



**HAL**  
open science

## Visa Policies, Networks and the Cliff at the Border

Simone Bertoli, Jesús Fernández-Huertas Moraga

► **To cite this version:**

Simone Bertoli, Jesús Fernández-Huertas Moraga. Visa Policies, Networks and the Cliff at the Border. 2015. halshs-01099863

**HAL Id: halshs-01099863**

**<https://shs.hal.science/halshs-01099863>**

Preprint submitted on 5 Jan 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



CENTRE D'ETUDES  
ET DE RECHERCHES  
SUR LE DEVELOPPEMENT  
INTERNATIONAL

## SERIE ETUDES ET DOCUMENTS

### **Visa Policies, Networks and the Cliff at the Border**

Simone Bertoli

Jesús Fernández-Huertas Moraga

*Etudes et Documents* n° 27

December 2014

To cite this document:

Bertoli S., Fernández Huertas Moraga J. (2014) "Visa Policies, Networks and the Cliff at the border", *Etudes et Documents*, n°27, CERDI.  
[http://www.cerdi.org/back.php/production/show/id/1608/type\\_production\\_id/XXXX](http://www.cerdi.org/back.php/production/show/id/1608/type_production_id/XXXX)

CERDI  
65 BD. F. MITTERRAND  
63000 CLERMONT FERRAND – FRANCE  
TEL. + 33 4 73 17 74 00  
FAX + 33 4 73 17 74 28  
[www.cerdi.org](http://www.cerdi.org)

## The authors

Simone Bertoli

Associate Professor

Clermont Université, Université d'Auvergne, CNRS, UMR 6587, CERDI, F-63009 Clermont Ferrand

Email: [simone.bertoli@udamail.fr](mailto:simone.bertoli@udamail.fr)

Jesús Fernández-Huertas Moraga

Assistant Professor

Universidad Autónoma de Madrid, Ciudad Universitaria de Cantoblanco, 28049, Madrid

Email: [jesus.fernandez-huertas@uam.es](mailto:jesus.fernandez-huertas@uam.es)

**Corresponding author:** Simone Bertoli

*Etudes et Documents* are available online at: <http://www.cerdi.org/ed>

Director of Publication: Vianney Dequiedt

Editor: Catherine Araujo Bonjean

Publisher: Chantal Brige-Ukpong

ISSN: 2114 - 7957

## Disclaimer:

*Etudes et Documents* is a working papers series. Working Papers are not refereed, they constitute research in progress. Responsibility for the contents and opinions expressed in the working papers rests solely with the authors. Comments and suggestions are welcome and should be addressed to the authors.

**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.

**Key words:** international migration, networks, visa policies, multiple destinations, externalities

**JEL codes:** F22, O15, J61

**Acknowledgment**

The authors are grateful to Pedro Albarrán, Manuel Arellano, Michel Beine, Stéphane Bonhomme, Michael Clemens, Frédéric Docquier, Herbert Grubel, Dominik Hangartner, Fabio Mariani, Francesc Ortega, Çağlar Özden, Chris Parsons, Giovanni Peri and Hillel Rapoport for their comments, and to the participants in the 11<sup>th</sup> Journées Louis-André Gérard-Varet, in the 5<sup>th</sup> Migration and Development Conference, in the VI Workshop on Migration and Labor Economics, in the 2nd CEP-II-OECD Conference on Immigration in OECD Countries, in the XXXVII Symposium of the Spanish Economic Association, in the 1<sup>st</sup> CEMIR Conference on "International Migration: Competition for Talent and Brain Circulation", in the CESifo Workshop on "Migration Policies" and in seminar presentations at CERDI, at the University of Alicante and at IRES; the usual disclaimers apply.

*“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 (2013) provide empirical evidence that an aggregate measure of the restrictiveness of immigration policies reduces incoming flows from all origin countries. Still, some relevant host-country policies are bilateral in nature, so that potential 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 individuals to move across borders at a substantially lower cost, and with greatly reduced uncertainty with respect to their admission at destination.<sup>1</sup> This, in turn, suggests that the bilateral visa regime can also influence the scale of migration flows, as it determines the cost of entering legally into the country of

---

<sup>1</sup>Neumayer (2010) provides evidence on the negative impact of visa requirements on the number of travelers between countries.

destination, and then overstaying there beyond the period for which admission was granted. The US General Accounting Office (2004) reports that overstayers amounted to 2.3 million in 2000, accounting for at least 27 percent of illegal immigrants in the US. Six EU member states addressed in 2012 a complaint to the European Commissioner for Home Affairs about the alleged increase in migration inflows from the five Eastern European countries whose citizens had been granted a visa-free access to the Schengen area between 2009 and 2010.<sup>2</sup>

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 *et al.* (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, and Bertoli and Fernández-Huertas Moraga (2013) 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 exert a positive but only 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? Two closely related factors, namely the endogeneity of immigration policies and the dependence of bilateral flows on the attractiveness of other destinations, can explain this puzzle. With respect to visas, Chiswick (1988) observes that “the careful scrutiny given visa applicants, which offends many foreign students and visitors to the United States, is intended to ferret out those most likely to violate their visas” (p. 104), and the European legislation explicitly refers to the potential for illegal immigration from an origin country as one of the key criteria that is used to determine the visa policy toward its citizens.<sup>3</sup> Hence,

---

<sup>2</sup>“L’afflux de migrants des Balkans prèoccupe l’Union européenne”, *Le Monde*, October 24, 2012. The complaint was related to the increase in the number of asylum seekers from these countries, the same reason that also induced Canada to reintroduce in July 2009 the visa requirement on Czech citizens that had been lifted in October 2007, as described by Citizenship and Immigration Canada (source: <http://www.cic.gc.ca/english/department/media/backgrounders/2009/2009-07-13a.asp>, accessed on December 16, 2012).

<sup>3</sup>“The determination of those third countries whose nationals are subject to the visa requirement, and those exempt from it, is governed by a considered, case-by-case assessment of a variety of criteria relating *inter alia* to illegal immigration, public policy and security” (Council Regulation (EC) No 539/2001, March 15, 2001).

the bilateral visa regime can be correlated with unobservable factors that also shape the scale of migration flows.

Bilateral visa policies toward a given country are closely correlated across different destinations. The visa regimes that citizens from different origin countries face are 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.<sup>4</sup> Such a similarity in the bilateral visa policies toward the citizens of any given country can come from a policy coordination at the supranational level, as it occurs within the Schengen area, from a shared perception of the potential for illegal immigration or from the anticipation of an externality due to the policies adopted by other countries (Boeri and Brücker, 2005; Giordani and Ruta, 2013).

This can, in turn, create a further key analytical challenge to the identification of the effects of bilateral migration policies due to the complex dependence of bilateral migration flows upon the opportunities to migrate to other countries. The influence exerted by the attractiveness of alternative destinations upon the bilateral migration rate was labeled multilateral resistance to migration by Bertoli and Fernández-Huertas Moraga (2013), as the concept bears a close resemblance with the multilateral resistance that arises in gravity models of international trade (Anderson and van Wincoop, 2003). However, as it is the case in the international finance literature (Okawa and van Wincoop, 2012), the type of gravity equation<sup>5</sup> implied by a general specification of the behavior of migration flows goes beyond the simple extension proposed by Anderson (2011) and requires a different methodology. Even in the absence of general equilibrium effects or market clearing conditions, multilateral resistance to migration arises when unobservables prevent the econometrician from fully modeling the attractiveness of each destination. In such a case, the identification of the effect of the visa regime can be confounded by the visa policies adopted by other countries: potential migrants' destination choices depend on the *relative* size of the cliffs that characterize different borders rather than on their absolute size.

These arguments entail that the limited evidence on the effectiveness of bilateral immigration policies could 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 dyadic variables on bilateral migration flows

---

<sup>4</sup>“No visa required: Who has more freedom to travel?”, *The Economist*, August 25, 2010.

<sup>5</sup>See Beine *et al.* (2014) on this.

while controlling for such a confounding effect in a cross-sectional setting,<sup>6</sup> while at the same time greatly reducing the concerns related to differences in unobservables across countries that are subject to different visa regimes. In addition, we use our estimates to measure the diversion of the flows to other countries that is produced by the introduction of a visa requirement by one destination. This represents a migration policy externality that we are able to quantify. To our knowledge, this paper presents the first empirical estimate of an explicit migration policy externality.

We employ a Poisson pseudo-maximum likelihood, PPML, estimator that allows us (i) to be consistent with underlying random utility maximization, RUM, models with different patterns of dependency of the bilateral flows on the attractiveness of other destinations (Guimaraes *et al.*, 2004; Schmidheiny and Brülhart, 2011), and (ii) to deal with heteroscedastic disturbances (Santos Silva and Tenreyro, 2006) and with the presence of zeros (Santos Silva and Tenreyro, 2011).

The consistency of the PPML estimator with an underlying RUM model was first established by Guimaraes *et al.* (2003), and then extended by Guimaraes *et al.* (2004) and Schmidheiny and Brülhart (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 Brülhart (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 Brülhart (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 potential 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.<sup>7</sup> Second, the literature on discrete choice models (McFadden, 1974, 1978; Cardell, 1997; Train, 2003); third, the papers establishing the consistency of aggregate count data models with individual-level utility maximizing behavior (Guimaraes *et al.*, 2003, 2004; Schmidheiny and Brülhart,

---

<sup>6</sup>Bertoli and Fernández-Huertas Moraga (2013) propose a more general econometric approach that requires a longitudinal dimension that is often unavailable with international migration data.

<sup>7</sup>Other relevant empirical papers include Clark *et al.* (2007), Lewer and den Berg (2008), Belot and Hatton (2012), McKenzie *et al.* (2013), Bratsberg *et al.* (2012), Beine and Parsons (2014) and Bertoli *et al.* (2013a).



2011).

Our econometric analysis draws on the international migration data assembled by Docquier *et al.* (2009), which we combine with the dataset by Ozden *et al.* (2011) to obtain information on the size of the migration networks back 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 *et al.* (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 exerted by the attractiveness of alternative destinations, that is, multilateral resistance to migration.

As far as migration policy externalities are concerned, we estimate that a visa requirement imposed by one destination can increase bilateral migration flows to other destinations from the origin country subject to the visa 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 RUM model that describes the location decision problem that potential migrants face. Section 3 presents our estimation approach. The data sources and the basic descriptive statistics are presented in Section 4. Section 5 contains the results from the econometric analysis, while Section 6 presents a number of robustness checks. 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 choose their preferred location among the  $n$  countries belonging to the set  $D$ , including  $j$  itself.<sup>8</sup> Let  $m_{jk}$  represent the scale of the bilateral gross migration flow from country  $j$  to country  $k$ , and  $\mathbf{m}_j$  be the  $n \times 1$  vector that collects all the bilateral migration flows originating from country  $j$ . We can express its  $k$ -th element,  $m_{jk}$ , as:

$$m_{jk} = s_j p_{jk} \eta_{jk} \quad (1)$$

where  $p_{jk}$  is the expected 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 dyads  $(j, k)$ .

### 2.1 Choice probabilities

The  $n$  elements of the vector  $\mathbf{p}_j$  are the outcome of a location decision problem that individuals face and that we describe through a RUM model. Specifically, the utility  $U_{ijk}$  that the individual  $i$  from country  $j$  obtains from opting for destination  $k$  is given by the sum of a dyad-specific deterministic component  $V_{jk}$  and of an individual-specific stochastic term  $\epsilon_{ijk}$ :

$$U_{ijk} = V_{jk} + \epsilon_{ijk} \quad (2)$$

The vector  $\mathbf{p}_{ij} = (p_{ij1}, \dots, p_{ijn})$  that collects the choice probabilities for individual  $i$  over the  $n$  locations depends on the assumptions about the distribution of the stochastic term. These are, in turn, closely related with the specification of the deterministic component of utility in (2), which is usually assumed to be a linear function of a vector  $\mathbf{x}_{jk}$ :

$$V_{jk} = \mathbf{x}_{jk}' \boldsymbol{\beta} \quad (3)$$

Since the seminal contributions by McFadden (1974, 1978), the literature on discrete choice models has typically assumed that  $\epsilon_{ijk}$  follows an Extreme Value Type-1 marginal distribution, with this assumption allowing to write down the choice probability  $p_{jk}$  as a

---

<sup>8</sup>We present the RUM model omitting the time dimension of the location decision problem that potential migrants face, but the analysis can be extended to allow for such a dimension.

destination-specific analytic function of the vector  $\mathbf{V}_j$  of the deterministic components of utility, i.e.,  $p_{jk} = f_k(\mathbf{V}_j)$ . If and only if the stochastic component of utility is i.i.d. EVT-1, then the choice probabilities can be represented through a conditional logit model and they satisfy the property of the independence from irrelevant alternatives, which implies that the ratio  $p_{jk}/p_{jh}$  is *not* a function of the *full*  $n \times 1$  vector  $\mathbf{V}_j$ , but only of the difference between  $V_{jk}$  and  $V_{jh}$ .

As Train (2003) observes, the independence from irrelevant alternatives can be seen “as the natural outcome of a well-specified model that captures all sources of correlation over alternatives into representative utility, so that only white noise remains.” (p. 76). Thus, this analytically convenient property fails if data constraints prevent a full specification of the deterministic component of utility. Then, the realization of  $\epsilon_{ijk}$  contains information on the realizations of the stochastic components of utility for other alternatives. This entails, in turn, that the choice probability  $p_{jk}$  becomes more sensitive to variations in the deterministic component of utility of alternative locations with which location  $k$  shares some unobserved determinants of utility than the conditional logit model would predict.

The literature on discrete choice models has introduced a variety of distributional assumptions that allow for a non-zero correlation across alternatives in the stochastic component of utility. These models give rise to a more flexible pattern of responses of the elements of the vector  $\mathbf{p}_j$  when the elements of  $\mathbf{V}_j$  vary. The richer structure of the stochastic component of utility in (2) is precisely meant to correct for the bias in the estimation of the coefficients of the determinants of  $V_{jk}$  that would otherwise be produced by its incorrect specification.

While modeling migration decisions, the need to introduce more general distributional assumptions naturally arises from the presence of unobserved determinants of the attractiveness of a country, and from the estimation of the determinants of bilateral migration flows on aggregate data. Imagine, for instance, that cultural proximity with the origin country  $j$ , which is unobservable for the econometrician, influences the attractiveness of the various destinations; then, a potential migrant from the origin  $j$  receives a utility, conditional upon observables, from locating in culturally close destinations that systematically differs from the utility associated to culturally distant destinations.

The assumption, which we retain from the literature, that the vector of parameters  $\beta$  in (3) does not vary across individuals implies that any heterogeneity in the relationship between the elements of  $\mathbf{x}_{jk}$  and  $U_{ijk}$  ends up in  $\epsilon_{ijk}$ , introducing a correlation in the stochastic

component of utility across destinations.<sup>9</sup> Suppose, for instance, that one of the elements of  $\mathbf{x}_{jk}$  is represented by a dummy variable that signals whether country  $j$  and country  $k$  share an official language, as in Grogger and Hanson (2011) or Beine *et al.* (2011). The specification of  $V_{jk}$  in (3) implies that the deterministic component of utility that a Belgian would-be migrant obtains from locating in any destination does not depend on whether she is Walloon (French-speaking) or Flemish (Dutch-speaking). This, in turn, implies that the higher (lower) utility that a Walloon (Flemish) receives from locating in any country that has French among its official languages introduces a positive correlation in the stochastic component of utility across French-speaking destinations. Along the same lines, destination countries differ with respect to the average level and the dispersion of wages. Location-specific utility  $U_{ijk}$  is increasing with the average level of wages in  $k$  for all potential migrants from  $j$ , while the dispersion of wages can produce an opposite influence on the attractiveness of this destination for individuals characterized by different levels of ability (Borjas, 1987).

This is why the presence of unobservables and the specification of utility in (2) that is adopted in the literature calls for relaxing the assumption that the stochastic component of utility is independently distributed across countries when deriving equations to be estimated on aggregate bilateral migration data.

### 2.1.1 Distributional assumptions

Let the set of possible locations  $D$  be partitioned into  $m \geq 1$  subsets  $b$ , also called nests, and let  $b(k) \subseteq D$  denote the unique subset to which location  $k$  belongs to. Nests are groups of countries that share some observed or unobserved characteristics that influence their attractiveness, and whose impact can be heterogeneous across individuals. The individual stochastic component  $\epsilon_{ijk}$  of utility is assumed to be a mixture of a nest-specific and of a location-specific term:

$$\epsilon_{ijk} = (1 - \tau)\nu_{ijb(k)} + \tau v_{ijk} \quad (4)$$

where  $\tau \in (0, 1]$  is the weight associated to the location-specific term,  $v_{ijk} \stackrel{\text{iid}}{\sim}$  EVT-1 and  $\nu_{ijb(k)}$  is the unique random variable, whose distribution depends on  $\tau$ , that ensures that also  $\epsilon_{ijk}$  follows an EVT-1 marginal distribution (Cardell, 1997). The presence of the nest-

---

<sup>9</sup>“That heterogeneity is a special type of correlation amongst choice situations is not well understood.” (Hensher and Greene, 2003, p. 160).

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. The higher the weight  $\tau$  associated to the location-specific term, the lower the within-nest correlation of the stochastic component of utility in (2).

### 2.1.2 The vector of choice probabilities

The element  $k$  in the vector of choice probabilities  $\mathbf{p}_{ij}$  is equal to:<sup>10</sup>

$$p_{ijk} = \frac{e^{\mathbf{x}_{jk}'\beta/\tau} \left( \sum_{l \in b(k)} e^{\mathbf{x}_{jl}'\beta/\tau} \right)^{\tau-1}}{\sum_q \left( \sum_{l \in b_q} e^{\mathbf{x}_{jl}'\beta/\tau} \right)^\tau} \quad (5)$$

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

$$m_{jk} = s_j \frac{e^{\mathbf{x}_{jk}'\beta/\tau} \left( \sum_{l \in b(k)} e^{\mathbf{x}_{jl}'\beta/\tau} \right)^{\tau-1}}{\sum_q \left( \sum_{l \in b_q} e^{\mathbf{x}_{jl}'\beta/\tau} \right)^\tau} \eta_{jk} \quad (6)$$

The assumptions on the stochastic component of location-specific utility in (4) are more general than those adopted by other papers in the literature; specifically, our distributional assumptions reduce to those adopted by Grogger and Hanson (2011) if we further assume that all locations belong to a unique nest, i.e.,  $b(k) = D, \forall k \in D$ . Similarly, we can obtain the distributional assumptions in Ortega and Peri (2013), Beine *et al.* (2011) and McKenzie *et al.* (2013) by imposing the restriction that all locations but the origin belong to a unique nest, i.e.  $b(j) = \{j\}$  and  $b(k) = D/\{j\}$  for any  $k \in D$ . This assumption implies that, conditional upon the deterministic component of location-specific utility, potential migrants regard all possible countries except the origin as being close substitutes, and it can accommodate for differences in unobservables between migrants and stayers.

The assumptions on the stochastic component  $\epsilon_{ijk}$  that we introduced in (4) allow for a richer pattern of cross-elasticities, as potential migrants can perceive a destination  $h$  to be a close substitute only for a subset of all the potential destinations, represented by the nest  $b(h) \subset D$ . Specifically, we can use (6) to derive the elasticity of the bilateral migration flow

<sup>10</sup>See, for instance, Train (2003) for the derivation of the choice probability.

from  $j$  to  $k$  with respect to the attractiveness of a destination  $l \neq k$  for potential migrants from  $j$ :

$$\frac{\partial \ln(m_{jk})}{\partial \ln(V_{jl})} = - \left[ \tau p_{jk} + (1 - \tau) \frac{p_{jk} p_{jb(k)} p_{jl|b(k)}}{p_{jl}} \right] V_{jl} / \tau \quad (7)$$

where  $p_{jb(k)}$  is the probability that a potential migrant from  $j$  opts for a destination in the nest  $b(k)$ , and  $p_{jl|b(k)}$  is the probability of choosing destination  $l$  conditional upon opting for the nest  $b(k)$ .<sup>11</sup>

If destination  $l \in b(k)$ , then (7) simplifies to  $-p_{jk} V_{jl} / \tau$ , while if the destination  $l$  does not belong to the nest  $b(k)$ , and hence  $p_{jl|b(k)} = 0$ , then the indirect elasticity stands at  $-p_{jk} V_{jl}$ . As the weight  $\tau$  associated to the location-specific stochastic term in (4) lies between 0 and 1, then the indirect elasticity is larger in magnitude when  $l \in b(k)$ , and it is monotonically decreasing in  $\tau$ . Intuitively, the higher the weight  $(1 - \tau)$  associated to the nest-specific stochastic component in (4), the greater the sensitivity of bilateral migration flows to a variation in the attractiveness of another destination within the same nest.

### 2.1.3 The equation to be estimated

We can rewrite (6) more compactly as follows:

$$m_{jk} = \exp \left( \alpha_j + \mathbf{x}_{jk}' \boldsymbol{\beta} / \tau + \gamma_{jb(k)} + \ln(\eta_{jk}) \right) \quad (8)$$

where the origin specific term  $\alpha_j$  is equal to:

$$\alpha_j = \ln(s_j) - \ln \left[ \sum_{q=1}^m \left( \sum_{l \in b_q} e^{\mathbf{x}_{jl}' \boldsymbol{\beta} / \tau} \right)^\tau \right]$$

and the origin-nest specific term  $\gamma_{jb(k)}$  is given by:

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

Notice that the vector of parameters  $\boldsymbol{\beta}$  of the determinants of location-specific utility in (3) always appears in (8) scaled by the dissimilarity parameter  $\tau$ . The estimation of (8) does not separately identify  $\boldsymbol{\beta}$  and  $\tau$ , but only the ratio  $\boldsymbol{\beta} / \tau$ . As the direct and indirect elasticities

<sup>11</sup>The corresponding expression for the direct elasticity of migration flows from  $j$  to  $k$  with respect to  $V_{jk}$  is  $\partial \ln(m_{jk}) / \partial \ln(V_{jk}) = [\tau(1 - p_{jk}) + (1 - \tau)(1 - p_{jk|b(k)})] V_{jk} / \tau$ .

presented above depend on  $\tau$ , this introduces a fundamental uncertainty on their true size that is independent from the precision with which  $\beta/\tau$  is estimated. Section 5.2 below introduces bounds for the values of the two elasticities, extending the proposed approach by Schmidheiny and Brülhart (2011).

A further key analytical challenge in the estimation of (8) is represented by the correlation between the elements of the vector  $\mathbf{x}_{jk}$  and the term  $\gamma_{jb(k)}$  that reflects the expected utility for potential migrants from  $j$  from choosing a location in the nest  $b(k)$  (Train, 2003). Such a correlation arises, for instance, if the destinations belonging to the nest  $b(k)$  adopt similar bilateral immigration policies *vis-à-vis* the citizens of the origin country  $j$  (Giordani and Ruta, 2013); in such a case, the attractiveness of destination  $k$  would be positively correlated with the expected utility of opting for other destinations that are close substitutes to country  $k$  via the similarity of their bilateral immigration policies towards  $j$ . If  $\gamma_{jb(k)}$  is not adequately controlled for, then the estimation of the vector  $\beta/\tau$  would be exposed to an omitted variable bias.

### 3 PPML estimation of bilateral migration flows

The estimation approach that represents the industry-standard in the international migration literature involves a logarithmic transformation of the bilateral migration *rates* that can be derived from (8).<sup>12</sup> As discussed in Santos Silva and Tenreyro (2006), the assumption that  $E(\eta_{jk}) = 1$  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 (8); this, in turn, implies that, if  $\eta_{jk}$  is heteroskedastic with a variance that depends on the regressors in (6), then the transformed error term  $\ln(\eta_{jk}/\eta_{jj})$  will be correlated with the regressors, creating a serious threat to identification.

Santos Silva and Tenreyro (2006) proposed to resort to the estimation of (8) through Poisson pseudo-maximum likelihood estimation; this approach allows to deal with the presence of zeros in the dependent variable, and it is gaining momentum in the international migration literature. PPML estimation performs well even when the data are not Poisson-

---

<sup>12</sup>This approach has been adopted, *inter alia*, by Clark *et al.* (2007), Lewer and den Berg (2008), Ortega and Peri (2013), Mayda (2010), McKenzie *et al.* (2013), Simpson and Sparber (2013), Beine *et al.* (2011), Grogger and Hanson (2011) and Bertoli and Fernández-Huertas Moraga (2013).

distributed,<sup>13</sup> and when the data present a mass point at zero (Santos Silva and Tenreyro, 2011).<sup>14</sup>

We discuss here on the consistency of the Poisson estimation with the RUM model that gives us the expected scale of the observed bilateral migration flows. This requires going back to the expression for  $m_{jk}$  in (8): if one assumes, as before, that  $b(k) = D$ ,  $\forall k \in D$ , i.e., the IIA assumption holds, then we can simplify the expressions for  $\alpha_j$  and  $\gamma_{jb(k)}$ , as we have that  $\alpha_j = \ln(s_j) - \ln\left(\sum_{l \in D} e^{x_{jl}'\beta}\right)$  and  $\gamma_{jb(k)} = 0$ . Hence, when IIA characterizes the underlying RUM model, we can rewrite  $m_{jk}$  as follows:

$$m_{jk} = \exp\left(x_{jk}'\beta + \ln(s_j) - \ln\left(\sum_{l \in D} e^{x_{jl}'\beta}\right) + \ln(\eta_{jk})\right) \quad (9)$$

Some key observations emerge from the inspection of (9). 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  $k$ .

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 (9) in a non-linear way.<sup>15</sup>

Third, a RUM-consistent estimation of (9) 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  $x_{jk}$  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.

Guimaraes *et al.* (2003) demonstrate that the estimation of (9) through PPML delivers the same estimate for  $\beta$  as a conditional logit model estimated on individual-level data on

<sup>13</sup>This estimation technique produces consistent estimates as long as the conditional mean is correctly specified (Gourieroux *et al.*, 1984).

<sup>14</sup>The two-part model proposed by Egger *et al.* (2011) is an alternative way to handle the presence of a mass point at zero in the data; our estimates are robust to the adoption of this technique. Results are available from the authors upon request.

<sup>15</sup>Observe that  $V_{jj}$  enters linearly into the exponential of the ratio of the conditional means for  $m_{jk}$  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  $m_{jj}$  by Jensen's inequality, independently on the distributional assumptions on the underlying data-generating process), and this, in turn, violates the condition that is required to obtain consistent estimates with PPML (Gourieroux *et al.*, 1984).



the same determinants of location-specific utility, as the log-likelihood functions of the two models are identical up to a constant.<sup>16</sup> Hence, this estimation technique is fully consistent with the underlying RUM model that describes the choice of the utility-maximizing location.

Schmidheiny and Brülhart (2011) generalize this result under the same assumptions as in Ortega and Peri (2013), so that the model to be estimated becomes:

$$m_{jk} = \exp \left( \mathbf{x}_{jk}'\boldsymbol{\beta} + \ln(s_j) + \tau \ln \left( \sum_{l \in D} e^{\mathbf{x}_{jl}'\boldsymbol{\beta}/\tau} \right) - \ln \left[ e^{\mathbf{x}_{jj}'\boldsymbol{\beta}} + \left( \sum_{l \in D/\{j\}} e^{\mathbf{x}_{jl}'\boldsymbol{\beta}/\tau} \right)^\tau \right] + \ln(\eta_{jk}) \right) \quad (10)$$

PPML estimation of (10) 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 Brülhart, 2011). Observe that the origin fixed effects suffice to restore independence across observations both in (9) and (10), although the stochastic properties of the two underlying RUM models differ.<sup>17</sup> This, in turn, implies that PPML estimation is always characterized by a fundamental uncertainty about the magnitude of the elasticity of migration flows, which is connected to the inability to identify  $\tau$ .

### 3.1 Consistency with more general RUM models

Schmidheiny and Brülhart (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 (2013). Here, we go one step further, showing that the same consistency characterizes the estimation of (8), which we derived from (4). We reproduce here (8):

$$m_{jk} = \exp \left( \alpha_j + \mathbf{x}_{jk}'\boldsymbol{\beta}/\tau + \gamma_{jb(k)} + \ln(\eta_{jk}) \right)$$

---

<sup>16</sup>Guimaraes *et al.* (2003) focus on location-decisions taken from a single origin, so that  $\alpha$  does not vary with  $j$ ; Guimaraes *et al.* (2004) show that, with multiple time periods, the inclusion of origin-time dummies suffices to restore the parallel between the conditional logit and Poisson.

<sup>17</sup>The key is that the term describing the expected utility from migration to any destination in the nest  $D/\{j\}$  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 *et al.*, 2004).

PPML estimation requires observations to be cross-sectionally independent and, as discussed in Guimaraes *et al.* (2004), this can be achieved with the inclusion of a richer structure of dummies. Specifically, the inclusion of origin-nest dummies suffices to control for  $\gamma_{jb(k)}$ , and restore cross-sectional independence of the residuals.<sup>18</sup> This, in turn, will produce a consistent estimate of  $\beta/\tau$ , 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  $\beta/\tau$  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 (8) increases, the available variability that can be exploited to estimate  $\beta/\tau$  goes down. On the other hand, choosing a too parsimonious specification with few nests may not be able to fully restore the cross-sectional 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 cross-sectional dependence of the residuals in equation (8).

### 3.2 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  $e_{jk}$  represent the Pearson residual associated to the migration flow from the origin  $j$  to the destination  $k$ ,<sup>19</sup> and  $\mathbf{e}_k$  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  $E(\mathbf{e}_k' \mathbf{e}_l) = 0$  for  $l \neq k$ , while the presence of a nest-specific stochastic component of utility would entail that  $E(\mathbf{e}_k' \mathbf{e}_l) > 0$  if  $b(l) = b(k)$ . 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

<sup>18</sup>A similar use of nests can be found in the analysis of firms' location choice by Head *et al.* (1995) and Levinson (1996); see also the other papers cited by Guimaraes *et al.* (2004).

<sup>19</sup>Hsiao *et al.* (2012) provide evidence on the reliability of the Pearson residuals when testing for cross-sectional dependence in non-linear models.

by Pesaran (2004).<sup>20</sup> Let  $\hat{\rho}_{kl}$  denote the correlation between the vectors  $\mathbf{e}_k$  and  $\mathbf{e}_l$ ; the  $CD$  test statistic is given by:

$$CD = \left( \frac{2n_o}{n_d(n_d - 1)} \right)^{1/2} \sum_{k=1}^{n_d-1} \sum_{l=k+1}^{n_d} \hat{\rho}_{kl} \quad (11)$$

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 restore the cross-sectional independence of the residuals in equation (8).

## 4 Data

### 4.1 Data sources

We draw our data from three main sources. The first one is Docquier *et al.* (2009), who provide information on the size of bilateral migration stocks of individuals aged 25 and above in 31 countries of destination in 1990 and 2000. This dataset provides a proxy for the scale of gross migration flows that is represented by the variation in stocks,<sup>21</sup> a proxy that has been used, *inter alia*, by Beine *et al.* (2011).

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 based on census and register data and it might fail to capture undocumented migrants; if visa waivers do offer an option for a legal admission at destination for undocumented migrants, then this measurement error in the dependent variable would go against finding a significant relationship between bilateral visa policies and

---

<sup>20</sup>“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 independence when the data present both patterns of positive and negative correlation (Frees, 1995).

<sup>21</sup>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, p. 725).

the scale of gross migration flows.<sup>22</sup> The data by Docquier *et al.* (2009) are matched with the ones assembled by Ozden *et al.* (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 monthly publication of the International Air 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.<sup>23</sup> Observe that visa policies are based on citizenship rather than on country of birth, which is the basis for the migration data provided by 26 out of 31 destination countries in Docquier *et al.* (2009); 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) and Perkins and Neumayer (2013), refers to December 2004.<sup>24</sup> 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).<sup>25</sup>

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.

---

<sup>22</sup>Notice that although a visa overstayer becomes an undocumented immigrant after the expiry of the period for which admission was granted, the overstayer might have been regularized by the time of the following population census if the destination country adopts an amnesty for undocumented migrants.

<sup>23</sup>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).

<sup>24</sup>We repeatedly contacted the customer service of the IATA to obtain earlier editions of the manual, but regrettably the December 2004 edition is the oldest that is currently available.

<sup>25</sup>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.

## 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 variation in bilateral migration stocks between 1990 and 2000 takes strictly positive values. The sample size goes down from 5,611 origin-destination observations to just 3,466, fully dropping three destinations: Hungary, Korea and Poland.<sup>26</sup> Thus, 62 percent of the observations remain for OLS regressions on the logarithm of the bilateral migration rate. The largest increase in the bilateral stock (3.7 million) corresponds to the Mexican migration to the US whereas the minimum (-189,660) refers to the decline of the stock of German migrants in 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 in the Appendix 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. These figures suggest that, on average, the citizens of the origin

---

<sup>26</sup>The size of the 1990 migrant stock are estimated rather than observed for 10 destination countries, including Hungary, Korea and Poland (Docquier *et al.*, 2009, p. 317), and this introduces an additional measurement error in the variable; more specifically, the size of the estimated stocks for these three destinations are higher for all origin countries than observed stocks in 2000.

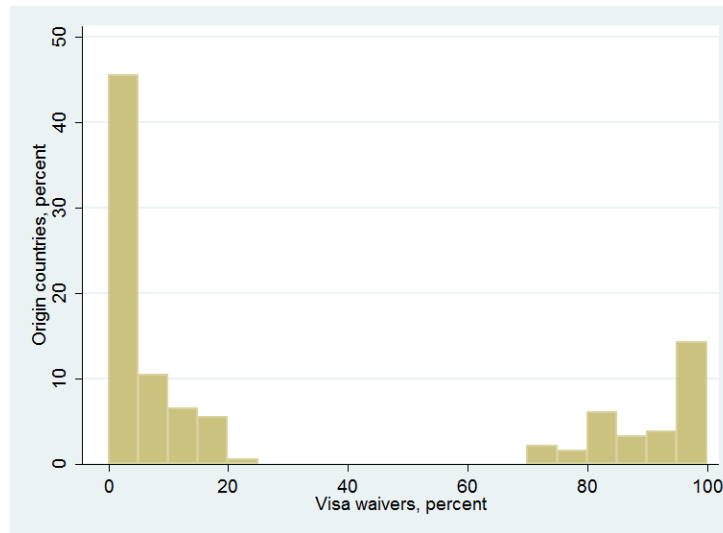
Table 1: Descriptive statistics

	<i>Full sample</i> (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	
	<i>Positive variations in stocks</i> (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 *et al.* (2009) for flows and migration networks in 1990; Ozden *et al.* (2011) for migration networks in 1960; Neumayer (2006) for the visa requirement, and Mayer and Zignago (2011) for the rest of the variables.

countries in our sample require a visa to be admitted in 69 percent of the destinations; this average hides considerable variability across origins, as revealed by Figure 1. As mentioned in the introduction, the opportunities for non-immigrant admission at destination are highly polarized, with 64 countries facing a visa requirement in all countries in our sample, and 13 countries benefiting from a visa waiver in all destinations.

Figure 1: Distribution of the countries of origin by visa regime



Source: authors' elaboration on Neumayer (2006).

The following variable in Table 1 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 (9 of the 31 destination countries in this period) adopted a common visa policy towards any origin country in our sample in 2004,<sup>27</sup> so that the inclusion of this variable, which is introduced following the main specification in Beine *et al.* (2011), could, if anything, limit the ability of the models below to identify the effect of the visa variable.<sup>28</sup> 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

<sup>27</sup>This was not the case in earlier years; for instance, 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, 2013).

<sup>28</sup>In fact, the results below are not sensitive to the exclusion of the Schengen variable.

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 introduced in Section 2. In order to closely tie the results to the existing literature, we begin by reproducing the OLS estimation in Beine *et al.* (2011) in the first data column in Table 2. The specification is exactly the same as in Beine *et al.* (2011) but for the addition of the visa requirement variable introduced in the previous section. It 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 (2013), 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). Origin and destination fixed effects also allow us to control for the dependency of current bilateral migration flows on the future attractiveness of all destinations, as discussed in Bertoli *et al.* (2013a). 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.<sup>29</sup> The inclusion of origin dummies entails that the identifying variation for the effect of the bilateral visa policy comes from the 105 origin countries in our sample that face different visa regimes across the 31 destinations that we include in the analysis.

Reassuringly, this specification produces the same results as in Beine *et al.* (2011) for all of the variables that they also included. Distance, colonial links and common language appear as significant correlates of the log of immigration rates. In particular, the coefficient on the log of networks in 1990 exactly coincides with that in Beine *et al.* (2011): a highly significant 0.62. The introduction of the visa requirement variable as an additional explanatory variable

---

<sup>29</sup>We follow Beine *et al.* (2011) in adding one to the size of the 1990 migration networks so as not to discard zero observations.



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

<i>Specification</i>	(1)	(2)	(3)
<i>Dependent variable</i>	ln(flow)	flow	flow
<i>Model</i>	OLS	PPML	PPML
ln(networks+1)	0.621*** [0.018]	0.658*** [0.042]	0.567*** [0.049]
Visa requirement	-0.051 [0.106]	0.017 [0.161]	-0.667*** [0.215]
Schengen	0.278 [0.179]	0.651* [0.381]	0.034 [0.235]
Colony	0.313** [0.137]	-0.290 [0.217]	0.451* [0.256]
Common language	0.420*** [0.076]	0.333** [0.130]	0.302* [0.161]
ln(distance)	-0.396*** [0.046]	-0.382*** [0.098]	-0.121 [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 variation in stocks and zero; standard errors are robust in specifications (1) to (3); the pseudo  $R^2$  for specifications (2)-(3) is defined as one minus the ratio between the log-likelihood of the model over the log-likelihood of a restricted model which only includes a constant.

does not have any effect on the rest of parameters, and the variable itself shows as non-significant.

This specification might be exposed to an inconsistency with the assumptions on the stochastic component of location-specific utility in the underlying RUM model. If the vector of regressors  $\mathbf{x}_{jk}$ , which we augmented with the inclusion of bilateral visa policies, fails to include all relevant dyadic determinants of migration or if some observed factors have an heterogeneous impact across potential migrants, 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, 2013), with the elements of  $\mathbf{x}_{jk}$  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 (De Hoyos and Sarafidis, 2006).

The second key problem with the OLS specification is precisely the need to discard non-positive values,<sup>30</sup> which can bias the estimated coefficients (Santos Silva and Tenreyro, 2006, 2011). This problem can be directly dealt with by using the Poisson regression model on the full sample from Table 1.<sup>31</sup> Specification (2) in Table 2 shows the result from running a Poisson regression on exactly the same variables as in specification (1). The estimates in specifications (1) and (2) are very similar,<sup>32</sup> 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 cross-sectional dependence in the error term. The presence of cross-sectional dependence in the residuals would imply that the coefficients from specification (2) in Table 2

---

<sup>30</sup>We also estimated the model with scaled-OLS, as in Ortega and Peri (2013); the coefficient of the visa variable is insignificant also in this case. Results available from the authors upon request.

<sup>31</sup>For the purposes of estimation, the 7 percent of negative values are set to zero, as variations in bilateral stocks are used as a proxy for unobserved gross flows, which are nonnegative by definition.

<sup>32</sup>We report robust standard errors for specification (2), which, as demonstrated by Gourieroux *et al.* (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.

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) using the *xtcd* Stata command introduced by Eberhardt (2011). Table 2 shows a statistic of 17.35, which strongly rejects the null of cross-sectional independence.

### 5.1.1 Reducing cross-sectional dependence

The approach that we adopt here is to restore cross-sectional independence by reducing the variability in the data that is used for identification. Specifically, the inclusion of origin-nest dummies allows us to control for unobservable nest-specific components of location-specific utility that have a differential impact on potential migrants from different countries of origin. This approach is not 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.<sup>33</sup> Why is there a need to specify assumptions, rather than opting for a systematic exploration of all alternative nest structures of the 31 destinations in our sample? The answer is represented by the Bell number  $B_n$ , which gives the number of possible partitions of a set of  $n$  elements: for a set of 31 elements we have that  $B_{31} \approx 5.4 \times 10^{23}$ ,<sup>34</sup> so that a systematic exploration of all alternative nest structures is computationally unfeasible. Bonhomme and Manresa (2012) recently proposed an approach that leads to an optimal (in terms of fit rather than cross-sectional dependence) partition of the units of analysis into non-overlapping nests for the estimation of a linear model. Their approach could in principle be extended to nonlinear models (page 10) but, as they acknowledge, the properties of such an estimator are yet to be studied and are out of the scope of this paper.

The adequacy of a partition of the destination countries into nests can be gauged by its ability to restore the cross-sectional 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.

---

<sup>33</sup>See, for instance, the discussion on the composition of the nests in Head *et al.* (1995).

<sup>34</sup>The Bell number is defined in a recursive way as follows:

$$B_{n+1} = \sum_{k=0}^n \binom{n}{k} B_k$$

with  $B_0 = 1$ .

Coarser nests, with the unique nest of destinations *à la* Ortega and Peri (2013) representing the limiting case of coarseness, have more identification power at the expense of a greater risk of an incorrect specification. 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 cross-sectional independence but delivers no identification in the cross section as it would be equivalent to origin-destination dyadic fixed effects.

This trade-off leads us to propose the following approach: if the  $CD$  test rejects the null of cross-sectional independence on the basis of a specification with  $m \geq 1$  nests, then we opt for a specification with  $m+1$  nests. This requires us to determine the criteria that inform how we define finer nests, and we opted for geographical proximity of the destinations 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 cross-sectional independence, with the sign or significance of the estimated coefficients being irrelevant for this stopping rule. The sequence of estimates that we obtained under progressively finer partitions of the set of destinations  $D$  is reported in Table A.1 in the Appendix A.

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 cross-sectional 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 US), and 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.<sup>35</sup> The following step was to split the nest  $b_{21}$  between a nest  $b_{211}$  for North America (Canada and the US) 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 still rejected the null at the 1 percent confidence level.

---

<sup>35</sup>Notice that the coefficient of the visa variable is negative and significant in this specification.

Finally, we ran a six-nests 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 led to a rejection of the null. Specification (3) in Table 2 reports the estimates,<sup>36</sup> obtained after adding controls that interact 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 cross-sectional correlation in the error term are nest-specific, with the destinations belonging to any of the six nests regarded as close substitutes by potential 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 and distance variables.<sup>37</sup> 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: 0.567. This fall is 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, such as (1)

---

<sup>36</sup>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). All of our results are robust to the exclusion of the last nest. Notice that our estimation approach does *not* require that other destinations that are not included in our sample do not belong to these six 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.

<sup>37</sup>The instability of the estimated coefficient for distance is not related to its correlation with the visa variable. Although the raw correlation between the two stands at 0.25, suggesting that distance to the destination country is positively correlated with the imposition of a visa requirement, this correlation declines to 0.04 once we partial out origin and destination fixed effects and to 0.03 after partialling out origin-nest fixed effects, so that multicollinearity cannot explain the change in the significance of distance.

and (2) in Table 2, that does not adequately control for multilateral resistance to migration, as shown by the test on the residuals, 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) in Table 2 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 toward any origin are closely correlated across destinations, as evidenced by Figure 1. This confirms one of the assumptions in the model by Giordani and Ruta (2013). The correlation in policies, in turn, introduces a correlation between the bilateral visa policy adopted by country  $k$  and the attractiveness of alternative destinations for potential 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 determinants 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 (2013) for 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 *et al.* (2004). A remaining threat to identification would be represented by the existence of differences in dyadic unobservables within a nest. For example, in the presence of reverse causality, with destinations requiring visas whenever migration flows from an origin are high, we would expect the magnitude of our coefficient to be downward biased.<sup>38</sup> This might suggest that our estimate of the visa effect might indeed be downward biased, though our estimation approach already greatly reduces the concerns related to unobservables, accounting for their influence on the pattern of correlation in the stochastic component of utility. The robustness checks presented in Section 6 below allow us to further reduce the empirical relevance of these legitimate concerns.

---

<sup>38</sup>Bertoli and Fernández-Huertas Moraga (2013), who can control for time-varying dyadic unobservables thanks to the frequency and to the longitudinal dimension of their data, find a larger effect of visa policies on migration flows to Spain.

## 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. The difficulty here is that Table 2 gives us estimates for  $\beta/\tau$  whereas we are unable to separately identify the elements of the vector  $\beta$  and  $\tau$ , with  $\tau$  entering separately in the expressions for the elasticities provided in Section 2.1.2 above.

Following the approach adopted in Schmidheiny and Brülhart (2011), we can define bounds for the two elasticities, conditional upon the estimated value of  $\widehat{\beta/\tau}$ , exploiting their monotonicity in  $\tau$ . Specifically, computing the direct elasticity for  $\tau$  converging to 0 and for  $\tau = 1$ , we can observe that:<sup>39</sup>

$$\left. \frac{\partial \ln(m_{jk})}{\partial \ln(V_{jk})} \right|_{\beta/\tau = \widehat{\beta/\tau}} = \left( (1 - p_{jk|b(k)}) \mathbf{x}'_{jk} \widehat{\beta/\tau}, (1 - p_{jk}) \mathbf{x}'_{jk} \widehat{\beta/\tau} \right] \quad (12)$$

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

$$\left. \frac{\partial \ln(m_{jk})}{\partial \ln(V_{jl})} \right|_{\beta/\tau = \widehat{\beta/\tau}} = \left( -p_{jk|b(k)} \mathbf{x}'_{jk} \widehat{\beta/\tau}, -p_{jk} \mathbf{x}'_{jk} \widehat{\beta/\tau} \right] \quad (13)$$

The indirect elasticity in (13) represents an externality. We are measuring the effect of factors related to the attractiveness of an alternative destination  $l$ , included in  $V_{jl}$ , on the migration flows not between  $j$  and  $l$  but between  $j$  and  $k$ . Whenever one of this factors is the migration policy set by  $l$  on the potential migrants coming from  $j$ , we will be quantifying the externality effect of migration policies.

### 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 (12) and (13) for the direct and indirect elasticity respectively. The summary statistics for the upper and lower bound of this direct elasticity can be observed in the upper panel of Table 3, while each dot in Figure 2 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

<sup>39</sup>Without loss of generality, we have ordered the extremes of the two intervals under the assumption that  $\mathbf{x}'_{jk} \widehat{\beta/\tau} \geq 0$ .

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, count 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 quite substantial. 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.

Table 3: Direct and indirect elasticities of networks and visa

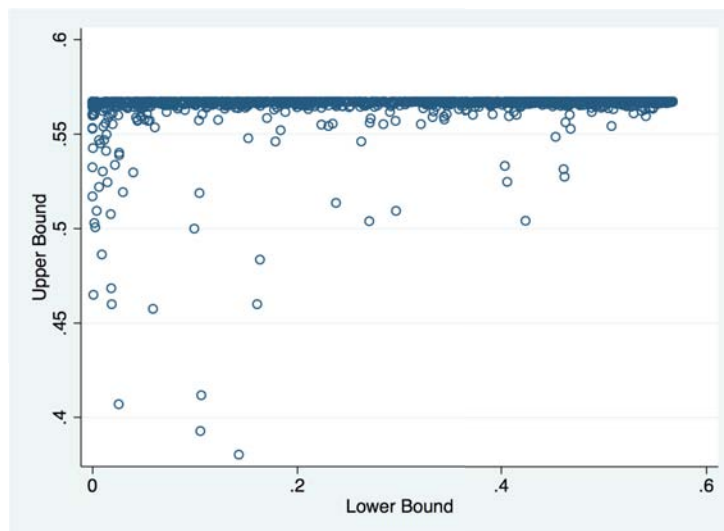
<i>Bound</i>	lower	upper
<i>Networks</i>		
Direct effect	0.459 (0.156)	0.567 (0.002)
Indirect effect	-0.108 (0.156)	0.000 (0.002)
<i>Visa</i>		
Direct effect	-0.473 (0.045)	-0.399 (0.131)
Indirect effect	0.028 (0.088)	0.169 (0.245)

Note: standard deviations in parentheses. The bounds correspond to averages, weighted by population at origin, over equations (12), (13) and (B.1)-(B.3) based on the estimates in specification (3) in Table 2.

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 (13) generates



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



Note: see Table 3 for the average values.

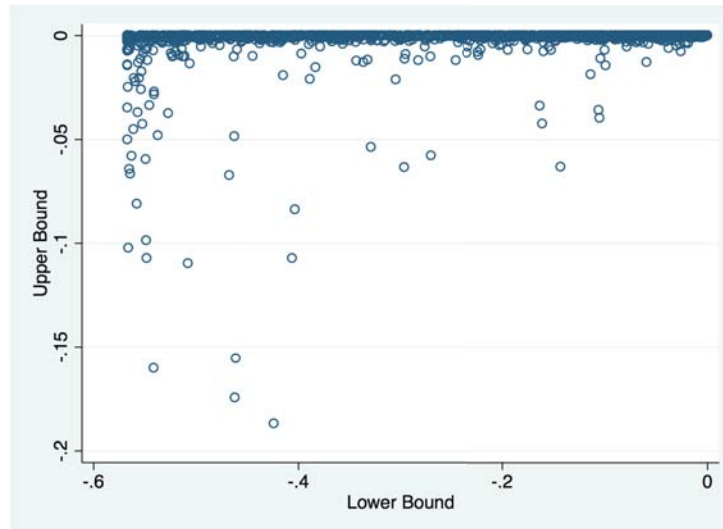
the bounds for the cross-elasticity that has typically been absent from the literature:<sup>40</sup> the elasticity of the migration flow from the origin  $j$  to the destination  $k$  with respect to the migration networks of  $j$  in another country  $l \in b(k)$ . The upper panel of Table 3 presents the averages of the upper and lower bound for this cross-elasticity, while Figure 3 represents the clouds of dyad-specific cross-elasticities.<sup>41</sup> The average upper bound for the cross-elasticity is almost zero. As for the lower bound, which corresponds to the largest within-nest correlation, the average cross-elasticity is higher in absolute value: -0.11.<sup>42</sup>

<sup>40</sup>Bertoli *et al.* (2013b) represent an exception in this respect.

<sup>41</sup>We only report here and in the next section within-nest cross-elasticities. Observe that (13) does not vary with  $l \in b(k)$ , so that we have the same number of direct and cross-elasticities. The cross-elasticity with a country out of the nest is not subject to uncertainty and it is just given by the upper bounds in Figure 3.

<sup>42</sup>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 (12) and (13) is independent from  $\tau$ . For instance, the lowest upper bounds in Figures 2 and 3 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 3: Bounds for the indirect elasticity of migration flows with respect to networks



Note: see Table 3 for the average values.

### 5.2.2 Visa effects

Differently from networks, the visa variable is dichotomous, so that we adjusted the formulas presented in Section 3 to account for the discrete nature of this variable, as shown in the Appendix B. 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.

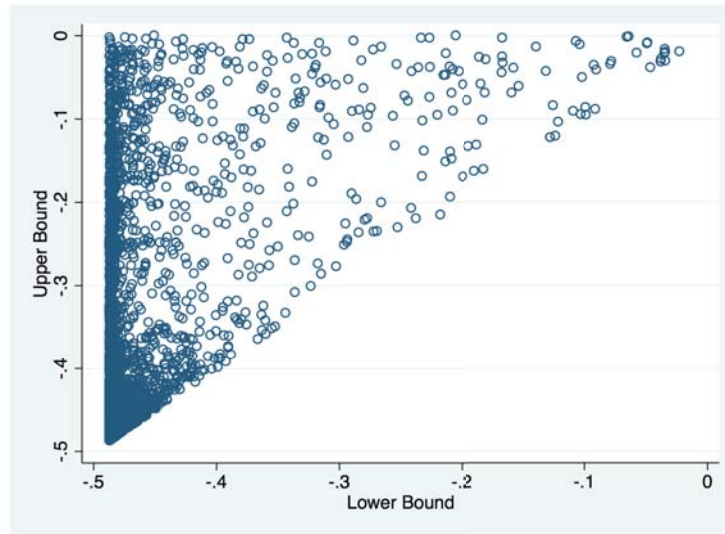
The most remarkable aspect that deserves to be commented about the direct and indirect effects of visas is their magnitude. The average bounds mean that we can expect the imposition of a visa requirement by country  $k$  on country  $j$  to correlate with a decrease of 40 to 47 percent of the level of migration flows from  $j$  to  $k$  with respect to the level that prevails when a visa waiver is applied.<sup>43</sup> 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 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 4, which

<sup>43</sup>Bertoli and Fernández-Huertas Moraga (2013) estimate that the introduction of a visa requirement reduces bilateral migration flows to Spain by up to 76 percent.

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).

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

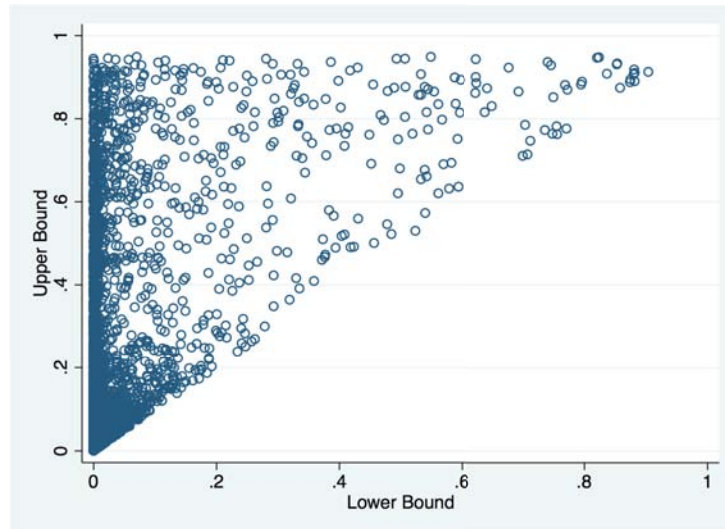


Note: see Table 3 for the average values.

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 Figure 5 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.

For instance, consider Canada, which received little more than 12,000 migrants from

Figure 5: Bounds for the indirect effect of the visa requirement on migration flows



Note: see Table 3 for the average values.

Mexico between 1990 and 2000; 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 indirect effect 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 minus 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 Mexicans.

## 6 Robustness

This section presents a number of robustness analyses on the main results presented in specification (3) of Table 2. First, we address the concern that our estimates of the effect of visa policies could be confounded by the unobserved cultural proximity between the origin and the destination country. Second, we tackle the concerns related to the measurement error in the visa variable, using data on gross migration flows to a subset of 15 destinations over 2005 and 2006 from Ortega and Peri (2013). Third, we re-estimate the models with

different samples. Fourth, we redefine the dependent variable in order to analyze separately the determinants of low and high-skill migration. Fifth, we propose an approach to deal with the potential endogeneity problem related to the inclusion of the networks variable.

## 6.1 Cultural proximity

Our interpretation of the estimated coefficient for the visa requirement in our benchmark specification is that this variable indirectly reveals the height of the policy-induced “cliff at the border,” as legal provisions for non-immigrant admission should exert no influence on the size of bilateral migration flows in the absence of severe restrictions on labor mobility across countries. An alternative interpretation of the estimated effect of visa policies is that it could actually be driven by the correlation between this variable and the (for us, unobserved) cultural proximity between countries: destination countries could be more likely to grant a visa waiver to the citizens of culturally close origin countries, with cultural proximity also exerting a direct influence on the scale of bilateral migration flows (Belot and Ederveen, 2012), so that our estimates would be picking up this spurious correlation.

The data by Neumayer (2006) give us a nice opportunity to deal with this relevant threat to identification, using data on the *reciprocal* visa policy: we can model the bilateral migration flow from country  $j$  to country  $k$  as a function of the visa requirement imposed by the origin country  $j$  on the citizens of country  $k$ . The underlying idea is that, say, the visa policy adopted by Ghana towards British citizens should have *no* impact on the scale of migration from Ghana to the United Kingdom, but this variable might be capturing the effect of cultural proximity between the two countries.<sup>44</sup>

Specifications (1) and (2) in Table 4 reveal that the reciprocal visa requirement is, as expected, negative but only marginally significant when we control for multilateral resistance to migration through the inclusion of origin-*nest* fixed effects. More interestingly, specifications (3) and (4) in Table 4 add the reciprocal visa policy to our benchmark specification: the estimated coefficient for the visa requirement imposed from the destination  $k$  to the origin  $j$  is virtually unaffected by the inclusion of the reciprocal visa policy, and it is negative and

---

<sup>44</sup>Once we partial out the correlation with the other regressors, the correlation between the visa policy of country  $k$  on  $j$  and the visa policy of country  $j$  on  $k$  stands at 0.23 (the correlation in the raw data stands at 0.62), so that the estimated effect of the reciprocal visa policy could also be reflecting this positive correlation.

Table 4: Determinants of migration flows (reciprocal visa policy)

<i>Specification</i>	(1)	(2)	(3)	(4)
<i>Dependent variable</i>	flow	flow	flow	flow
<i>Model</i>	PPML	PPML	PPML	PPML
ln(networks+1)	0.658*** [0.042]	0.560*** [0.050]	0.658*** [0.042]	0.563*** [0.050]
Visa requirement	-	-	0.017 [0.160]	-0.639*** [0.211]
Reciprocal visa requirement	0.017 [0.304]	-0.463* [0.253]	0.016 [0.302]	-0.436* [0.256]
Schengen	0.653* [0.379]	0.057 [0.233]	0.651* [0.381]	0.041 [0.235]
Colony	-0.292 [0.221]	0.464* [0.253]	-0.291 [0.221]	0.451* [0.255]
Common language	0.332** [0.130]	0.305* [0.161]	0.333** [0.131]	0.298* [0.160]
ln(distance)	-0.381*** [0.099]	-0.130 [0.116]	-0.382*** [0.099]	-0.123 [0.114]
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
Pseudo $R^2$	0.988	0.996	0.988	0.996
Log pseudo-likelihood	-4,294,724	-2,218,243	-4,294,666	-2,202,439
Pesaran (2004) CD test	17.35	-1.30	17.34	-1.70
p-value	0.000	0.193	0.000	0.089

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 is equal to the maximum between the variation in stocks and zero; standard errors are robust; the nest structure in specifications (2) and (4) is the same as in specification (3) in Table 2; the pseudo  $R^2$  is defined as one minus the ratio between the log-likelihood of the model over the log-likelihood of a restricted model which only includes a constant.

highly statistically significant in the specification that restores the cross-sectional independence of the residuals. This is reassuring about the ability of the visa policy to provide us with information about the height of the “cliff at the border,” as the estimated effect of the visa requirement does not appear to be driven by unobserved cultural proximity between countries.

## 6.2 Measurement error in the visa variable

Our visa variable refers to December 2004, while our proxy for gross migration flows has been measured between 1990 and 2000: this timing creates the concern that the measurement error in the visa variable could be systematically correlated with the dependent variable in our econometric analysis. Specifically, destination countries that used to grant a visa waivers to some origins in the 1990s and that faced a substantial immigration of visa over-stayers might have then imposed a visa requirement. This was, for instance, the case for Spain, which lifted the visa waiver that it used to grant to Colombian and Ecuadorian citizens in 2001 and 2003, after experiencing substantial immigration flows in the late 1990s from these two origin countries (Bertoli *et al.*, 2011; Bertoli and Fernández-Huertas Moraga, 2013). This type of policy change would clearly bias the estimated effect of visa requirements towards zero.

We deal this challenge by drawing on the data on yearly gross migration flows to 15 OECD countries by Ortega and Peri (2013) for the years 2005 and 2006.<sup>45,46</sup> This allows us to re-estimate our model using as dependent variable gross migration flows measured *subsequently* to our visa variable. Table 5 report the estimates obtained with OLS on the subset of origin-destination pairs with positive gross migration flows, and the estimates obtained with PPML on the full sample of 2,700 observations. The dummy for the visa requirement is negative and highly statistically significant in specification (3), which restores the cross-sectional independence of the residuals through the inclusion of origin times nest dummies, and the

---

<sup>45</sup>The 15 destination countries are: Australia, Belgium, Canada, Denmark, Finland, Germany, Italy, Netherlands, Norway, New Zealand, Spain, Sweden, Switzerland, United Kingdom and United States; they are all included in the dataset by Docquier *et al.* (2009), which represents our main data source for migration data.

<sup>46</sup>2005 and 2006 are the last two years with data on bilateral migration flows in the dataset by Ortega and Peri (2013). Our dependent variable aggregates over the two years in order to keep the same structure as in the previous estimates.

Table 5: Determinants of migration flows (2005-2006)

<i>Specification</i>	(1)	(2)	(3)
<i>Dependent variable</i>	ln(flow)	flow	flow
<i>Model</i>	OLS	PPML	PPML
ln(networks+1)	0.393*** [0.021]	0.597*** [0.037]	0.583*** [0.048]
Visa requirement	-0.244* [0.132]	-0.986*** [0.169]	-0.668*** [0.133]
Schengen	0.297 [0.271]	0.068 [0.210]	-0.118 [0.213]
Colony	0.788*** [0.208]	0.010 [0.194]	0.176 [0.187]
Common language	0.646*** [0.107]	0.474*** [0.130]	0.636*** [0.172]
ln(distance)	-0.581*** [0.063]	-0.429*** [0.082]	-0.479*** [0.114]
Destination fixed effects	Yes	Yes	Yes
Origin fixed effects	Yes	Yes	Yes
Origin*nest fixed effects	No	No	Yes
Observations	1,981	2,700	2,700
Adjusted (pseudo) $R^2$	0.864	0.907	0.941
Log pseudo-likelihood	-	-1,616,062	-908,216
Pesaran (2004) CD test	-	7.11	-0.90
p-value	-	0.000	0.370

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 refers to gross flows from Ortega and Peri (2013); the nest structure in specification (3) is the same as in specification (3) in Table 2, although with a reduced number of destinations; the pseudo  $R^2$  for specifications (2)-(3) is defined as one minus the ratio between the log-likelihood of the model over the log-likelihood of a restricted model which only includes a constant.



size of the estimated effect is in line with the one obtained in our benchmark specification in Table 2. Getting rid of the incorrect timing between the measurement of the migration flows and of the visa variable allows the effect of visa policies to emerge more clearly from the estimates: the coefficient of the visa requirement is also negative and significant, although only marginally so when the model is estimated with OLS, in the other two specifications, whose RUM-consistency is however either not testable or rejected by the test on the residuals.

### 6.3 Different samples

The results from Table 2 do not depend on the particular coverage of the dataset described in Table 1. Table 6 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 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 6. Specification (1) does not include origin-nest fixed effects and the Pesaran *CD* test shows that the residuals could be cross-sectionally 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). 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.<sup>47</sup> We are then left with 4,708 observations. Again, we reject the cross-sectional 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 cross-sectional 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 smaller in magnitude than in the baseline specification

---

<sup>47</sup>This specification omits the Schengen variable, as the restriction of the sample leaves us with no variability in the data to identify its effect.

Table 6: Determinants of migration flows (1990-2000), different samples

<i>Specification</i>	<i>Population size</i>		<i>Developing countries</i>	
	(1)	(2)	(3)	(4)
<i>Dependent variable</i>	flow	flow	flow	flow
<i>Model</i>	PPML	PPML	PPML	PPML
ln(networks+1)	0.654*** [0.043]	0.569*** [0.051]	0.662*** [0.057]	0.609*** [0.065]
Visa requirement	0.016 [0.166]	-0.716*** [0.243]	0.0004 [0.178]	-0.358** [0.163]
Schengen	0.671* [0.381]	0.041 [0.236]	- -	- -
Colony	-0.282** [0.220]	0.462* [0.258]	-0.005 [0.213]	0.445 [0.341]
Common language	0.320** [0.133]	0.287* [0.162]	0.574*** [0.143]	0.261 [0.210]
ln(distance)	-0.381*** [0.100]	-0.116 [0.116]	-0.612*** [0.157]	-0.674*** [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
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 variation in stocks and zero; standard errors are robust; the pseudo  $R^2$  is defined as one minus the ratio between the log-likelihood of the model over the log-likelihood of a restricted model which only includes a constant.

from Table 2. Still, these differences are not statistically significant. We can also observe that the distance variable regains significance in this smaller sample.

## 6.4 Estimation by skill levels

In this part, we take advantage from the fact that the dataset by Docquier *et al.* (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 7 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 are taken into account, so that multilateral resistance to migration is controlled for. As before, we can disregard specification (1) on the basis of the *CD* test, clearly rejected with a statistic of 19.91.<sup>48</sup> 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 cross-sectional 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 7, we are focusing on low-skill migration. Again, specification (3), without origin-nest fixed effects, has problems of cross-sectional 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 cross-sectional 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.<sup>49</sup>

---

<sup>48</sup>This suggests that the disaggregation of migration flows by skill level does not suffice to remove the heterogeneity across potential migrants that induces a correlation in the stochastic component of utility.

<sup>49</sup>These findings are confirmed when we disaggregate the bilateral migration rates by skill level and by

Table 7: Determinants of migration flows by skill level (1990-2000)

<i>Specification</i>	<i>High-skill</i>		<i>Low-skill</i>	
	(1)	(2)	(3)	(4)
<i>Dependent variable</i>	flow	flow	flow	flow
<i>Model</i>	PPML	PPML	PPML	PPML
ln(networks+1)	0.615*** [0.038]	0.496*** [0.037]	0.703*** [0.052]	0.608*** [0.056]
Visa requirement	-0.073 [0.131]	-0.559*** [0.238]	0.110 [0.213]	-0.718*** [0.268]
Schengen	0.629** [0.281]	-0.305* [0.178]	0.960* [0.565]	0.879* [0.467]
Colony	-0.152 [0.162]	0.327 [0.221]	-0.238 [0.260]	0.652** [0.279]
Common language	0.548*** [0.114]	0.619*** [0.131]	0.084 [0.158]	0.048 [0.207]
ln(distance)	-0.175* [0.092]	-0.220** [0.088]	-0.470*** [0.120]	-0.053 [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
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 *et al.* (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 variation in stocks and zero; standard errors are robust.

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 8, based in specifications (2) and (4) from Table 7. 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.

## 6.5 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 *et al.* (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 *et al.* (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,<sup>50</sup> given that the Poisson model is non-linear and 2SPS would generally be inconsistent (Terza *et al.*, 2008).

The results are presented in Table 9 for several correct specifications discussed in the paper.<sup>51</sup> 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 

---

gender; the visa effect for both genders is larger in magnitude for low- than for high-skilled individuals, and slightly larger for men than for women. Results are available from the authors upon request.

<sup>50</sup>Recent applications of the 2SRI estimator within the migration literature can be found in Beine *et al.* (2012), Marchetta (2012) and Bertoli and Marchetta (2015).

<sup>51</sup>We present bootstrapped standard errors after 1,000 replications to account for the two-step estimation.

Table 8: Direct and indirect elasticities of networks and visa by skill level

Flow	<i>High-skill</i>		<i>Low-skill</i>	
	<i>Bound</i>	lower	upper	lower
	<i>Networks</i>			
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)
	<i>Visa</i>			
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 (12), (13) and (B.1)-(B.3) based on the estimates in specifications (2) and (4) in Table 7.

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, coinciding with the result in Beine *et al.* (2011). A possible interpretation of the direction of the bias, which is also reflected in the negative and significant coefficient for the first stage residuals, relates to return migration: a larger network can be associated with a larger scale of return migration, which influences a dependent variable that captures variation in stocks rather than gross migration flows. As for the visa, the coefficient remains negative and significant at the 90 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 whereas the visa coefficient is virtually unaffected although the larger standard errors make it marginally sig-

Table 9: Determinants of migration flows (1990-2000), two-stage residual inclusion

<i>Sample</i>	<i>Baseline</i>	<i>Population</i>	<i>Developing</i>	<i>High-</i>	<i>Low-</i>
<i>Specification</i>	(1)	<i>Size</i>	<i>Countries</i>	<i>Skill</i>	<i>Skill</i>
<i>Dependent variable</i>	flow	flow	flow	flow	flow
<i>Model</i>	2SRI PPML	2SRI PPML	2SRI PPML	2SRI PPML	2SRI PPML
ln(networks+1)	0.766*** [0.095]	0.780*** [0.096]	0.791*** [0.130]	0.661*** [0.078]	0.799*** [0.112]
Visa requirement	-0.621* [0.462]	-0.682* [0.559]	-0.329 <sup>a</sup> [0.327]	-0.543* [0.473]	-0.653 <sup>b</sup> [0.498]
Schengen	-0.032 [0.334]	0.016 [0.340]	- -	-0.325 [0.244]	0.810* [0.547]
Colony	-0.082 [0.363]	-0.042 [0.360]	-0.004 [0.467]	-0.163 [0.289]	0.177 [0.402]
Common language	0.098 [0.236]	0.051 [0.243]	0.037 [0.348]	0.469** [0.208]	-0.161 [0.308]
ln(distance)	0.191 [0.198]	0.234 [0.203]	-0.240 [0.421]	0.020 [0.147]	0.264 [0.236]
First stage residual	-0.248** [0.099]	-0.265** [0.101]	-0.224* [0.135]	-0.208*** [0.078]	-0.237* [0.125]
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
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: bootstrapped standard errors after 1,000 replications in brackets; \*\*\* significant at the 99 percent level, \*\* significant at the 95 percent level, \* significant at the 90 percent level, <sup>a</sup> test that the coefficient is positive p-value = 0.115, <sup>b</sup> test that the coefficient is positive p-value = 0.100. Specification (1) as specification (3) in Table 2; specifications (2) and (3) and specifications (2) and (4) in Table 6; specifications (4) and (5) as specifications (2) and (4) in Table 7. The dependent variable is equal to the maximum between the variation in stocks and zero; the excluded instrument is the natural logarithm of one plus the size of migration networks in 1960 (see Table 1).

nificant. None of the specifications rejects the null of cross-sectional independence in the residuals.

## 7 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 by bilateral visa policies on international migration flows, which can be identified only when the confounding effect due to 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 changing its policy 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 (2013), and are based on an estimation technique with minimal data requirements, which is well-suited for most of the existing international migration datasets.

While we have shown our results to be quite robust to a number of variations in our estimation strategy, it is true that our estimated externality depends on the precise definition of the nests of alternative destinations that potential migrants perceive as close substitutes. An obvious topic for further research is to make the choice of these nests completely data-driven along the lines proposed by Bonhomme and Manresa (2012) and Bai and Ando (2013) for the case of non-linear estimation techniques while restricting the errors of the resulting models to be cross-sectionally independent.



## References

- ANDERSON, J. (2011): “The Gravity Model,” *Annual Review in Economics*, 3(1), 133–160.
- ANDERSON, J., AND E. VAN WINCOOP (2003): “Gravity with Gravitas: A Solution to the Border Puzzle,” *The American Economic Review*, 93(1), 170–192.
- BAI, J., AND T. ANDO (2013): “Panel data models with grouped factor structure under unknown group membership,” MPRA Paper 52782, University Library of Munich, Germany.
- BEINE, M., S. BERTOLI, AND J. FERNÁNDEZ-HUERTAS MORAGA (2014): “A practitioners’ guide to gravity models of migration,” *The World Economy*, forthcoming.
- BEINE, M., F. DOCQUIER, AND Ç. ÖZDEN (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*, 42(5), 844–857.
- BEINE, M., AND C. PARSONS (2014): “Climatic factors as determinants of International Migration,” *Scandinavian Journal of Economics*, DOI: 10.1111/sjoe.12098.
- 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*, 114(4), 1105–1128.
- BERTOLI, S., H. BRÜCKER, AND J. FERNÁNDEZ-HUERTAS MORAGA (2013a): “The European Crisis and Migration to Germany: Expectations and the Diversion of Migration Flows,” CERDI Discussion Paper No. 2013/21, Clermont-Ferrand.
- BERTOLI, S., AND J. FERNÁNDEZ-HUERTAS MORAGA (2013): “Multilateral Resistance to Migration,” *Journal of Development Economics*, 102, 79–100.
- BERTOLI, S., J. FERNÁNDEZ-HUERTAS MORAGA, AND F. ORTEGA (2011): “Immigration Policies and the Ecuadorian Exodus,” *World Bank Economic Review*, 25(1), 57–76.

- (2013b): “Crossing the Border: Self-Selection, Earnings and Individual Migration Decisions,” *Journal of Development Economics*, 101(1), 75–91.
- BERTOLI, S., AND F. MARCHETTA (2015): “Bringing It All Back Home: Return migration and fertility choices,” *World Development*, 65, 27–40.
- BOERI, T., AND H. BRÜCKER (2005): “Why Are Europeans So Tough on Migrants?,” *Economic Policy*, 20(44), 631–703.
- BONHOMME, S., AND E. MANRESA (2012): “Grouped Patterns Of Heterogeneity In Panel Data,” CEMFI Working Paper No. 2012-1208, Madrid.
- BORJAS, G. J. (1987): “Self-Selection and the Earnings of Immigrants,” *American Economic Review*, 77(4), 531–553.
- BRATSBERG, B., J. E. DØLVIK, AND O. RAAUM (2012): “Economic Shocks, the Legal Environment and Work-related Migration,” paper presented at the VI Workshop on Migration and Labor Economics organized by INSIDE-MOVE, NORFACE and CReAM, IAE (CSIC), October 18-19.
- 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.
- CHISWICK, B. R. (1988): “Illegal Immigration and Immigration Control,” *Journal of Economic Perspectives*, 2(3), 101–115.
- 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?,” *Journal of Economic Perspectives*, 25(3), 83–106.
- DE HOYOS, R., AND V. SARAFIDIS (2006): “Testing for cross-sectional dependence in panel-data models,” *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*, 50(3), 681–730.
- EBERHARDT, M. (2011): “XTCD: Stata module to investigate Variable/Residual Cross-Section Dependence,” Statistical Software Components, Boston College Department of Economics.
- EGGER, P., M. LARCH, K. E. STAUB, AND R. WINKELMANN (2011): “The Trade Effects of Endogenous Preferential Trade Agreements,” *American Economic Journal: Economic Policy*, 3(3), 113–143.
- FREES, E. W. (1995): “Assessing cross-sectional correlation in panel data,” *Journal of Econometrics*, 69(2), 393–414.
- GIORDANI, P. E., AND M. RUTA (2013): “Coordination failures in immigration policy,” *Journal of International Economics*, 89(1), 55–67.
- 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.
- 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.
- HENSHER, D. A., AND W. H. GREENE (2003): “The Mixed Logit model: The state of practice,” *Transportation*, 30(2), 133–176.

- 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.
- 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.
- MARCHETTA, F. (2012): “Return migration and the survival of entrepreneurial activities in Egypt,” *World Development*, 40(10), 1999–2013.
- 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 (2013): “Distortions in the International Migrant Labor Market: Evidence from Filipino Migration and Wage Responses to Destination Country Economic Shocks,” *American Economic Journal: Applied Economics*, 6(2), 49–75.
- 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.
- OKAWA, Y., AND E. VAN WINCOOP (2012): “Gravity in International Finance,” *Journal of International Economics*, 87(2), 205–215.
- ORTEGA, F., AND G. PERI (2013): “The Role of Income and Immigration Policies in Attracting International Migrants,” *Migration Studies*, 1(1), 27–46.
- ÖZDEN, Ç., 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.
- PERKINS, R., AND E. NEUMAYER (2013): “Geographies of educational mobilities: exploring the uneven flows of international students,” *Geographical Journal*, 180(3), 246–259
- PESARAN, M. H. (2004): “General Diagnostic Tests for Cross Section Dependence in Panels,” IZA Discussion Paper No. 1204, Bonn.
- 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. BRÜLHART (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 (2013): “The Short- and Long-Run Determinants of Unskilled Immigration into U.S. States,” *Southern Economic Journal*, 80(2), 414–438.
- 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.

# Appendix

## A Estimates under different nest structures

Table A.1: Determinants of migration flows (1990-2000), different nests

<i>Specification</i>	(1)	(2)	(3)	(4)	(5)	(6)
ln(networks+1)	0.658*** [0.042]	0.645*** [0.042]	0.627*** [0.045]	0.577*** [0.047]	0.570*** [0.048]	0.567*** [0.049]
Visa requirement	0.017 [0.161]	0.064 [0.235]	-0.706*** [0.184]	-0.666*** [0.211]	-0.663*** [0.212]	-0.667*** [0.215]
Schengen	0.651* [0.381]	0.286 [0.237]	0.270 [0.242]	0.288 [0.242]	0.290 [0.244]	0.034 [0.235]
Colony	-0.290 [0.217]	0.060 [0.221]	0.148 [0.229]	0.444* [0.247]	0.455* [0.249]	0.451* [0.256]
Common language	0.333** [0.130]	0.225 [0.154]	0.260* [0.145]	0.221 [0.153]	0.229 [0.154]	0.302* [0.161]
ln(distance)	-0.382*** [0.098]	0.027 [0.106]	-0.012 [0.113]	-0.115 [0.113]	-0.113 [0.116]	-0.121 [0.116]
Destination fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Origin*nest fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,611	5,611	5,611	5,611	5,611	5,611
Pseudo $R^2$	0.988	0.993	0.994	0.995	0.995	0.996
Log pseudo-likelihood	-4,294,695	-3,160,509	-2,667,687	-2,340,175	-2,323,708	-2,213,844
Pesaran (2004) CD test	17.35	5.52	3.92	3.82	3.30	-1.57
p-value	0.000	0.000	0.000	0.000	0.001	0.117
<i>Assignment of destination countries to nests</i>						
Austria	1	1	1	1	1	1
Belgium	1	1	1	1	1	1
Denmark	1	1	1	1	1	1
Finland	1	1	1	1	1	1
France	1	1	1	1	1	1
Germany	1	1	1	1	1	1
Greece	1	1	1	1	1	1
Ireland	1	1	1	1	1	1
Italy	1	1	1	1	1	1
Luxembourg	1	1	1	1	1	1
Netherlands	1	1	1	1	1	1
Portugal	1	1	1	1	1	1
Spain	1	1	1	1	1	1
Sweden	1	1	1	1	1	1
United Kingdom	1	1	1	1	1	1
Iceland	1	1	1	1	2	2
Norway	1	1	1	1	2	2
Switzerland	1	1	1	1	2	2
Czech Republic	1	1	1	1	2	3
Hungary	1	1	1	1	2	3
Poland	1	1	1	1	2	3
Slovakia	1	1	1	1	2	3
Canada	1	2	2	2	3	4
United States	1	2	2	2	3	4
Australia	1	2	2	3	4	5
Japan	1	2	2	3	4	5
New Zealand	1	2	2	3	4	5
Korea	1	2	3	4	5	6
Mexico	1	2	3	4	5	6
South Africa	1	2	3	4	5	6
Turkey	1	2	3	4	5	6

Note: the dependent variable is the size of the flow, and all specifications are estimated with PPML; 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 is equal to the maximum between the variation in stocks and zero; standard errors are robust.

## B Direct and indirect effects of visas on migration flows

Let  $\widehat{V}_{jk}/\tau$  represent the estimated value of the deterministic component of location-specific utility, and  $\widehat{V}_{jk}^1/\tau$  and  $\widehat{V}_{jk}^0/\tau$  represent the corresponding values when a visa requirement is imposed from country  $k$  upon country  $j$  and when a visa waiver is granted. The ratio  $\widehat{v}_{kjk}$  between the size of the bilateral migration flow conditional upon  $\widehat{V}_{jk}^1/\tau$  and the size of the flow conditional upon  $\widehat{V}_{jk}^0/\tau$  measures the effect of the introduction of a visa requirement by country  $k$  on  $m_{jk}$ . Similarly, we can define  $\widehat{v}_{kjl}$  as the impact on  $m_{jl}$  due to the imposition of a visa requirement by country  $k$  upon the citizens of country  $j$ .

From (5), we have that:<sup>52</sup>

$$\widehat{v}_{kjk} = e^{\widehat{\beta}_1/\tau} \left[ \frac{\sum_{l \in b(k)/\{k\}} e^{\widehat{V}_{jl}/\tau} + e^{\widehat{V}_{jk}^0/\tau}}{\sum_{l \in b(k)/\{k\}} e^{\widehat{V}_{jl}/\tau} + e^{\widehat{V}_{jk}^1/\tau}} \right]^{1-\tau} \left[ \frac{\sum_{l \in D/\{k\}} e^{\widehat{V}_{jl}/\tau} + e^{\widehat{V}_{jk}^0/\tau}}{\sum_{l \in D/\{k\}} e^{\widehat{V}_{jl}/\tau} + e^{\widehat{V}_{jk}^1/\tau}} \right]^\tau$$

$\widehat{v}_{kjk}$ , which is monotonic in  $\tau$  for any given estimate  $\widehat{\beta}/\tau$  of the vector  $\beta/\tau$ , is equal to a weighted geometric average of the effect of the introduction of a visa requirement when  $\tau$  converges to 0 and when  $\tau = 1$ . When  $\tau$  converges to 0,  $\widehat{v}_{kjk}$  converges to:

$$\widehat{v}_{kjk}|_{\tau=0} = e^{\widehat{\beta}_1/\tau} \frac{\sum_{l \in b(k)/\{k\}} e^{\widehat{V}_{jl}/\tau} + e^{\widehat{V}_{jk}^0/\tau}}{\sum_{l \in b(k)/\{k\}} e^{\widehat{V}_{jl}/\tau} + e^{\widehat{V}_{jk}^1/\tau}} \quad (\text{B.1})$$

When  $\tau = 1$ , this becomes:

$$\widehat{v}_{kjk}|_{\tau=1} = e^{\widehat{\beta}_1/\tau} \frac{\sum_{l \in D/\{k\}} e^{\widehat{V}_{jl}/\tau} + e^{\widehat{V}_{jk}^0/\tau}}{\sum_{l \in D/\{k\}} e^{\widehat{V}_{jl}/\tau} + e^{\widehat{V}_{jk}^1/\tau}} \quad (\text{B.2})$$

If  $\widehat{\beta}_1/\tau < 0$ , then the effect is larger in magnitude when  $\tau$  converges to zero (upper bound) than when  $\tau = 1$  (lower bound).<sup>53</sup> With respect to the effect  $\widehat{v}_{kjl}$  due to the introduction of a visa requirement by another country  $l$  upon  $m_{jk}$ , we have that

$$\widehat{v}_{kjl} \in \left[ \frac{\sum_{l \in D/\{k\}} e^{\widehat{V}_{jl}/\tau} + e^{\widehat{V}_{jk}^0/\tau}}{\sum_{l \in D/\{k\}} e^{\widehat{V}_{jl}/\tau} + e^{\widehat{V}_{jk}^1/\tau}}, \frac{\sum_{l \in b(k)/\{k\}} e^{\widehat{V}_{jl}/\tau} + e^{\widehat{V}_{jk}^0/\tau}}{\sum_{l \in b(k)/\{k\}} e^{\widehat{V}_{jl}/\tau} + e^{\widehat{V}_{jk}^1/\tau}} \right) \quad (\text{B.3})$$

<sup>52</sup>Without loss of generality, we denote with  $\beta_1$  the parameter in (2) associated to the variable denoting the bilateral visa policy.

<sup>53</sup>The percentage change induced by the introduction of a visa requirement is simply  $\widehat{v}_{kjk} - 1$ .