007), Belot and Hatton (2012), Lewer and den Berg (2008) and McKenzie, Theoharides, and Yang (2012).

ment imposed by one destination on one origin can increase bilateral migration flows to other destinations by between 3 and 17 percent on average. In some particular cases, this externality effect might even be larger than the own-country effect. These results are robust when we estimate our model for each skill group, and we find that low-skill migration flows respond slightly more to changes in visa requirements than high-skill flows.

The rest of the paper is structured as follows: Section 2 develops a simple random utility maximization model that describes the location choice that would-be migrants face. Section 3 discusses the two main RUM-consistent approaches to the estimation, presenting the arguments that justify our choice to rely on PPML estimation. The data sources and the basic descriptive statistics are presented in Section 4. Section 5 contains the results from the econometric analysis, and Section 6 draws the main conclusions.

2 A RUM model of international migration

Consider a population of s_j individuals originating from a country $j \in H$, who have to chose their preferred location among the n countries belonging to the set D, including j itself.[8] Let m_{jk} represent the scale of the bilateral gross migration flow from country j to country k, and $\boldsymbol{m}_j$ be the $n \times 1$ vector that collects all the bilateral migration flows originating from country j.[9] We can express m_{jk} as:

$$m_{jk} = s_j p_{jk} \eta_{jk} \tag{1}$$

where p_{jk} is the probability that an individual from country j will move to country $k \in D$ and $\boldsymbol{\eta}_j$ is a vector of spatially uncorrelated errors, with $E(\eta_{jk}) = 1$ for all k.

2.1 Choice probabilities

The n elements of the vector $\boldsymbol{p}_j$ are the outcome of a location decision problem that individuals face, which we describe through a random utility maximization problem. Specifically, the utility that the individual i from country j obtains from opting for destination k is given by:

[8] We present the RUM model omitting the time dimension of the location decision problem that would-be migrants face, but the analysis can be extended to allow for such a dimension.

[9] Observe that m_{jj} represents the number of individuals who opted to stay in the home country j.

$$U_{ijk} = V_{jk} + \epsilon_{ijk} = \boldsymbol{x}_{jk}{}'\boldsymbol{\beta} + \epsilon_{ijk} \tag{2}$$

where the deterministic component of utility V_{jk} is a linear function of the vector $\boldsymbol{x}_{jk}$, and ϵ_{ijk} represents an individual-specific stochastic component.[10] The vector $\boldsymbol{p}_{ij} = (p_{ij1}, ..., p_{ijk}, ...)$ that collects the choice probabilities for individual i over the n locations depends on the assumptions about the distribution of the stochastic term. We assume that ϵ_{ijk} follows an Extreme Value Type-1 marginal distribution, not i.i.d. like most of the literature but with a stochastic term that can be positively correlated across destinations; ϵ_{ijk} can thus be obtained from a Generalized Extreme Value generating function (McFadden, 1978), as most of the econometric approaches adopted in the literature are all consistent with different GEV models.[11]

2.1.1 Distributional assumptions

Let the set of possible locations D be partitioned into m subsets b, also called nests, and let $b(k) \subseteq D$ denotes the unique subset to which location k belongs to. Nests are groups of destination countries that share some unobservable characteristics from the point of view of potential migrants. The stochastic component ϵ_{ijk} of utility for country k is given by:

$$\epsilon_{ijk} = (1 - \tau)\nu_{ijb(k)} + \tau\upsilon_{ijk} \tag{3}$$

where $\tau \in (0, 1]$, $\upsilon_{ijk} \overset{\text{iid}}{\sim}$ EVT-1 and $\nu_{ijb(k)}$ is the unique random variable, whose distribution depends on τ, which ensures that also ϵ_{ijk} follows a marginal distribution EVT-1 (Cardell, 1997). The presence of the nest-specific stochastic component $\nu_{ijb(k)}$ introduces a positive correlation in the realizations of the stochastic component of utility for the locations belonging to the same nest; specifically, we have that $\text{corr}(\epsilon_{ijk}, \epsilon_{ijh}) = \sqrt{1 - \tau^2}$ if $b(k) = b(h)$, and zero otherwise.

[10] We are assuming here, as the literature does, that the vector $\boldsymbol{\beta}$ is not origin-specific, so that pooling migration data across origins poses no problem; Llull (2011) represents an exception in this respect.

[11] Partial exceptions are also represented by Clark, Hatton, and Williamson (2007) and Mayda (2010) who assume normality of the stochastic component in their theoretical model but then adopt an estimation approach that is consistent with an identically and independently distributed EVT-1 error term.

2.1.2 The vector of choice probabilities

The element k in the vector of choice probabilities $\boldsymbol{p}_{ij}$ is equal to:

$$p_{ijk} = \frac{e^{\boldsymbol{x}_{jk}'\boldsymbol{\beta}/\tau}\left(\sum_{l\in b(k)} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau}\right)^{\tau-1}}{\sum_q \left(\sum_{l\in b_q} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau}\right)^{\tau}} \tag{4}$$

Averaging over individual decisions, we have that $\boldsymbol{p}_{ij} = \boldsymbol{p}_j$, which in turn allows us to rewrite the element k of the vector $\boldsymbol{m}_j$ of bilateral migration flows as follows:

$$m_{jk} = s_j \frac{e^{\boldsymbol{x}_{jk}'\boldsymbol{\beta}/\tau}\left(\sum_{l\in b(k)} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau}\right)^{\tau-1}}{\sum_q \left(\sum_{l\in b_q} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau}\right)^{\tau}} \eta_{jk} \tag{5}$$

Our key interest is to understand whether, and under which hypotheses, we can recover a consistent estimate of the vector of parameters $\boldsymbol{\beta}$ in (2) from the observation of migration flows m_{jk} and of the vector $\boldsymbol{x}_{jk}$.

The assumptions on the stochastic component of location-specific utility in (3) are more general than those adopted by other papers in the literature; specifically, our distributional assumptions coincide with those adopted by Grogger and Hanson (2011) if we further assume that each location belongs to a singleton, i.e. $b(k) = \{k\}$ for any $k \in D$.[12] Similarly, we can obtain the distributional assumptions in Ortega and Peri (2009, 2012), Beine, Docquier, and Ozden (2011) and McKenzie, Theoharides, and Yang (2012) by imposing the restriction that all locations but the origin belong to a unique nest, i.e. $b(j) = \{j\}$ and $b(k) = b(h)$ for any $h, k \in D$.

The distributional assumption by Ortega and Peri (2009, 2012) implies that, conditional upon the deterministic component of location-specific utility, would-be migrants regard all possible countries but the origin as being close substitutes. The potential countries of destination share some unobservable factors that introduce a positive correlation in the realization of the stochastic component of utility that is higher the lower is the dissimilarity parameter τ. A lower τ is associated with a higher elasticity of migration flows to a destination h with respect to a change in the utility associated to any other destination $k \in D$, as discussed in Bertoli, Fernández-Huertas Moraga, and Ortega (2012).

[12] An alternative equivalent assumption is that $b(k) = D$ for all $k \in D$, as the inclusion of a stochastic component that is common to all locations does not affect the vector of choice probabilities that only depends on the differences in utility across locations and not on their levels.

The assumptions on the stochastic component ϵ_{ijk} that we introduced in (3) allow for a richer pattern of cross-elasticities, as would-be migrants can perceive a destination h to be a close substitute only for a subset of all the potential destinations, represented by the nest $b(h) \subset D$. The presence of a nest-specific stochastic component implies that the probability to opt for destination h responds more to a variation in the utility associated to country $k \in b(h)$ than to a variation in the utility associated to country $l \notin b(h)$. Intuitively, we are allowing the unobservable factors that influence would-be migrants' utility to vary across nests of potential destinations.

3 Two main approaches to the estimation

The estimation of the determinants of bilateral migration flows with aggregate data has to deal with some key analytical challenges, and we focus here on two of them: (i) its consistency with a general underlying random utility maximization model, and (ii) the presence of zero bilateral flows in the data.

We discuss here the two main approaches to the estimation that can be followed, and how they allow to deal with points (i)-(ii) above. The first one, which probably represents the industry-standard in the international migration literature,[13] involves a logarithmic transformation of the bilateral migration *rates* that can be derived from (5), while the second resorts to non-linear count data models to estimate directly the determinants of the bilateral migration *flows* described in (5).[14]

3.1 Estimation of bilateral migration rates

The first approach to the estimation adopts the logarithm of the bilateral gross migration rate, i.e. $y_{jk} = \ln(m_{jk}/m_{jj})$, as the dependent variable. From (5) we have that:

[13] This approach has been adopted, *inter alia*, by Clark, Hatton, and Williamson (2007), Lewer and den Berg (2008), Ortega and Peri (2009, 2012), Mayda (2010), McKenzie, Theoharides, and Yang (2012), Simpson and Sparber (2012), Beine, Docquier, and Ozden (2011), Grogger and Hanson (2011) and Bertoli and Fernández-Huertas Moraga (2011).

[14] This approach, which has a long-standing tradition in the internal migration literature, see Flowerdew and Aitkin (1982) for an early application, has been applied by Egger and Radulescu (2009), Beine, Noel, and Ragot (2011), Belot and Ederveen (2012) and Beine and Parsons (2012) in the international migration literature.

$$y_{jk} = (\boldsymbol{x}_{jk} - \boldsymbol{x}_{jj})'\boldsymbol{\beta}/\tau + (\tau - 1)\ln\Big(\sum_{l \in b(k)} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau}\Big) + \ln(\eta_{jk}/\eta_{jj})$$

which can be more compactly rewritten as follows:

$$y_{jk} = (\boldsymbol{x}_{jk} - \boldsymbol{x}_{jj})'\boldsymbol{\beta}/\tau + \varepsilon_{jk} \tag{6}$$

where:

$$\varepsilon_{jk} = (\tau - 1)\ln(\gamma_{jb(k)}) + \ln(\eta_{jk}/\eta_{jj})$$

Under the distributional assumptions in (3), y_{jk} depends on the deterministic components of location-specific utility in country k and j and on an error term ε_{jk} that is a function of an origin-nest specific term $\gamma_{jb(k)}$ reflecting the expected utility from migration to locations that belong to the same nest as k, and of a logarithmic transformation of the error term in (5). As discussed in Santos Silva and Tenreyro (2006), the assumption that $E(\eta_{jk}) = 0$ for any j, k does not suffice to conclude that $E[\ln(\eta_{jk}/\eta_{jj})] = 0$, as the latter will be, in general, a function of higher-order moments of the distribution of the error term in (5); this, in turn, implies that, if η_{jk} is heteroskedastic with a variance that depends on the regressors in (5), then the error term ε_{jk} in (6) will be correlated with the regressors, casting doubts on the unbiasedness of the estimates.

If one assumes that each destination is assigned to a singleton nest, then one can estimate bilateral migration flows as a function of origin and destination characteristics only, as in Clark, Hatton, and Williamson (2007), Mayda (2010) and Grogger and Hanson (2011).[15] Under this assumption, $\ln(\gamma_{jb(k)}) = \boldsymbol{x}_{jk}'\boldsymbol{\beta}/\tau$, so that the equation to be estimated simplifies to:

$$y_{jk} = (\boldsymbol{x}_{jk} - \boldsymbol{x}_{jj})'\boldsymbol{\beta} + \ln(\eta_{jk}/\eta_{jj})$$

The assumption that the stochastic components follow identically and independently distributed EVT-1 distributions, so that IIA holds, allows to recover the vector $\boldsymbol{\beta}$ that appears in the deterministic component of location-specific utility. Still, the appropriateness

[15] Mayda (2010) includes among the regressors an "atheoretical measure" (p. 1270) of the attractiveness of other locations, represented by the average of GDP per capita.

of this assumption critically hinges on the ability to correctly specify location-specific utility, as omitted variables could introduce spatial correlation in the error term.

A first approach that allows to relax IIA involves assuming that $b(j) = \{j\}$ and $b(k) = b(h)$ for any $h, k \in D$, with all destinations belonging to one unique nest. Under these assumptions, the error term simplifies to:

$$\varepsilon_{jk} = (\tau - 1)\ln(\gamma_j) + \ln(\eta_{jk}/\eta_{jj})$$

where $\gamma_j = \sum_{l \in D} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau}$ does *not* vary across destinations. This implies that the inclusion of origin dummies[16] among the regressors, which was proposed by Ortega and Peri (2009) and adopted in McKenzie, Theoharides, and Yang (2012) and Beine, Docquier, and Ozden (2011),[17] suffices to remove the spatial and serial dependence of the error term, and its correlation with the regressors.[18] This estimation strategy, which is valid under the restrictions on (3) that we just discussed,[19] does not allow to separately identify the value of the vector of coefficients $\boldsymbol{\beta}$ and of the dissimilarity parameter τ, but just the ratio $\boldsymbol{\beta}/\tau$. The inclusion of the origin-time dummies controls for the component of the error term ε_{jk} where τ enters separately from $\boldsymbol{\beta}$; if the true value of τ is unknown, $\boldsymbol{\beta}$ can be recovered from individual-level data as in Bertoli, Fernández-Huertas Moraga, and Ortega (2012), as these data allow to estimate the within-nest correlation in the stochastic component of utility.

A different approach has been adopted by Bertoli and Fernández-Huertas Moraga (2011), who show that longitudinal data allow to estimate (6) under more general distributional assumptions with the adoption of the Common Correlated Effects estimator proposed by Pesaran (2006). Bertoli and Fernández-Huertas Moraga (2011) apply their estimation approach

[16]These dummies have to be interacted with time dummies whenever the data have a longitudinal dimension.

[17]McKenzie, Theoharides, and Yang (2012) include origin-time dummies as the focus of their analysis is on the pull factors of Filipino migration, while Beine, Docquier, and Ozden (2011) include origin dummies in a cross-sectional setting to control for "the combined effect of all unobserved characteristics of the origin country" (p. 34) on the bilateral migration rate.

[18]See the discussion in Bertoli and Fernández-Huertas Moraga (2011) of the reason why the error term ε_{jk} is correlated with the regressors in (6) when the error term depends on the attractiveness of other destinations, which entails that the correction proposed by Driscoll and Kraay (1998) is not sufficient here.

[19]McKenzie, Theoharides, and Yang (2012), who estimate the determinants of Filipino migration to 107 destinations between 1998 and 2009, include both (origin)-time and (origin)-destination dummies, but their approach is still consistent with the same assumptions on the stochastic term as in Ortega and Peri (2009).

to an administrative monthly dataset, the *Estadística de Variaciones Residenciales*, which records gross migration inflows to Spain, and that is characterized by a very low incidence of zeros in the data. This feature of the data allows them to deal with the first analytical challenge discussed above, i.e. the consistency with a general micro-foundation of migration flows, while leaving the second challenge aside.

3.1.1 Interpretation of the estimated coefficients

The inability to separately identify $\boldsymbol{\beta}/\tau$ creates uncertainty about the elasticity of bilateral migration rates, both to k and to any other destination l, with respect to any of the elements of the vector $\boldsymbol{x}_{jk}$ that determines the utility associated to destination k. To understand why this occurs, consider the direct elasticity of the migration flow m_{jk} with respect to the element c of $\boldsymbol{x}_{jk}$ for destination k:[20]

$$\frac{\partial \ln(m_{jk})}{\partial \ln(x_{jkc})} = \Big[\tau(1-p_{jk}) + (1-\tau)(1-p_{j[k|b(k)]})\Big](\beta_c/\tau)x_{jkc} \tag{7}$$

where $p_{j[k|b(k)]}$ represents the probability of opting for destination k conditional upon choosing an element of the nest to which k belongs to. Similarly, the cross-elasticity for destination l is given by:

$$\frac{\partial \ln(m_{jl})}{\partial \ln(x_{jkc})} = -\Big[\tau p_{jk} + (1-\tau)p_{j[k|b(k)]}\Big](\beta_c/\tau)x_{jkc} \tag{8}$$

This cross-elasticity represents an externality. We are measuring the effect of a bilateral variable x_{jkc} on the migration flows not between j and k but between j and an alternative destination l. Whenever x_{jkc} refers to a migration policy variable, we will be quantifying the externality effect of migration policies.

Both elasticities in (7) and (8) are not a function of β_c/τ only, but they also depend on the unknown value of the dissimilarity parameter τ, which belongs to the interval $(0, 1]$. Following the approach adopted in Schmidheiny and Brulhart (2011), we can define bounds for the two elasticities, conditional upon the estimated value of β_c/τ, exploiting their monotonicity in τ. Specifically, by computing the values of (7) and (8) for τ converging to 0 and

[20] If a variable enters the vector $\boldsymbol{x}_{jk}$ after a logarithmic transformation, then x_{jkc} does not appear on the right hand side of (7)-(10).

for $\tau = 1$, we can observe that:[21]

$$\frac{\partial \ln(m_{jk})}{\partial \ln(x_{jkc})}\Big|_{\beta_c/\tau=\widehat{\beta_c/\tau}} \in \Big((1-p_{j[k|b(k)]})(\widehat{\beta_c/\tau})x_{jkc}, (1-p_{jk})(\widehat{\beta_c/\tau})x_{jkc}\Big] \tag{9}$$

Similarly, with respect to the cross-elasticity, we can define the following interval:

$$\frac{\partial \ln(m_{jl})}{\partial \ln(x_{jkc})}\Big|_{\beta_c/\tau=\widehat{\beta_c/\tau}} \in \Big(-p_{j[k|b(k)]}(\widehat{\beta_c/\tau})x_{jkc}, -p_{jk}(\widehat{\beta_c/\tau})x_{jkc}\Big] \tag{10}$$

The two intervals in (9) and (10) reflect an uncertainty about the true elasticities that cannot be narrowed down by increasing the precision in the estimate of β_c/τ, as the uncertainty arises from the impossibility of separately identifying $\boldsymbol{\beta}$ and τ.

Consider, for instance, the 0.62 estimated coefficient for networks[22] in the basic specification of the migration equation (Beine, Docquier, and Ozden, 2011, p. 37) that, as discussed above, corresponds to $\widehat{\beta_c/\tau}$, as the estimation strategy they adopted is consistent with a departure from the IIA assumption.[23] According to the data by Docquier, Lowell, and Marfouk (2009), migration flows from Mexico to the US between 1990 and 2000 represented 99.2 percent of total Mexican migration ($p_{j(k|D)t}$), and approximately 3.5 percent of the population of Mexico-born individuals (p_{jk}). From (9), these figures imply that the elasticity of Mexican migration to the US with respect to the local networks of Mexicans ranges between 0.022 and 0.615. Observe that, for a given $\widehat{\beta_c/\tau}$, introducing an assumption on τ also pins down the value of β_c. When τ converges to 0, the value of β_c also needs to be converging to 0 to keep $\widehat{\beta_c/\tau}$ unchanged, and this explains that low direct elasticity. The range of the possible elasticities that are consistent with the model is not specific to the example we chose, nor it is related to the precision with which β_c/τ is estimated. It just reflects the fact that $p_{j(k|D)} \gg p_{jk}$, which occurs whenever only a small share of the population migrates.

3.1.2 Zero flows in the data

The second difficulty related to the adoption of the logarithm of the bilateral migration rate as the left hand side relates to the fact that the incidence of zeros in migration datasets is,

[21]Without loss of generality, we ordered the two extremes of the proposed bounds under the assumption that $\widehat{\beta_c/\tau} \geq 0$.

[22]Networks are defined as the logarithm of one plus the size of the bilateral migrant stock in 1990.

[23]The estimation strategy in Beine, Docquier, and Ozden (2011) is consistent with $b(k) = D/\{j\}$ for all $k \neq j$, and we can use this relationship to simplify (9).

in general, very high. The share of zero observations stands at 9 percent in Grogger and Hanson (2011), 36 percent in Beine, Docquier, and Ozden (2011), 70 percent in Ortega and Peri (2009),[24] and up to 95 percent in Simpson and Sparber (2012). Some analyses have been carried out on the sample sample restricted to non-zero observations only (McKenzie, Theoharides, and Yang, 2012; Grogger and Hanson, 2011; Bertoli and Fernández-Huertas Moraga, 2011). Ortega and Peri (2009) have resorted to a scaled-OLS estimation, while Simpson and Sparber (2012) and Beine, Docquier, and Ozden (2011) have resorted to a threshold Tobit model and to a two-step Heckman selection control procedure respectively. The first two approaches are exposed to the criticism expressed by Santos Silva and Tenreyro (2006).[25]

The threshold Tobit model developed by Eaton and Tamura (1994) was proposed by Martin and Pham (2008) as a superior alternative to the PPML estimation favored by Santos Silva and Tenreyro (2006), but Santos Silva and Tenreyro (2011) have recently questioned the ability of this estimator to deal with a large share of zeros in the data. The Monte Carlo evidence that they provided reveals that the bias associated to the threshold Tobit estimator is large, and close to the one that characterizes the estimation on the truncated sample of positive observations.

The reliance on a two-step Heckman selection control procedure is confronted with two different sorts of difficulties, which relate to the exclusion restriction and to the consistency of the first stage estimation with the structure of the error term in (6). Specifically, finding a variable that, conditional on the other regressors, exerts a significant influence on the probability of observing a positive flow, but that is uncorrelated with the size of the flow once this is positive is a challenging task,[26] all the more so when the dataset has a longitudinal dimension.

A further challenge relates to the consistency of the assumptions on the error term that underlie the second stage equation, and the assumption of normality in the first stage equation. If the second stage equation is characterized by the presence of spatial correlation

[24]The figure goes down to 10 percent when Ortega and Peri (2009) use migrant stocks rather than flows as the dependent variable, as Grogger and Hanson (2011) did; bilateral stocks can be regarded as a proxy for the unobserved bilateral gross flows, which is the theoretically relevant measure of migration.

[25]"These procedures will generally lead to inconsistent estimators of the parameters of interest. The severity of these inconsistencies will depend on the particular characteristics of the sample and model used, but there is no reason to believe that they will be negligible" (Santos Silva and Tenreyro, 2006, p. 643).

[26]Beine, Docquier, and Ozden (2011) rely on bilateral diplomatic representation as an exclusion restriction in a cross-sectional setting.

in the residuals, then also the non-linear observations in the non-linear first stage will not be independent.[27] These arguments suggest that further concerns, beyond the doubts cast by Santos Silva and Tenreyro (2006), relate to the chance of satisfactorily dealing with the presence of zeros in the data under a general specification on the (unknown) underlying data-generation process.

3.2 Poisson estimation of bilateral migration flows

The estimation approach proposed by Santos Silva and Tenreyro (2006) is precisely meant to deal with the presence of a large share of zeros in the data, and it is gaining momentum in the international migration literature. PPML estimation performs well even when the data fail to satisfy the equidispersion property that characterizes the Poisson distribution (Santos Silva and Tenreyro, 2011),[28] and different approaches have been proposed to deal with the potential endogeneity of the regressors (Tenreyro, 2007; Terza, Basu, and Rathouz, 2008; Cameron and Trivedi, 2010).

Hence, we focus here on the consistency of the Poisson estimation with the RUM model underlying the scale of the observed bilateral migration flows. Going back to the expression for m_{jk} in (5), which we report here:

$$m_{jk} = s_j \frac{e^{\boldsymbol{x}_{jk}'\boldsymbol{\beta}/\tau}\left(\sum_{l\in b(k)} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau}\right)^{\tau-1}}{\sum_q \left(\sum_{l\in b_q} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau}\right)^{\tau}} \eta_{jk}$$

We can rewrite it more compactly as follows:

$$m_{jk} = e^{\alpha_j + \boldsymbol{x}_{jk}'\boldsymbol{\beta}/\tau + \gamma_{jb(k)}} \eta_{jk} \tag{11}$$

where:

$$\alpha_j = \ln(s_j) - \ln\left[\sum_q \left(\sum_{l\in b_q} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau}\right)^{\tau}\right]$$

and:

[27] See Fernández-Val and Vella (2011) and Arellano and Bonhomme (2011) for an overview of the challenges connected to the estimation of nonlinear panel models required for the first stage equation.

[28] This estimation technique will produce consistent estimates as long as the conditional mean is correctly specified (Gourieroux, Monfort, and Trognon, 1984).

$$\gamma_{jb(k)} = \ln\Big(\sum_{l\in b(k)} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau}\Big)^{\tau-1}$$

If one assumes, as before, that $b(k) = \{k\}$ for all $k \in D$, that is to say that the IIA assumption holds, then we can simplify the expressions for α_j and $\gamma_{jb(k)}$, as we have that $\alpha_j = \ln(s_j) - \ln\Big(\sum_{l\in D} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}}\Big)$ and $\gamma_{jb(k)} = 0$. Hence, when IIA characterizes the underlying RUM model, we can rewrite m_{jk} as follows:

$$m_{jk} = e^{\boldsymbol{x}_{jk}'\boldsymbol{\beta} + \ln(s_j) - \ln\left(\sum_{l\in D} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}}\right)}\eta_{jk} \tag{12}$$

Some key observations emerge from the inspection of (12).

First, the *scale* of the bilateral migration flow from j to k always depends on the utility associated to all possible destinations, and not only to the utility associated to the origin j and the destination k. When bilateral migration flows are estimated with PPML, the long-standing tradition of "estimating bilateral migration flows as a function of characteristics in the source and destination countries only" (Hanson, 2010, p. 4373) that characterizes the international migration literature no longer applies. An increase in the attractiveness of destination l will always reduce the scale of migration from j to k, though it leaves the migration rate from j to k unchanged if IIA holds.

Second, the adoption of the PPML estimator prevents the identification of the effect of the so-called push factors of international migration, as the deterministic component of utility at origin enters into the exponential term in (12) in a non-linear way.[29]

Third, a RUM-consistent estimation of (12) requires the inclusion of origin dummies to absorb the effect of population at origin s_j and of the attractiveness of all possible locations upon m_{jk}. The inclusion of origin dummies implies that the expected value of m_{jk} conditional upon $\boldsymbol{x}_{jk}$ and the set of dummies, is independent across all observations in the dataset, which represents a necessary condition for the estimation of the Poisson model.

[29] Observe that V_{jj} enters linearly into the exponential of the ratio of the conditional means for m_{jkt} and m_{jj}; still, the conditional mean of the ratio m_{jk}/m_{jj} never coincides with the ratio of the two conditional means (specifically, the conditional mean of m_{jk}/m_{jj} is higher than the ratio of the conditional means of m_{jk} and m_{jj}, independently on the distributional assumptions on the underlying data-generating process), and this violates the condition that is required to obtain consistent estimates with PPML (Gourieroux, Monfort, and Trognon, 1984).

Guimaraes, Figueiredo, and Woodward (2003) demonstrate that the estimation of (12) through PPML delivers the same estimate for $\boldsymbol{\beta}$ as a conditional logit model estimated on individual-level data on the same determinants of location-specific utility, as the log-likelihood functions of the two models are identical up to a constant.[30] Hence, this estimation technique is fully consistent with the underlying RUM model that describes the choice of the utility-maximizing location.

Schmidheiny and Brulhart (2011) generalize this result under the same assumptions as in Ortega and Peri (2009), so that the model to be estimated becomes:

$$m_{jk} = e^{\boldsymbol{x}_{jk}'\boldsymbol{\beta}+\ln(s_j)+\tau\ln\left(\sum_{l\in D} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau}\right)-\ln\left[e^{\boldsymbol{x}_{jj}'\boldsymbol{\beta}}+\left(\sum_{l\in D/\{j\}} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau}\right)^{\tau}\right]}\eta_{jk} \tag{13}$$

PPML estimation of (13) delivers the same estimate for $\boldsymbol{\beta}/\tau$ as the estimation of an individual-level nested logit model, with the nest structure that we just described (Schmidheiny and Brulhart, 2011). Observe that the origin fixed effects suffice to restore independence across observations both in (12) and (13), although the stochastic properties of the two underlying RUM models differ.[31] Hence, the consistency of the estimation through PPML with the underlying RUM model does not rest on the inclusion in the vector $\boldsymbol{x}_{jk}$ of all the factors that shape the utility of country k for a would-be migrant from country j.[32]

3.2.1 Interpretation of the estimated coefficients

Schmidheiny and Brulhart (2011) derive the bounds of the elasticities when one estimates (13) without any further *a priori* information on the true characteristics of the underlying RUM model. These bounds can also be derived for the more general equation (11). Specifically, we have that:

[30] Guimaraes, Figueiredo, and Woodward (2003) focus on location-decisions taken from a single origin, so that α does not vary with j; Guimaraes, Figueiredo, and Woodward (2004) show that, with multiple time periods, the inclusion of origin-time dummies suffices to restore the parallel between the conditional logit and Poisson.

[31] The key is that the term describing the expected utility from migration to any destination in the nest D does not vary across destinations, so that it is absorbed by the origin fixed effect, which is always to be included in the estimation (Guimaraes, Figueiredo, and Woodward, 2004).

[32] More specifically, RUM-consistency is ensured and the IIA assumption holds even if the vector $\boldsymbol{x}_{jk}$ does not include factors that are common to all destinations $k \in D/\{j\}$.

$$\frac{\partial \ln(m_{jk})}{\partial \ln(x_{jkc})}\Big|_{\beta_c/\tau=\widehat{\beta_c/\tau}} \in \Big((1-p_{j(k|D/\{j\})})(\widehat{\beta_c/\tau})x_{jkc}, (1-p_{jk})(\widehat{\beta_c/\tau})x_{jkc}\Big]$$

A similar expression can be derived for the cross-elasticity, which we will later use to quantify externality effects of bilateral variables:

$$\frac{\partial \ln(m_{jl})}{\partial \ln(x_{jkc})}\Big|_{\beta_c/\tau=\widehat{\beta_c/\tau}} \in \Big(-p_{j(k|D/\{j\})}(\widehat{\beta_c/\tau})x_{jkc}, -p_{jk}(\widehat{\beta_c/\tau})x_{jkc}\Big]$$

This implies that PPML estimation is characterized by the same fundamental uncertainty about the magnitude of the elasticity of migration flows as the estimation of the determinants of the bilateral migration rates, which is connected to the inability to identify the dissimilarity parameter τ. As in the example discussed above about the elasticity of Mexican migration to the US with respect to networks, when $p_{j(k|D/\{j\})} \gg p_{jk}$, the bounds defined by Schmidheiny and Brulhart (2011) for the self- and cross-elasticity will diverge from each other. PPML offers no advantage, in this respect, over the traditional approach to the estimation reviewed above.

3.2.2 Consistency with RUM under less restrictive distributional assumptions

Schmidheiny and Brulhart (2011) established the consistency of the PPML estimation with an utility-maximizing behavior of the migrants under the same assumption on the stochastic properties of the RUM model used in Ortega and Peri (2009). Here, we go one step further, showing that the same consistency characterizes the estimation of (11), which we derived from (3). We reproduce here (11):

$$m_{jk} = e^{\alpha_j + \boldsymbol{x}_{jk}'\boldsymbol{\beta}/\tau + \gamma_{jb(k)}}\eta_{jk}$$

which depends on a origin-nest specific component $\gamma_{jb(k)}$. PPML estimation requires observations to be spatially independent and, as discussed in Guimaraes, Figueiredo, and Woodward (2004), this can be achieved with the inclusion of a richer structure of dummies.[33]

[33] Notice that the presence of nest-specific unobservables does not *per se* induce spatial correlation in the residuals if these factors exert an identical influence on the would-be migrants from any country of origin, as this effect could be controlled for by the inclusion of destination dummies α_k in (11); spatial dependence arises if would-be migrants from some origins are attracted by these nest-specific unobservables, while others are repelled by the same set of factors; see also the discussion in Head, Ries, and Swenson (1995), p. 240, on this.

Specifically, the inclusion of origin-nest dummies suffices to control for $\gamma_{jb(k)}$, and restore spatial independence of the residuals.[34] This, in turn, will produce a consistent estimate of $\boldsymbol{\beta}/\tau$, which is identified only out of within-nest variation.

Such an approach requires to specify the assumptions on the nests b, and it is feasible thanks to the absence of an incidental parameter problem in the estimation of a Poisson model (Trivedi and Munkin, 2010). The estimation delivers the same estimate for $\boldsymbol{\beta}/\tau$ as the individual-level estimation of a nested logit model with location-specific regressors.

The choice of nests b is a data-dependent empirical exercise with a clear trade-off. As the number of nests used to specify equation (11) increases, the available variability that can be exploited to estimate $\boldsymbol{\beta}/\tau$ goes down. On the other hand, choosing a too parsimonious specification with few nests may not be able to fully restore the spatial independence of the residuals that is needed to be able to interpret the coefficients of the model as coming from a RUM framework. In able to assess this trade-off, the next subsection introduces a test of the spatial dependence of the residuals in equation (11).

3.2.3 Tests for spatial dependence of the residuals

The spatial independence of the migration flows from a given origin to different destinations can be assessed through tests on the residuals generated by the various specifications of our estimates. Specifically, let $\widehat{e}_{jk}$ represent the Pearson residual associated to the migration flow from the origin j to the destination k,[35] and $\widehat{\boldsymbol{e}}_k$ be the vector of Pearson residuals for destination k. If the set of fixed effects introduced among the regressors suffices to restore spatial independence, then we should have $E(\widehat{\boldsymbol{e}}_k, \widehat{\boldsymbol{e}}_l) = 0$ for $l, k \in D$, with $l \neq k$, while the presence of a nest-specific stochastic component of utility would entail that $E(\widehat{\boldsymbol{e}}_k, \widehat{\boldsymbol{e}}_l) > 0$. As the RUM model gives us an expectation on the direction of the correlation if we do not have cross-sectional independence, we can adapt a modified version of the CD test proposed by Pesaran (2004).[36] Let $\widehat{\rho}_{kl}$ denote the correlation between the vectors $\widehat{\boldsymbol{e}}_k$ and $\widehat{\boldsymbol{e}}_l$; the CD

[34] A similar use of nests can be found in the analysis of firms' location choice by Head, Ries, and Swenson (1995) and Levinson (1996); see also the other papers cited by Guimaraes, Figueiredo, and Woodward (2004).

[35] Hsiao, Pesaran, and Pick (2012) provide evidence on the reliance of the Pearson residuals when testing for cross-sectional dependence in non-linear models.

[36] "The choice of the appropriate test should be supported by *a priori* information (e.g. from economic theory) on the way statistical units may be correlated" (Moscone and Tosetti, 2009, p. 558), and this is why we are not concerned here with the fact that the CD test might fail to reject the null of cross-sectional

test statistic is given by:

$$CD = \Big(\frac{2n_o}{n_d(n_d-1)}\Big)^{1/2} \sum_{k=1}^{n_d-1} \sum_{l=k+1}^{n_d} \widehat{\rho}_{kl} \tag{14}$$

where n_o and n_d represent respectively the number of origins and destinations in the dataset. Under the null of no cross-sectional correlation in the residuals, the CD test statistic is asymptotically distributed as a standard Normal variable.

In the empirical part of the paper, we will use the CD statistic to choose a model that is parsimonious enough while being able to fully restore the spatial independence of the residuals in equation (11).

4 Data

4.1 Data sources

We draw our data from three main sources. The first one is represented by Docquier, Lowell, and Marfouk (2009), who provide information on the size of bilateral migration stocks in 31 countries of destination in 1990 and 2000. This dataset provides a proxy for the scale of bilateral gross migration flows that is represented by the variation in stocks,[37] that has been used, *inter alia*, by Beine, Docquier, and Ozden (2011). In fact, our objective is to start by replicating the results in Beine, Docquier, and Ozden (2011) so as to make our methodology directly comparable with a well known paper in the literature. Bilateral migrant stocks are defined on the basis of country of birth for all but five destinations (Germany, Hungary, Italy, Japan and Korea), which resort to citizenship to identify immigrants. This dataset is matched with the one assembled by Ozden, Parsons, Schiff, and Walmsley (2011), giving us the size of bilateral migration stocks in 1960, which will be used as an instrument for the size of networks in 1990.

With respect to bilateral visa policies, we use the dataset by Neumayer (2006), which is based on the *Travel Information Manual*, a yearly publication of the International Air

independence when the data present both patterns of positive and negative correlation (Frees, 1995).

[37]This is a common practice in the migration literature, which implies that "it is impossible to know how exactly these changes balance attrition (and whether attrition is caused by death, return migration or emigration to a third country) and new entry flows." (Docquier and Rapoport, 2012).

Transport Association, IATA. The *Travel Information Manual* contains information on all the legal requirements related to transit or non-immigration admission into all countries of the world, including visa requirements. Neumayer (2006) built a dichotomous variable signaling whether the citizens of country j are requested to have a visa for entering into country k or they benefit from a visa waiver.[38] Observe that visa policies are based on citizenship rather than on country of birth, which is the basis for most of the migration data used in our analysis; the measurement error induced by this discrepancy is likely to be negligible as citizenship and country of birth are likely to coincide for the vast majority of the population in each origin country. This dataset, which has been used also in Neumayer (2010, 2011), refers to the year 2004.[39] As we will be using the information contained in this dataset to estimate the determinants of migration flows between 1990 and 2000, this introduces an additional source of measurement error related to the changes in visa policies that might have occurred between our period of analysis and 2004, but "this measurement error is small because the number of changes to visa restrictions is likely to be very small compared to the total number of restrictions in place" (Neumayer, 2010, p. 173).[40]

We also draw on Mayer and Zignago (2011) for the time-invariant dyadic variables such as distance, common language, colonial relationship and contiguity, which can influence bilateral migration costs.

4.2 Descriptive statistics

Table 1 presents the summary statistics for the variables that will be used in the estimation below. The first panel presents the full dataset of 31-destinations-times-182-origins dyads while the second focuses on those observations for which the dependent variable takes strictly positive values. The sample size goes down from 5,611 origin-destination observations[41]

[38] Visas that need *not* to be requested before traveling are considered as visa waivers, as a visa that can be obtained upon arrival "typically does not represent any restriction at all because the procedure of getting it is extremely simple and does not involve any major check on the applicant." (Neumayer, 2010, p. 173).

[39] We directly contacted the customer service of the IATA to obtain earlier editions of the manual, but the December 2004 edition is the oldest that is currently available for sale.

[40] We can also observe that a similar measurement error occurs in Grogger and Hanson (2011), who include the bilateral visa policies in 1999 among the determinants of the size of bilateral migration stocks in 2000.

[41] The sample does not include the 31 dyads for which the origin and the destination country coincide.

to just 3,466, fully dropping three destinations: Hungary, Korea and Poland.[42] Thus, 62 percent of the observations remain for OLS regressions on the logarithm of the bilateral migration rate. The maximum of the dependent variable (3.7 million) corresponds to the Mexican migration to the US whereas the minimum (-189,660) refers to the net flow between Germany and the US. Incidentally, only 7 percent of the observations take negative values, which means that the share of strict zeros is 31 percent. The average value is less than 3,000 immigrants per origin-destination pair in the total sample and it goes up to more than 5,000 immigrants in the strictly positive sample. The standard deviations are in both cases notably larger than the means (52,910 and 66,991 respectively), pointing out to a high level of dispersion in the data.

The first independent variable in Table 1 is the size of migration networks for each origin-destination pair in the year 1990. The average in this case is over 7,000 immigrants with a maximum of 2.7 million corresponding again to the Mexican network in the US. On the lower end, up to one third of the sample corresponds to zero values in the first panel, number reduced to just 6 percent in the lower panel. Some of the regressions below also use the 1960 size of the networks. In this case, the average is lower (5,867 immigrants) although the maximum is still quite high, corresponding this time to the 2.2 million Polish-origin individuals living in Germany. The number of zeros in this variable is 35 percent in the full sample and 21 percent in the lower panel.

Next, the dummy variable representing the visa requirement to enter a given destination from a given origin has an average value of 0.69 in the full sample and a slightly lower 0.67 in the lower panel. Thus, its variability does not hinge on the inclusion of zero-flow observations in the sample. The following variable refers to the Schengen treaty. It takes value 1 when both the origin and the destination country belonged to the Schengen area at some point in the 1990s and 0 otherwise. The members of the Schengen area (nine of the 31 destination countries in this period) adopted a common visa policy toward any origin country in our sample in 2004,[43] so that the inclusion of this variable, which is introduced

[42] The size of the 1990 migrant stock are estimated rather than observed for 10 destination countries, including Hungary, Korea and Poland (Docquier, Lowell, and Marfouk, 2009, p. 317); more specifically, the size of the estimated stocks for these three destinations are lower, for all origin countries, than observed stocks in 2000.

[43] Spain granted a visa waiver to Colombians up to 2001 and to Ecuadorians up to 2003, when a visa requirement was imposed by the European Council regulation (Bertoli and Fernández-Huertas Moraga,

following the main specification in Beine, Docquier, and Ozden (2011), could, if anything, limit the ability of the models below to identify the effect of the visa variable.[44] Finally, three other classical variables from the literature are presented: colonial links, the existence of a common language and the distance in kilometers between each origin and each destination. None of the three appears very different in the two samples.

5 Estimation results

We present first the estimates of the various specifications that we run, and we then discuss the interpretation of the coefficients following the lines proposed in Section 3.

5.1 Estimates

This section presents the results from estimating several versions of the model presented in Section 2, following some of the different strategies presented in Section 3. In order to closely tie the results to the existing literature, we begin by reproducing the OLS estimation in Beine, Docquier, and Ozden (2011) in the first data column in Table 2. The specification is exactly the same as in Beine, Docquier, and Ozden (2011) but for the addition of the visa requirement variable introduced in the previous section. This specification includes both origin and destination country dummies; the inclusion of origin dummies suffices to make the estimates consistent with an underlying RUM model as in Ortega and Peri (2009), and it controls for all origin-specific push factors of bilateral migration flows. The inclusion of destination dummies absorbs destination-specific pull factors and general immigration policies as those considered by Mayda (2010). Hence, the structure of dummies included among the regressors entails that we can only identify the effects of dyadic variables, with migration networks and bilateral visa policies representing the two key variables of interest.[45]

2011).

[44]In fact, the results below are not sensitive to the exclusion of the Schengen variable.

[45]In the absence of a credible instrument for visa policies, we follow the literature's standard practice of treating the migration policies as exogenous; see Clark, Hatton, and Williamson (2007), Mayda (2010), Beine, Docquier, and Ozden (2011), Grogger and Hanson (2011) or Ortega and Peri (2009, 2012), Bertoli and Fernández-Huertas Moraga (2011) and Belot and Ederveen (2012) among others; notice that the inclusion of origin-nest dummies among the regressors entails that we will be controlling for the fact that nests of destinations might determine their visa policies toward a given origin country on the basis of the perceived

Reassuringly, this specification produces the same results as in Beine, Docquier, and Ozden (2011) for all of the variables that they also included. Distance, colonial links and common language appear as significant correlates of the log of net immigration rates. In particular, the coefficient on the log of networks in 1990 exactly coincides with that in Beine, Docquier, and Ozden (2011):[46] a highly significant 0.62. The introduction of the visa requirement variable as an additional explanatory variable does not have any effect on the rest of parameters, and the variable itself shows as non-significant.[47]

Section 3 described how the estimation of the OLS model suffers from two key limitations. The first relates to the possible inconsistency with the assumptions on the stochastic component of location-specific utility in the underlying RUM model. If the vector of regressors $\boldsymbol{x}_{jk}$, which we augmented with the inclusion of bilateral visa policies, fails to include all relevant dyadic determinants of migration,[48] then this would introduce correlation between the realizations of the stochastic component of location-specific utility. This, in turn, would give rise to multilateral resistance to migration (Bertoli and Fernández-Huertas Moraga, 2011), with the elements of $\boldsymbol{x}_{jkt}$ being correlated with the error term, and with the bilateral migration rate between j and k being still dependent on the attractiveness of destinations other than k. While in principle one could address this concern by testing whether the residuals are characterized by cross-sectional dependence, the highly unbalanced structure of the dataset, which is caused by the exclusion of observations with non-positive flows, hinders the adoption of these tests.[49]

The second key problem with the OLS specification is precisely the need to discard non-positive values, which can bias the estimated coefficients (Santos Silva and Tenreyro, 2006). This problem can be directly dealt with by using the Poisson regression model on the full

potential for illegal immigration from that country.

[46]We follow Beine, Docquier, and Ozden (2011) in adding one to the size of the 1990 migration networks so as not to discard zero observations.

[47]This result resembles the one obtained by Grogger and Hanson (2011), where the variable describing the 1999 bilateral visa policy has just a marginally significant effect on the size of the 2000 migration stock in their preferred specification.

[48]Observe that the inclusion of both origin and destination dummies implies that the only possible source of spatial dependence in the error term is related to the omission of dyadic variables; Belot and Ederveen (2012) recently pointed to the role of bilateral cultural proximity in shaping the size of intra-OECD migration flows, and migration flows could also respond to bilateral capital (Kugler and Rapoport, 2007) or trade flows.

[49]See De Hoyos and Sarafidis (2006).

sample from Table 1.[50] Specification (2) in Table 2 shows the result from running a Poisson regression on exactly the same variables as in specification (1). As discussed in Section 3, the inclusion of origin dummies is a necessary, though not sufficient, condition to restore the independence across observations (Guimaraes, Figueiredo, and Woodward, 2004), and it also implies that the consistency of the estimates does not hinge upon IIA (Schmidheiny and Brulhart, 2011).

The estimates in specifications (1) and (2) are very similar,[51] with just two minor changes. PPML estimation makes the colonial variable become insignificant whereas the Schengen variable turns marginally significant. The visa requirement variable is still insignificant in this specification.

The RUM-consistency of the Poisson estimates depends, as discussed in Section 3, on the absence of spatial dependence in the error term. The presence of cross-sectional dependence in the residuals would imply that the coefficients from specification (2) cannot be interpreted as being consistent with the underlying RUM model. In this case, they should be rather seen as the outcome of an atheoretical specification. To check whether this is the case, we computed the CD statistic for specification (2). Table 2 shows a statistic of 17.35,[52] with a corresponding 0.00 p-value that leads to the rejection of the null of spatial independence. Thus, the coefficients in specification (2) cannot be correctly interpreted as derived from a RUM model *à la* Ortega and Peri (2009), as the inclusion of origin dummies does not suffice to restore independence of the residuals across origin-destination dyads.

5.1.1 Reducing spatial dependence

The alternative approach that we adopt here is to restore spatial independence by reducing the variability in the data that is used for identification. Specifically, as discussed in Section 3.2.3, the inclusion of origin-nest dummies allows us to control for unobservable nest-specific components of location-specific utility that have a differential impact would-be migrants

[50] For the purposes of estimation, the 7 percent of negative values are set to zero, as net flows are used as a proxy for unobserved gross flows, which are always nonnegative.

[51] We report robust standard errors for specification (2), which, as demonstrated by Gourieroux, Monfort, and Trognon (1984), make the estimates from the Poisson regression consistent even when the data are not characterized by the equality between mean and variance; the test on the residuals proposed by Cameron and Trivedi (2010) reveals that the equi-dispersion property is indeed not satisfied by our model.

[52] We calculate the tests with the *xtcd* command in Stata, introduced by Eberhardt (2011).

from different countries of origin.[53] This approach is much less demanding in terms of data requirements, but it needs to specify assumptions about the composition of the nests of destinations that share unobserved components of location-specific utility.[54]

While the composition of the nests is necessarily arbitrary, its adequacy can still be measured through its ability to restore the spatial independence of the residuals of the model. There is a clear trade-off between the fineness of the nests and the loss of identification power. Coarser nests, with the unique nest of destinations *à la* Ortega and Peri (2009) representing the limiting case of coarseness, have more identification power at the expense of a greater risk of remaining spatial dependence. Finer nests, like the ones presented here, run the risk of saturating the model and losing much of the identification power in the data. In the limit, the finest partition, which is represented by single-destination nests, ensures spatial independence but delivers no identification in the cross section as they would be equivalent to origin-destination fixed effects.

This trade-off suggested us the following approach: if the CD test rejects the null of spatial independence on the basis of a specification with $m \geq 1$ nests, then we opt for a specification with $m + 1$ nests. This requires to determine the criteria that inform how we define finer nests, and we opted for geographical proximity and income per capita as the two guiding factors. We stop once the nest structure produces residuals that do not lead to the rejection, at conventional confidence levels, of the null hypothesis of spatial independence.

As the CD test conducted on the residuals from specification (2) in Table 2 where $m = 1$ rejected the null, we opted for a specification with two nests, the nest b_1 including Europe and the nest b_2 including all the other destinations. This specification reduced the CD test to 5.52, but it still leads to reject the null of spatial independence at the 1 percent confidence level. We then divided the nest b_2 into a nest b_{21} containing high-income countries (Australia, Canada, Japan, New Zealand and the United States), and into a nest b_{22} for emerging countries (Korea, Mexico, South Africa and Turkey). This specification with $m = 3$ generated a CD statistic of 3.92, still rejecting at the 1 percent confidence level. The following step was to split the nest b_{21} between a nest b_{211} for North America (Canada and the United

[53]Cultural proximity might be, as discussed above, one of these unobservables; a nest of destinations that share similar cultural traits would be attractive (unattractive) for would-be migrants from origin countries with a culture that is close (distant) from the one of that nest of destinations.

[54]See, for instance, the discussion on the composition of the nests in Head, Ries, and Swenson (1995), p. 241.

States) and a nest b_{212} for the other countries, but this only reduced the value of the CD test to 3.82 (p-value of 0.000). We then divided the large European nest b_1 between the nest b_{11} for Western European countries and a nest b_{12} for Eastern European countries. The value of the CD statistic went further down with this five-nest specification to 3.30, but it sill rejected the null at the 1 percent confidence level.[55]

Finally, our we run a six-nest specification, further dividing the Western European nest into a nest b_{111} for the EU-15 countries, and a nest b_{112} for the three members of the European Free Trade Association, namely Iceland, Norway and Switzerland. Here we stopped, as the residuals generated from this specification of the model no longer lead to a rejection of the null. Specification (3) in Table 2 reports the estimates,[56] obtained interacting the origin dummies with the nest dummies, so that the coefficients are identified only out of within-nest variability in the data.

As discussed above, this identification strategy works under the assumption that the unobserved components of location-specific utility that induce a spatial correlation in the error term are nest-specific, with the destinations belonging to any of the six nests are regarded as close substitutes by would-be migrants. Their location choices within each nest are more sensitive than the decision to migrate to variations in the attractiveness of any other destinations in the nest.

The loss of identification power is reflected in the lack of precision in the estimates for the Schengen,[57] and distance variables. On the other hand, the colonial and common language variables become marginally significant. The migration networks variable remains highly significant although the value of the coefficient falls in this specification: 0.57. This fall is

[55] Auxiliary regressions are available from the authors upon request.

[56] The origin dummies are interacted with the following six nests: b_{111} (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden and the United Kingdom), b_{112} (Iceland, Norway and Switzerland), b_{12} (Czech Republic, Hungary, Poland and Slovakia), b_{211} (Canada and the US), b_{212} (Australia, Japan and New Zealand) and b_{22} (Korea, Mexico, South Africa, and Turkey); notice that our estimation approach does *not* require that other destinations that are not included in our sample do not belong to these five nests; for instance, Romania could belong to the Eastern European nest b_{12}, or Brazil could belong to the nest b_{22} of emerging countries.

[57] With respect to the Schengen variable, observe that this is also insignificant in several specifications in Beine, Docquier, and Ozden (2011), and that the "Schengen accord has had little effect on the scale of migration among signatory countries" (Grogger and Hanson, 2011, p. 51), and "Schengen had no effect at all" (Ortega and Peri, 2009, p. 17).

what we could expect from the existence of a problem of multilateral resistance to migration that is addressed by the use of the appropriate nest structure. The reason is that a larger network from one origin to a particular destination will be typically correlated with lower networks to destinations that are perceived as substitutes. In a specification that does not control for multilateral resistance to migration, the network variable might be picking up the own larger network effect together with the other destinations lower network effects, leading to an upward bias in the coefficient, which appears to be limited in this case.

Still, the most notable change in specification (3) relates to the coefficient of the visa variable, which turns highly significant with a value of -0.667. The economic interpretation of the observed change is clear: visa policies can be coordinated at the supranational level, as it occurs within the Schengen area, so that bilateral visa policies towards any origin are closely correlated across several destinations. This, in turn, introduces a correlation between the bilateral visa policy adopted from country k upon country j, and the attractiveness of alternative destinations for would-be migrants from country j. Once we account for the attractiveness of alternative destinations through the inclusion of origin-nest dummies, bilateral visa policies become significant predictors of the scale of bilateral migration flows. This change in the estimated effect of visa policies once multilateral resistance to migration is controlled for is in line with the results found by Bertoli and Fernández-Huertas Moraga (2011) for migration to Spain.

Are the results in specification (3) preferable to those in specification (2) in terms of their RUM-consistency? They are, since the CD test performed on the residuals from specification (3) does not, by construction, reject the null of spatial independence, as the value of the statistic stands at -1.57 (p-value of 0.117).The much larger value of the log pseudo-likelihood function with respect to specification (2) also represents another reason to favor specification (3), as pointed out by Guimaraes, Figueiredo, and Woodward (2004).

5.2 Uncertainty on the elasticities

Once we have an estimation technique that is well micro-founded and thus consistent with the theory, such as the one presented in specification (3) in Table 2, our objective is to provide an economic interpretation of the estimates. However, we have seen in Section 3 how none of the presented techniques allows us to escape from a fundamental uncertainty on the calculation of the elasticities implied by the theoretical model. The reason is that Table 2 gives us estimates for $\boldsymbol{\beta}/\tau$ whereas we are unable to separately identify the elements of the vector $\boldsymbol{\beta}$ or τ. This subsection provides the empirical counterpart of 3.1.1 and 3.2.1, calculating the elasticity bounds implied by this fundamental uncertainty for the RUM model. We concentrate on the two key variables of interest, migration networks and the visa requirement, using the estimates from specification (3) in Table 2.

5.2.1 Network elasticities

Calculating the bounds of the elasticity of migration flows with respect to the size of the networks is a straightforward task. We just need to follow equations (9) and (10). Equation (9) gives us the bounds for the elasticity of migration flows between the origin j and the destination k with respect to the networks of migrants from j and residing in k. The summary statistics for the upper and lower bound of this direct elasticity can be observed in the upper panel of Table 3,[58] while each dot in Figure 1 represents the two bounds for an origin-destination dyad. Since we chose to represent the lower bounds in the horizontal axis, this implies automatically that all the observations are above the 45 degree line. The figure shows how the upper bound tends to be quite similar for most countries. The reason is that the upper bound depends on unconditional probabilities of emigration which, for most countries, counts for a fairly small share of the total population. On the contrary, the lower bound depends on conditional probabilities of migration within the nest which, for many countries, are bound to be quite substantial (e.g., Mexican migration to the US). All in all, Table 3 shows that the average upper bound is 0.57, with this figure coinciding with the estimated coefficient. On the other extreme, under a high correlation in the unobserved component of utility between destinations of the same nest, the average lower bound for the elasticity of migration with respect to networks would stand at 0.46.

The heterogeneity of the results does not stop at the direct elasticities. Our simple RUM migration model also has implications for the cross-elasticity. Equation (10) generates the bounds for the cross-elasticity that has typically been absent from the literature:[59] the elasticity of the migration flow from the origin j to the destination l with respect to the migration networks of j in another destination k. The upper panel of Table 3 presents the averages of the upper and lower bound for this cross-elasticity, while Figure 2 represents the

[58] The averages of the various bounds are virtually unaffected if we resort to an unweighted averaging.

[59] Bertoli, Fernández-Huertas Moraga, and Ortega (2012) represents an exception.

clouds of dyad-specific cross-elasticities.[60] The average upper bound for the cross-elasticity is almost zero.[61] As for the lower bound, the furthest away from the IIA assumption, the average cross-elasticity is -0.11.[62]

5.2.2 Visa effects

Differently from networks, the visa variable is dichotomous, so that we cannot use directly the elasticity formulas from the previous subsection. We derive the appropriate expressions in the Appendix. The bottom panel of Table 3 presents the averages of these effects implied by the point estimates taken out of specification (3) in Table 2.

The most remarkable aspect that deserves to be commented about the direct and indirect effects of visas is their magnitude. The average bounds mean that we can expect the imposition of a visa requirement by country k on country j to correlate with a decrease of 40 to 47 percent of the level of migration flows from j to k with respect to the level that prevails when a visa waiver is applied. We can recall from the Introduction that Pritchett (2006) argued that host-country policies could be decreasing migration flows by a factor of two to five; Table 3 shows is that visas might be a big part of that "cliff at the border", being able to almost halve migration flows by themselves.

As it was the case with network elasticities, there is a great deal of heterogeneity in the visa effects. The full extent of this heterogeneity can be observed in Figure 3, which represents the whole range of visa effects calculated for each origin-destination pair. The concentration of points in the lower part of the triangle explains the relatively high level of the visa effect bounds (in absolute value).

The requirement of a visa from country k to the citizens of country j also has effects on the migration flows going to alternative destinations, that is, it creates an externality. The bottom panel of Table 3 presents the average values that quantify this externality whereas

[60]Observe that (10) does not vary with l, so that we have the same number of direct and cross-elasticities.

[61]Remember that the upper bound corresponds to the satisfaction of the IIA assumption so this would imply an exact zero if we were looking at the cross-elasticity of migration rates instead of that of flows.

[62]Notice that, logically, the instances of very large (in absolute value) lower bound cross-elasticities correspond to instances of very low lower bound direct elasticities, as the difference between (7) and (8) is independent from τ. For instance, the lowest upper bounds in Figures 1 and 2 correspond both to the Grenada-US dyad, and the difference between the upper bounds for any pair of points that correspond to any origin-destination dyads in the two figures is always 0.567, which corresponds to the estimated coefficient for networks in Table 2.

Figure 4 represents all of the visa cross-effects bounds for each origin-destination dyad. As in the previous section, the cross-effects are the inverse image of the direct effects. The magnitude of the average bounds ranges between 3 and 17 percent, describing the size of the increase in migration flows from j to l generated by the imposition of a visa requirement by a third country k upon the citizens of j. To our knowledge, these calculated bounds represent the first measure of the possible magnitude of migration policy externalities, that is, the effect of the migration policy of one destination on the migration flows going to another destination. The implication is that countries whose visa policies may have a small effect on the migration flows going out of a particular country may, on the contrary, generate large effects on the migration flows from that particular country to an alternative destination.[63]

For instance, consider Canada, which received little more than 12,000 migrants from Mexico; our estimates suggest that this bilateral flow is highly sensitive to the policies adopted in the US, which represent the largest destination for Mexican migrants. The estimated cross-effects of the US visa policy on Mexicans upon the migration flow from Mexico to Canada range between 90 and 91 percent of the actual flow. This figure is much larger than the direct effect of the Canadian visa policy toward Mexicans, which is estimated at -48 percent: hence, the flow of Mexicans to Canada would respond less to a change in the Canadian visa policy than to a change in the US visa policy toward Mexico.

5.3 Robustness

This subsection presents three types of robustness analysis on the main results presented in specification (3) of Table 2. First, we re-estimate the models with different samples. Second, we redefine the dependent variable in order to study low and high-skill migration. Then, we discuss the potential endogeneity problem related to the inclusion of the networks variable.

5.3.1 Different samples

The results from Table 2 do not depend on the particular coverage of the dataset described in Table 1. Table 4 reruns specifications (2) and (3) from Table 2 while restricting the sample in two different ways: by population size and by income level.

In terms of population size, the objective of the exercise is to guarantee that the main results are not driven by the inclusion of very small origin countries in the sample. To this

[63] Bertoli, Fernández-Huertas Moraga, and Ortega (2012) make a similar point for the effect of income.

end, we drop observations with origin countries whose population is lower than one million inhabitants in 1990. With this, the sample size goes down from 5,611 to 4,497 dyads but the main results are virtually unaffected, as it can be observed in specifications (1) and (2) from Table 4. Specification (1) does not include origin-nest fixed effects and the Pesaran CD test shows that the residuals could be spatially correlated. The statistic is 15.55 (p-value of 0.000). The appropriate structure of the residuals is obtained in specification (2), where the CD statistic is -1.89 (p-value of 0.059), where, if anything, we can observe a larger coefficient for the visa requirement than that presented in Table 2.

We can also restrict the sample by income level, so that we focus more particularly in South-North migration. We do this in specifications (3) and (4) by dropping high-income OECD countries from the set of origins.[64] We are then left with 4,708 observations. Again, we reject the spatial independence of the residuals in specification (3) where we do not include origin-nest fixed effects. In specification (4), where we include them, the value of the CD statistic is -1.55 (p-value of 0.121) so that we can be confident that we have been able to restore spatial independence and we can interpret the results as coming from a RUM model. In this case, the coefficient on the networks variable is slightly higher while the coefficient on the visa variable is lower than in the baseline specification from Table 2. Still, these differences are not statistically significant. We can also observe that the distance variable regains significance in this smaller sample.

5.3.2 Estimation by skill levels

In this part, we take advantage from the fact that the dataset by Docquier, Lowell, and Marfouk (2009) allows us to compute migration flows (and rates) by skill level. We define the tertiary educated in their dataset as high-skill whereas the primary and secondary educated are grouped together as low-skill. Table 5 re-estimates the model with and without origin-nest fixed effects for both high- and low-skill versions of the dependent variable.

Starting with high-skill migration, specifications (1) and (2) confirm that the visa requirement variable only becomes significant once the opportunities to migrate to alternative destinations, multilateral resistance to migration, are taken into account. As before, we can disregard specification (1) on the basis of the CD test, clearly rejected with a statistic of

[64] This specification omits the Schengen variable, as the restriction of the sample leaves us with no variability in the data to identify its effect.

19.91. In this case, specification (2) is on the verge of rejecting the null (p-value of 0.050) but we can still have some confidence that this specification has less problems of spatial dependence than the first one. It must be noted that both the network and the visa variable have lower coefficients in absolute levels than the baseline specification although the differences are not statistically significant.

When we turn to specifications (3) and (4) in Table 5, we are focusing on low-skill migration. Again, specification (3), without origin-nest fixed effects, has problems of spatial dependence since the CD test rejects the null with a value of 9.40. Specification (4) does not have this problem since the statistic is -0.61 and we cannot reject the spatial independence of the residuals (p-value of 0.543). The bias induced by multilateral resistance to migration on low-skill flows seems to be of the same nature as the one we observed in the baseline: lower coefficient on networks and larger on the visa requirement variable in absolute levels. However, it is interesting to note that the absolute values of both coefficients are larger than those observed for high-skill migration flows. This is consistent with the idea that low-skill migration flows might be more sensitive to changes in the costs of migration than high-skill migration flows.

It is useful to compare the different sensitivity of migration flows by skill level to networks and migration policies by looking again at the implied elasticities according to a RUM model. This is done in Table 6, based in specifications (2) and (4) from Table 5. We can see that the bounds for the direct elasticity of bilateral migration flows with respect to bilateral networks differ between 0.40 to 0.50 for high-skill flows and between 0.50 and 0.61 for low-skill flows. For the cross elasticities, the differences are smaller: between -0.10 and 0.00 for high-skill flows and between -0.11 and 0.00 for low-skill flows. In the case of the direct effects of bilateral visas on migration flows, the bounds are between -0.42 and -0.35 for high-skill flows but they go up to between -0.50 and -0.42 for low-skill flows. Correspondingly, the externality effect of the visa requirement is also larger for low-skill flows, between 3 and 19 percent, compared with the interval for high-skill flows, between 2 and 14 percent.

5.3.3 Endogeneity of networks

An additional concern with the estimation of the determinants of migration flows is that of the endogeneity of migration networks. Factors that generated the networks up to 1990, such as migration flows between 1980 and 1990, are likely to be correlated with the determinants of 1990-2000 migration flows. To address this concern, Beine, Docquier, and Ozden (2011) applied two-stage least squares by instrumenting the size of migration networks in 1990 with old bilateral guest worker agreements and different proxies for the networks in 1960, which they did not observe. We have the advantage that a new dataset, created by Ozden, Parsons, Schiff, and Walmsley (2011), has come out, which provides us with more precise estimates of the size of the networks in 1960. Thus, we use the networks in 1960 as an instrument for the networks in 1990. We resort to two-stage residual inclusion,[65] given that the Poisson model is non-linear and the 2SLS estimates would be in general inconsistent, as demonstrated by Terza, Basu, and Rathouz (2008).

The results are presented in Table 7 for all the correct specifications discussed in the previous subsections.The first stage is very strong, with the size of networks in 1960 having substantial explanatory power over the 1990 variable (the correlation between the two variables is 0.76 in the full sample). Specification (1) reproduces the baseline correct specification (3) from Table 2. The main significant change is the notable increase in the coefficient of migration networks, which suggests that the previous estimate was downward biased. The new coefficient is a strongly significant 0.77, which coincides with the result in Beine, Docquier, and Ozden (2011). A possible interpretation of the direction of the bias relates to return migration: a larger network can be associated with a larger scale of return migration, which influences a dependent variable that captures net rather than gross migration flows. As for the visa, the coefficient remains negative and significant at the 95 percent confidence level: -0.62.

The rest of specifications are shown for robustness purposes and carry exactly the same message: there is some downward bias on the networks coefficient in the baseline specification whereas the visa coefficient is virtually unaffected. None of the specifications rejects the null of spatial independence in the residuals.

6 Concluding remarks

The migration of people across borders can be severely limited by the policies adopted at destination. Our paper provides a contribution to the understanding of the influence exerted

[65]Recent applications of the two-stage residual inclusion estimator within the migration literature can be found in Beine, Lodigiani, and Vermeulen (2012), Marchetta (2012) and Bertoli and Marchetta (2012).

by bilateral visa policies on international migration flows, which can be identified only when the confounding effect represented by the opportunities to migrate to alternative destinations, multilateral resistance to migration, is adequately controlled for. The prevailing visa regime significantly contributes to determine the height of the "cliff at the border" (Pritchett, 2009), and a change in the requirements for non-immigrant admission can influence the scale of migration flows directed both to the country implementing the policy change and to other destinations.

The estimation of the determinants of international migration on aggregate data does not allow us to recover the structural parameters of the underlying theoretical model, and this creates an unavoidable uncertainty on the estimated direct and indirect effect of the visa policy. Our estimates entail that, on average, the introduction of a visa requirement reduces direct bilateral flows between 40 and 47 percent, while increasing the flows toward other destinations between 3 and 17 percent. The uncertainty on the true size of the effect notwithstanding, these figures are strongly suggestive of the relevance of the legal framework for non-immigrant admission in shaping the scale and direction of international migration flows.

These results confirm and extend the findings in Bertoli and Fernández-Huertas Moraga (2011), and are based on an estimation technique with minimal data requirements, which are well-suited for most of the existing international migration datasets. Regrettably, a binding constraint upon further applications and extensions of the proposed econometric approach is currently represented by the scarcity of longitudinal data on bilateral immigration policies. The International Migration Policy and Law Analysis, IMPALA, database will fill this gap, allowing to identify the effects of the "cliff at the border" upon migration flows out of changes in bilateral policies over time.[66]

References

Arellano, M., and S. Bonhomme (2011): "Nonlinear Panel Data Analysis," *Annual Review of Economics*, 3(1), 395–424.

[66] "The IMPALA Database is a collaborative project, bringing together social science and legal researchers from Harvard University, the University of Luxembourg, the University of Amsterdam, the London School of Economics, and the University of Sydney" (source: http://projects.iq.harvard.edu/impala, accessed on March 13, 2012).

Beine, M., F. Docquier, and C. Ozden (2011): "Diasporas," *Journal of Development Economics*, 95(1), 30–41.

Beine, M., E. Lodigiani, and R. Vermeulen (2012): "Remittances and financial openness," *Regional Science and Urban Economics*, forthcoming.

Beine, M., R. Noel, and L. Ragot (2011): "The determinants of international mobility of students," paper presented at the Fourth World Bank-AFD International Conference on "Migration and Development", Harvard University, June 10-11.

Beine, M., and C. Parsons (2012): "Climatic factors as determinants of International Migration," CESifo Working Paper No. 3747, Munich.

Belot, M., and S. Ederveen (2012): "Cultural barriers in migration between OECD countries," *Journal of Population Economics*, 25(3), 1077–1105.

Belot, M., and T. Hatton (2012): "Skill Selection and Immigration in OECD Countries," *Scandinavian Journal of Economics*, forthcoming.

Bertoli, S., and J. Fernández-Huertas Moraga (2011): "Multilateral Resistance to Migration," IZA Discussion Paper No. 5958, Bonn.

Bertoli, S., J. Fernández-Huertas Moraga, and F. Ortega (2011): "Immigration Policies and the Ecuadorian Exodus," *World Bank Economic Review*, 25(1), 57–76.

——— (2012): "Crossing the Border: Self-Selection, Earnings and Individual Migration Decisions," *Journal of Development Economics*, forthcoming.

Bertoli, S., and F. Marchetta (2012): "Bringing It All Back Home: Return migration and fertility choices," CERDI Working Paper No. 2012.01, Clermont-Ferrand.

Cameron, A., and P. Trivedi (2010): *Microeconometrics Using Stata*. College Station: Stata Press, revised edition.

Cardell, N. S. (1997): "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity," *Econometric Theory*, 13(2), 185–213.

Clark, X., T. Hatton, and J. Williamson (2007): “Explaining U.S. immigration, 1971-1998,” *Review of Economics and Statistics*, 89(2), 359–373.

Clemens, M. (2011): “Economics and Emigration: Trillion-Dollar Bills on the Sidewalk?,” *The Journal of Economic Perspectives*, 25(3), 83–106.

De Hoyos, R., and V. Sarafidis (2006): “Testing for cross-sectional dependence in panel-data models,” *The Stata Journal*, 6(4), 482–496.

Docquier, F., B. L. Lowell, and A. Marfouk (2009): “A Gendered Assessment of Highly Skilled Emigration,” *Population and Development Review*, 35(2), 297–321.

Docquier, F., and H. Rapoport (2012): “Globalization, brain drain and development,” *Journal of Economic Literature*, forthcoming.

Driscoll, J. C., and A. C. Kraay (1998): “Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data,” *Review of Economics and Statistics*, 80(4), 549–560.

Eaton, J., and A. Tamura (1994): “Bilateralism and Regionalism in Japanese and U.S. Trade and Direct Foreign Investment Patterns,” *Journal of the Japanese and International Economies*, 8(4), 478–510.

Eberhardt, M. (2011): “XTCD: Stata module to investigate Variable/Residual Cross-Section Dependence,” Statistical Software Components, Boston College Department of Economics.

Egger, P., and D. M. Radulescu (2009): “The Influence of Labour Taxes on the Migration of Skilled Workers,” *World Economy*, 32(9), 1365–1379.

Fernández-Val, I., and F. Vella (2011): “Bias corrections for two-step fixed effects panel data estimators,” *Journal of Econometrics*, 163(2), 144–162.

Flowerdew, R., and M. Aitkin (1982): “A method of fitting the gravity model based on the Poisson distribution,” *Journal of Regional Science*, 22(2), 191–202.

Frees, E. W. (1995): “Assessing cross-sectional correlation in panel data,” *Journal of Econometrics*, 69(2), 393–414.

Gourieroux, C., A. Monfort, and A. Trognon (1984): "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 52(3), 701–720.

Grogger, J., and G. H. Hanson (2011): "Income maximization and the selection and sorting of international migrants," *Journal of Development Economics*, 95(1), 42–57.

Guimaraes, P., O. Figueiredo, and D. Woodward (2003): "A Tractable Approach to the Firm Location Decision Problem," *Review of Economics and Statistics*, 85(1), 201–204.

——— (2004): "Industrial Location Modeling: Extending the Random Utility Framework," *Journal of Regional Science*, 44(1), 1–20.

Hanson, G. (2010): "International Migration and the Developing World," in *Handbook of Development Economics*, ed. by D. Rodrik, and M. Rosenzweig, pp. 4363–4414. Amsterdam: North-Holland, Volume 5.

Head, K., J. Ries, and D. Swenson (1995): "Agglomeration benefits and location choice: Evidence from Japanese manufacturing investments in the United States," *Journal of International Economics*, 38(3-4), 223–247.

Hsiao, C., M. H. Pesaran, and A. Pick (2012): "Diagnostic Tests of Cross-section Independence for Limited Dependent Variable Panel Data Models," *Oxford Bulletin of Economics and Statistics*, 74(2), 253–277.

Kugler, M., and H. Rapoport (2007): "International labor and capital flows: Complements or substitutes?," *Economics Letters*, 94(2), 155–162.

Levinson, A. (1996): "Environmental regulations and manufacturers' location choices: Evidence from the Census of Manufactures," *Journal of Public Economics*, 62(1-2), 5–29.

Lewer, J. J., and H. V. den Berg (2008): "A gravity model of immigration," *Economics Letters*, 99(1), 164–167.

Llull, J. (2011): "Understanding International Migration: Evidence from a New Dataset of Bilateral Stocks (1960-2000)," mimeo, Universitat Autónoma de Barcelona.

Marchetta, F. (2012): "Return migration and the survival of entrepreneurial activities in Egypt," *World Development*, 40(10), 1999–2013.

Martin, W., and C. S. Pham (2008): "Estimating the Gravity Model When Zero Trade Flows are Frequent," Economic Series Working Paper No. 03, Deakin University.

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

Mayer, T., and S. Zignago (2011): "Notes on CEPII's distances measures: The GeoDist database," CEPII Working Paper No. 2011-25, Paris.

McFadden, D. (1974): "Conditional logit analysis of qualitative choice behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, pp. 105–142. New York: Academic Press.

——— (1978): "Modeling the Choice of Residential Location," in *Spatial interaction theory and planning models*, ed. by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, pp. 75–96. Amsterdam: North-Holland.

McKenzie, D., C. Theoharides, and D. Yang (2012): "Distortions in the International Migrant Labor Market: Evidence from Filipino Migration and Wage Responses to Destination Country Economic Shocks," CReAM Discussion Paper Series 09/12, London.

McKenzie, D. J., and H. Rapoport (2010): "Self-selection patterns in Mexico-U.S. migration: The role of migration networks," *Review of Economics and Statistics*, 92(4), 811–821.

Moscone, F., and E. Tosetti (2009): "A review and comparison of tests of cross-section independence in panels," *Journal of Economic Surveys*, 23(3), 528–561.

Neumayer, E. (2006): "Unequal access to foreign spaces: how states use visa restrictions to regulate mobility in a globalized world," *Transactions of the Institute of British Geographers*, 31(1), 72–84.

——— (2010): "Visa Restrictions and Bilateral Travel," *The Professional Geographer*, 62(2), 171–181.

——— (2011): "On the detrimental impact of visa restrictions on bilateral trade and foreign direct investment," *Applied Geography*, 31(3), 901 – 907.

Ortega, F., and G. Peri (2009): "The Causes and Effects of International Migrations: Evidence from OECD Countries 1980-2005," NBER Working Paper No. 14883, National Bureau of Economic Research, Cambridge MA.

——— (2012): "The Role of Income and Immigration Policies in Attracting International Migrants," *Migration Studies*, forthcoming.

Ozden, C., C. R. Parsons, M. Schiff, and T. L. Walmsley (2011): "Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960-2000," *World Bank Economic Review*, 25(1), 12–56.

Papola, A. (2004): "Some developments on the cross-nested logit model," *Transportation Research Part B: Methodological*, 38(9), 833–851.

Pesaran, M. H. (2004): "General Diagnostic Tests for Cross Section Dependence in Panels," IZA Discussion Paper No. 1204, Bonn.

——— (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74(4), 967–1012.

Pritchett, L. (2006): *Let Their People Come*. Washington: Center for Global Development.

——— (2009): "The Cliff at the Border," in *Equity and Growth in a Globalizing World*, ed. by R. Kanbur, and M. Spence, pp. 263–286. Washington: The World Bank and the Commission on Growth and Development.

Santos Silva, J. M. C., and S. Tenreyro (2006): "The Log of Gravity," *Review of Economics and Statistics*, 88(4), 641–658.

——— (2011): "Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator," *Economics Letters*, 112(2), 220–222.

Schmidheiny, K., and M. Brulhart (2011): "On the equivalence of location choice models: Conditional logit, nested logit and Poisson," *Journal of Urban Economics*, 69(2), 214–222.

Simpson, N. B., and C. Sparber (2012): "The Short- and Long-Run Determinants of Unskilled Immigration into U.S. States," *Southern Economic Journal*, forthcoming.

Tenreyro, S. (2007): "On the trade impact of nominal exchange rate volatility," *Journal of Development Economics*, 82(2), 485 – 508.

Terza, J., A. Basu, and P. Rathouz (2008): "A two stage residual inclusion estimation: addressing endogeneity in health econometric modeling," *Journal of Health Economics*, 27(3), 531–543.

Train, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press.

Trivedi, P., and M. Munkin (2010): "Recent Developments in Cross Section and Panel Count Models," in *Handbook of Empirical Economics and Finance*, ed. by A. Ullah, and D. Giles, pp. 87–131. Oxford: Taylor and Francis.

UN Population Division (2008): *Trends in International Migrant Stock: The 2008 Revision*. United Nations database, POP/DB/MIG/Stock/Rev.2008.

US General Accounting Office (2004): "Overstay Tracking: A Key Component of Homeland Security and a Layered Defense," Report to the Chairman, Committee on the Judiciary, House of Representatives, Washington.

Wen, C.-H., and F. S. Koppelman (2001): "The generalized nested logit model," *Transportation Research Part B*, 35(7), 627–641.

A Effects of visa requirements on migration flows

Let v_{kjk} be the effect of the introduction of a visa requirement by country k upon the citizens of country j upon the bilateral migration flow from j to k, defined as the variation in the size of the bilateral migration flow with respect to the size of the flow under a visa waiver:

$$v_{jjk} = \frac{m_{jk}|_{x_{1jk}=1}}{m_{jk}|_{x_{1jk}=0}} - 1 = \frac{p_{jk}|_{x_{1jk}=1}}{p_{jk}|_{x_{1jk}=0}} - 1$$

Without loss of generality, we denote the visa variable with x_{1jk}. v_{kjk} represents the size of the bilateral migration flow when a visa requirement is imposed as a share of the corresponding flow when citizens from j benefit from a visa waiver granted by country k. Let $\boldsymbol{x}^e$ represent the vector of determinants of location-specific utility excluding x_1; then, from (4), we have that:

$$v_{kjk} = e^{\beta_1/\tau}\Big[\frac{\sum_{l\in b(k),l\neq k} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau} + e^{\boldsymbol{x}^e_{jk}{}'\boldsymbol{\beta}/\tau}}{\sum_{l\in b(k),l\neq k} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau} + e^{\boldsymbol{x}^e_{jk}{}'\boldsymbol{\beta}/\tau}e^{\beta_1/\tau}}\Big]^{1-\tau}\Big[\frac{\sum_{l\in D_j,l\neq k} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau} + e^{\boldsymbol{x}^e_{jk}{}'\boldsymbol{\beta}/\tau}}{\sum_{l\in D_j,l\neq k} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau} + e^{\boldsymbol{x}^e_{jk}{}'\boldsymbol{\beta}/\tau}e^{\beta_1/\tau}}\Big]^{\tau} - 1$$

v_{kjk} depends on a weighted geometric average of 1 plus the values of the effects corresponding to $\tau = 0$ and $\tau = 1$, which is monotonic in τ for any given estimate of the vector $\boldsymbol{\beta}/\tau$. If $\beta_1 < 0$, then $v_{kjk} \in (-1, 0)$. When $\tau = 0$, v_{kjk} becomes:

$$v_{kjk}|_{\tau=0} = e^{\beta_1/\tau}\frac{\sum_{l\in b(k),l\neq k} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau} + e^{\boldsymbol{x}^e_{jk}{}'\boldsymbol{\beta}/\tau}}{\sum_{l\in b(k),l\neq k} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau} + e^{\boldsymbol{x}^e_{jk}{}'\boldsymbol{\beta}/\tau}e^{\beta_1/\tau}} - 1 \tag{15}$$

When $\tau = 1$, this becomes:

$$v_{kjk}|_{\tau=1} = e^{\beta_1/\tau}\frac{\sum_{l\in D_j,l\neq k} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau} + e^{\boldsymbol{x}^e_{jk}{}'\boldsymbol{\beta}/\tau}}{\sum_{l\in D_j,l\neq k} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau} + e^{\boldsymbol{x}^e_{jk}{}'\boldsymbol{\beta}/\tau}e^{\beta_1/\tau}} - 1 \tag{16}$$

If the coefficient β_1 associated to the visa requirement is negative, then the effect is larger in magnitude when $\tau = 0$ (upper bound) than when $\tau = 1$ (lower bound).

In addition to the direct effects, the calculation of the visa cross-effects is straightforward, and we report here bounds associated to v_{kjl}, which measures the influence of the visa policy adopted by country k upon country j on migration flows from j to $l \neq k$:

$$v_{kjl} \in \Big[\frac{\sum_{l\in D_j,l\neq k} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau} + e^{\boldsymbol{x}^e_{jk}{}'\boldsymbol{\beta}/\tau}}{\sum_{l\in D_j,l\neq k} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau} + e^{\boldsymbol{x}^e_{jk}{}'\boldsymbol{\beta}/\tau}e^{\beta_1/\tau}} - 1, \frac{\sum_{l\in b(k),l\neq k} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau} + e^{\boldsymbol{x}^e_{jk}{}'\boldsymbol{\beta}/\tau}}{\sum_{l\in b(k),l\neq k} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}/\tau} + e^{\boldsymbol{x}^e_{jk}{}'\boldsymbol{\beta}/\tau}e^{\beta_1/\tau}} - 1\Big) \tag{17}$$

Table 1: Descriptive statistics

	Full sample (5,611 observations)				
	mean	st. dev.	min	max	zeros
Immigration flows, 1990-2000	2,905	52,910	-189,660	3,718,828	0.31
Migration networks in 1990	7,213	55,022	0	2,655,997	0.33
Migration networks in 1960	5,867	55,648	0	2,226,485	0.35
Visa requirement	0.69	0.46	0	1	
Schengen countries during the 1990s	0.01	0.11	0	1	
Colonial links	0.03	0.18	0	1	
Common language	0.11	0.31	0	1	
Distance (km.)	7,212	4,297	59.62	19,586.18	
	Positive net flows (3,466 observations)				
	mean	st. dev.	min	max	zeros
Immigration flows, 1990-2000	5,173	66,991	1	3,718,828	0.00
Migration networks in 1990	8,057	59,421	0	2,655,997	0.06
Migration networks in 1960	5,112	51,212	0	2,226,485	0.21
Visa requirement	0.67	0.47	0	1	
Schengen countries during the 1990s	0.02	0.13	0	1	
Colonial links	0.04	0.20	0	1	
Common language	0.14	0.35	0	1	
Distance (km.)	6,690	4,309	60.00	19,586.18	

Sources: Authors' elaboration on Docquier, Lowell, and Marfouk (2009) for net flows and migration networks in 1990; Ozden, Parsons, Schiff, and Walmsley (2011) for migration networks in 1960; Neumayer (2006) for the visa requirement, and Mayer and Zignago (2011) for the rest of the variables.

Table 2: Determinants of migration flows (1990-2000)

Specification	(1)	(2)	(3)
Dependent variable	ln(flow)	flow	flow
Model	OLS	PPML	PPML
ln(networks+1)	0.621***	0.658***	0.567***
	[0.018]	[0.042]	[0.049]
Visa requirement	-0.051	0.017	-0.667***
	[0.106]	[0.161]	[0.215]
Schengen	0.278	0.651*	0.034
	[0.179]	[0.381]	[0.235]
Colony	0.313**	-0.290	0.451*
	[0.137]	[0.217]	[0.256]
Common language	0.420***	0.333**	0.302*
	[0.076]	[0.130]	[0.161]
ln(distance)	-0.396***	-0.382***	-0.121
	[0.046]	[0.098]	[0.116]
Destination fixed effects	Yes	Yes	Yes
Origin fixed effects	Yes	Yes	Yes
Origin*nest fixed effects	No	No	Yes
Observations	3,466	5,611	5,611
Adjusted (pseudo) R^2	0.867	0.988	0.996
Log pseudo-likelihood	-	-4,294,695	-2,213,844
Pesaran (2004) CD test	-	17.35	-1.57
p-value	-	0.000	0.117

Note: standard errors in brackets; *** significant at the 99 percent level, ** significant at the 95 percent level, * significant at the 90 percent level. The dependent variable in specifications (2)-(3) is equal to the maximum between the net flow and zero; standard errors are robust in specifications (1) to (3).

Table 3: Direct and indirect elasticities of networks and visa

Bound	lower	upper
	Networks	
Direct effect	0.459	0.567
	(0.156)	(0.002)
Indirect effect	-0.108	0.000
	(0.156)	(0.002)
	Visa	
Direct effect	-0.473	-0.399
	(0.045)	(0.131)
Indirect effect	0.028	0.169
	(0.088)	(0.245)

Note: standard deviations in parentheses. The bounds correspond to averages, weighted by population at origin, over equations (9), (10) and (15)-(17) based on the estimates in specification (3) in Table 2.

Table 4: Determinants of migration flows (1990-2000), different samples

	Population size		*Developing countries*	
Specification	(1)	(2)	(3)	(4)
Dependent variable	flow	flow	flow	flow
Model	PPML	PPML	PPML	PPML
ln(networks+1)	0.654***	0.569***	0.662***	0.609***
	[0.043]	[0.051]	[0.057]	[0.065]
Visa requirement	0.016	-0.716***	0.0004	-0.358**
	[0.166]	[0.243]	[0.178]	[0.163]
Schengen	0.671*	0.041	-	-
	[0.381]	[0.236]	-	-
Colony	-0.282**	0.462*	-0.005	0.445
	[0.220]	[0.258]	[0.213]	[0.341]
Common language	0.320**	0.287*	0.574***	0.261
	[0.133]	[0.162]	[0.143]	[0.210]
ln(distance)	-0.381***	-0.116	-0.612***	-0.674***
	[0.100]	[0.116]	[0.157]	[0.186]
Destination fixed effects	Yes	Yes	Yes	Yes
Origin fixed effects	Yes	Yes	Yes	Yes
Origin*nest fixed effects	No	Yes	No	Yes
Observations	4,497	4,497	4,708	4,708
Adjusted (pseudo) R^2	0.988	0.996	0.992	0.997
Log pseudo-likelihood	-4,236,609	-2,181,524	-2,363,592	-1,442,418
Pesaran (2004) CD test	15.55	-1.89	8.49	-1.55
p-value	0.000	0.059	0.000	0.121

Note: standard errors in brackets; *** significant at the 99 percent level, ** significant at the 95 percent level, * significant at the 90 percent level. Specifications (1) and (2) are estimated on a sample restricted to origin countries with a population of at least 1 million; specifications (3) and (4) are estimated on a sample that excludes high-income OECD origin countries. The dependent variable is equal to the maximum between the net flow and zero; standard errors are robust.

Table 5: Determinants of migration flows by skill level (1990-2000)

	High-skill		*Low-skill*	
Specification	(1)	(2)	(3)	(4)
Dependent variable	flow	flow	flow	flow
Model	PPML	PPML	PPML	PPML
ln(networks+1)	0.615***	0.496***	0.703***	0.608***
	[0.038]	[0.037]	[0.052]	[0.056]
Visa requirement	-0.073	-0.559***	0.110	-0.718***
	[0.131]	[0.238]	[0.213]	[0.268]
Schengen	0.629**	-0.305*	0.960*	0.879*
	[0.281]	[0.178]	[0.565]	[0.467]
Colony	-0.152	0.327	-0.238	0.652**
	[0.162]	[0.221]	[0.260]	[0.279]
Common language	0.548***	0.619***	0.084	0.048
	[0.114]	[0.131]	[0.158]	[0.207]
ln(distance)	-0.175*	-0.220**	-0.470***	-0.053
	[0.092]	[0.088]	[0.120]	[0.131]
Destination fixed effects	Yes	Yes	Yes	Yes
Origin fixed effects	Yes	Yes	Yes	Yes
Origin*nest fixed effects	No	Yes	No	Yes
Observations	5,611	5,611	5,611	5,611
Adjusted (pseudo) R^2	0.934	0.983	0.993	0.997
Log pseudo-likelihood	-1,560,807	-690,849	-3,189,084	-1,720,273
Pesaran (2004) CD test	19.91	-1.96	9.40	-0.61
p-value	0.000	0.050	0.000	0.543

Note: standard errors in brackets; *** significant at the 99 percent level, ** significant at the 95 percent level, * significant at the 90 percent level. The dependent variable in specifications (1) and (2) refers to migration flows that are tertiary educated in Docquier, Lowell, and Marfouk (2009); specifications (3) and (4) refer to migration flows that are primary and secondary educated. The dependent variable is equal to the maximum between the net flow and zero; standard errors are robust.

Table 6: Direct and indirect elasticities of networks and visa by skill level

Flow	*High-skill*		*Low-skill*	
Bound	lower	upper	lower	upper
		Networks		
Direct effect	0.398	0.495	0.495	0.608
	(0.143)	(0.007)	(0.185)	(0.002)
Indirect effect	-0.098	-0.001	-0.113	0.000
	(0.143)	(0.007)	(0.185)	(0.002)
		Visa		
Direct effect	-0.417	-0.350	-0.497	-0.420
	(0.034)	(0.114)	(0.048)	(0.140)
Indirect effect	0.019	0.138	0.031	0.189
	(0.060)	(0.200)	(0.099)	(0.287)

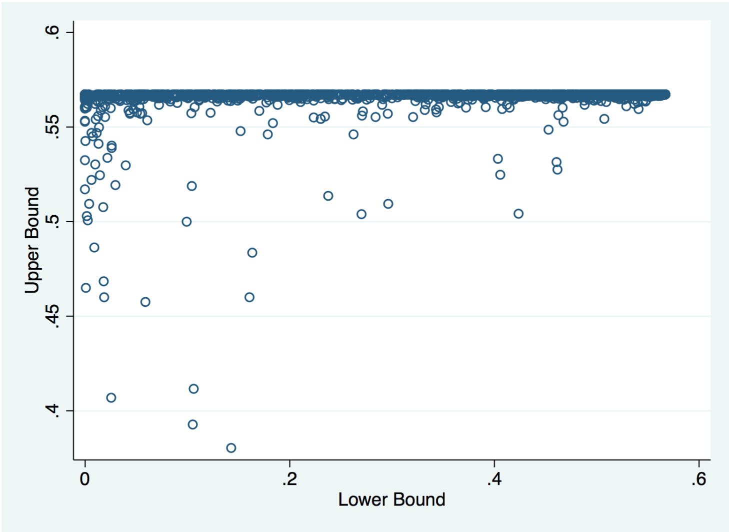
Note: standard deviations in parentheses. The bounds correspond to averages, weighted by population at origin, over equations (9), (10) and (15)-(17) based on the estimates in specifications (2) and (4) in Table 5.

Table 7: Determinants of migration flows (1990-2000), accounting for the endogeneity of networks

Sample	*Baseline*	*Population Size*	*Developing Countries*	*High Skill*	*Low Skill*
Specification	(1)	(2)	(3)	(4)	(5)
Dependent variable	flow	flow	flow	flow	flow
Model	2SRI PPML	2SRI PPML	2SRI PPML	2SRI PPML	2SRI PPML
ln(networks+1)	0.766***	0.780***	0.791***	0.661***	0.799***
	[0.066]	[0.067]	[0.088]	[0.053]	[0.080]
Visa requirement	-0.621***	-0.682***	-0.329**	-0.543**	-0.653**
	[0.231]	[0.258]	[0.156]	[0.256]	[0.283]
Schengen	-0.032	0.016	-	-0.325*	0.810*
	[0.245]	[0.246]	-	[0.170]	[0.478]
Colony	-0.082	-0.042	-0.004	-0.163	0.177
	[0.260]	[0.258]	[0.329]	[0.226]	[0.289]
Common language	0.098	0.051	0.037	0.469***	-0.161
	[0.178]	[0.182]	[0.266]	[0.143]	[0.230]
ln(distance)	0.191	0.234	-0.240	0.020	0.264
	[0.144]	[0.149]	[0.298]	[0.105]	[0.175]
First stage residual	-0.248***	-0.265***	-0.224**	-0.208***	-0.237**
	[0.076]	[0.078]	[0.107]	[0.055]	[0.101]
Destination fixed effects	Yes	Yes	Yes	Yes	Yes
Origin fixed effects	Yes	Yes	Yes	Yes	Yes
Origin*nest fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	5,611	4,497	4,708	5,611	5,611
Adjusted (pseudo) R^2	0.996	0.996	0.997	0.982	0.997
Log pseudo-likelihood	-2,168,416	-2,129,898	-1,421,063	-678,062	-1,694,777
Pesaran (2004) CD test	-1.60	-1.89	-1.48	-1.78	-0.38
p-value	0.110	0.059	0.139	0.074	0.701
First stage F-stat	540.37	445.71	359.70	540.37	540.37

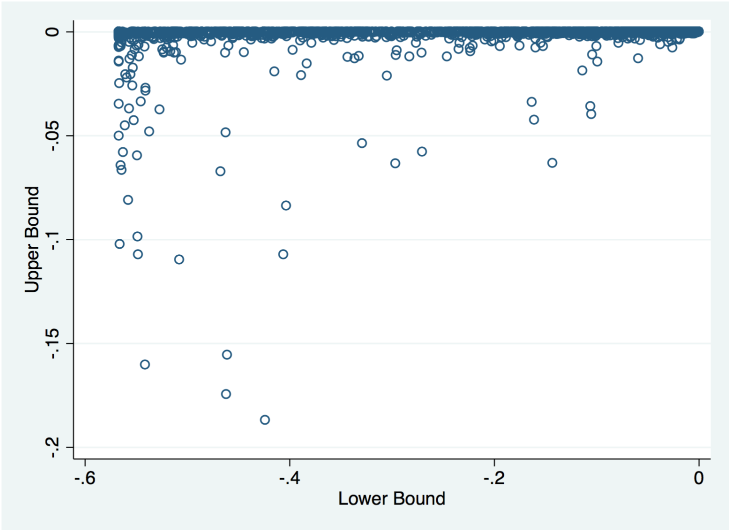
Note: standard errors in brackets; *** significant at the 99 percent level, ** significant at the 95 percent level, * significant at the 90 percent level. Specification (1) as specification (3) in Table 2; specifications (2) and (3) and specifications (2) and (4) in Table 4; specifications (4) and (5) as specifications (2) and (4) in Table 5. The dependent variable is equal to the maximum between the net flow and zero; standard errors are robust. The excluded instrument is the natural logarithm of one plus the size of migration networks in 1960 (see Table 1).

Figure 1: Bounds for the direct elasticity of migration flows with respect to networks



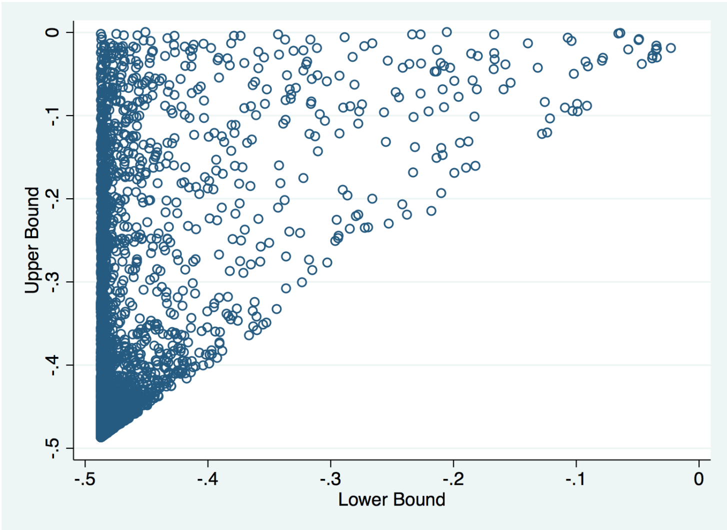
Note: see Table 3 for the average values.

Figure 2: Bounds for the indirect elasticity of migration flows with respect to networks



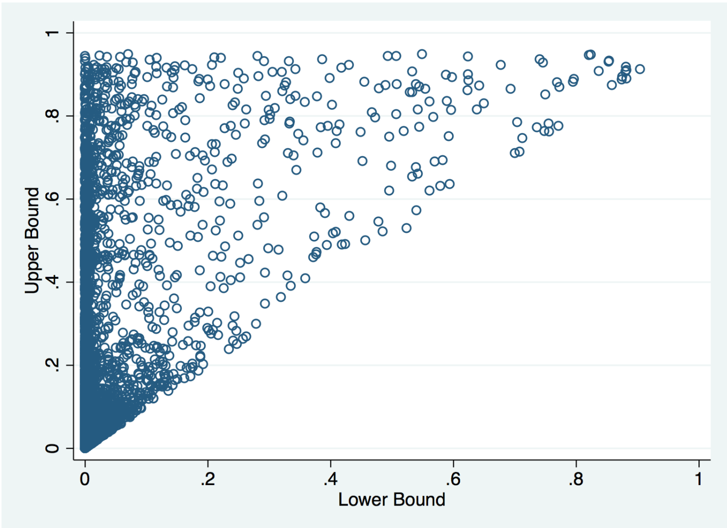
Note: see Table 3 for the average values.

Figure 3: Bounds for the direct effect of the visa requirement on migration flows



Note: see Table 3 for the average values.

Figure 4: Bounds for the indirect effect of the visa requirement on migration flows



Note: see Table 3 for the average values.