Regional *vs* long-distance international migration: The case of South American emigrants*

Gilles Spielvogel†

December 2014

Abstract

Regional migration represents about one third of all international migration. In this paper, we study the determinants of regional vs long-distance international migration, emphasizing the relative selectivity of these two types of mobility. We develop a model of destination choice accounting for both unobserved individual heterogeneity and different marginal utility of income for different destination types. We estimate this model using exhaustive data on emigration from South American countries. We find that the independence of irrelevant alternatives (IIA) assumption, commonly used in the literature, is not compatible with migration choices made by South Americans. Our estimates show that regional migration and long-distance migration are indeed different: the former is less selective and much more sensitive to international wage differentials.

JEL codes: F22, J61, O15, C25

Keywords: International migration, Economic determinants, Random utility model, Nested logit, South America.

*I am grateful to Sarah Widmaier for her participation in early stages of this research, and to Jean-Christophe Dumont for his thoughtful comments and suggestions on multiple versions of this paper. I also thank Philippe De Vreyer, Hubert Jayet, Hillel Rapoport, and especially Flore Gubert for helpful comments, as well as seminar participants at the University of Lille, São Paulo School of Economics, DIAL (Paris), and the 2013 Norface Migration Conference. The usual disclaimer applies.

The database DIOC-E used in this paper was compiled as the result of a joint effort of the OECD, the Development Research Group of the World Bank and the Research Department of the Agence Française de Développement and benefited from a special contribution from Afristat. This project has been supported by the Research Department of the Agence Française de Développement.

†University Paris 1 Panthéon-Sorbonne – IEDES, UMR Développement et Sociétés (U. Paris 1 - IRD) and UMR DIAL (U. Paris Dauphine - IRD). E-mail: gilles.spielvogel@univ-paris1.fr.

1 Introduction

According to recent estimates, there were about 220 million international migrants in the world in 2010 – about 3.2% of the total population (United Nations Population Division, 2013).[1] This global migrant population is strongly concentrated: 20% of all migrants live in the United States, 30% in European countries, while the remaining 50% are scattered throughout the rest of the world. Many migrants, however, do not travel far: about 31% of international migrants live in the part of the continent ("UN subregion") where they were born.[2]

The regional dimension of international migration has not been absent from academic and policy discussions, and attracts a growing attention. Focusing on developing areas, some early descriptive works include Gould (1974) on Sub-Saharan Africa, Kritz and Gurak (1979) on Latin America, and Seccombe (1985) on the Middle East. More recently, Bakewell (2009) has documented the predominance of regional migration in West Africa, as well as the central role played by South Africa in the regional labour migration system. According to Ratha and Shaw (2007), "South-South migration is overwhelmingly intraregional". Although these works usually note that the determinants of regional migration are likely to be different than those of long-distance moves, they do not offer any empirical investigation of this hypothesis. As far as we know, there is no rigorous answer in the economic literature to this basic question: What differentiates regional migrants from those who choose farther destinations?

In this paper, we study the determinants of regional vs long-distance international migration, emphasizing the relative selectivity associated with these two types of mobility. More specifically, using the *Database on Immigrants in OECD and non-OECD Countries* (DIOC-E) which provides data on immigrants stocks for the year 2000 (Dumont et al., 2010), we analyse emigration from 10 South American countries towards

[1] Unless otherwise specified, all the figures given in this section refer to the year 2010 and were computed using data from United Nations Population Division (2013).

[2] This proportion varies quite a lot across countries: for about one third of origin countries, the share of migrants living in their subregion of birth is above 50%, while it is below 10% for another third. A contiguity-based measure of proximity gives a similar result: about 34% of all international migrants live in a country that shares a border with their country of origin. There are 22 UN subregions, such as Eastern Africa, Caribbean, Western Asia or Southern Europe, see https://unstats.un.org/unsd/methods/m49/m49regin.htm for the complete list.

(almost) the whole world.[3]

Regional migration in South America is on par with the world average, with 31% of its emigrants living in the region. Compared to other developing regions, South American countries have a larger linguistic and cultural proximity, which should enhance labour mobility among them. While the existing regional agreements (Mercosur and the Andean Community) have some labour mobility provisions, they are relatively recent and their implementation has been quite slow, so that they are unlikely to have affected the migration stocks as of 2000, which result from flows which have taken place in the previous decades. In addition, the second half of the twentieth century has been marked in most South American countries by episodes of authoritarian rule, which have generally had restrictive immigration – and sometimes emigration – policies, thus hindering regional mobility in the region (Durand and Massey, 2010). However, in recent years, migration has increased within the region, as well as towards specific destination countries that are much farther away, such as the United States and Spain. South American countries thus constitute a particularly interesting case for the analysis of regional vs long-distance migration choices. Although our discussion and results only deal with South American emigration patterns, we believe that they would also constitute a relevant starting point to analyse emigration from other developing and emerging regions.

The lack of systematic quantitative research on regional migration and on its determinants hinges on the paucity of data on South-South migration. Indeed, in order to study the characteristics of international migrants in a worldwide comparative perspective, data are usually collected where migrants live. Since migrants are very numerous, and data more easily available, in OECD countries, existing international migration databases were until recently restricted to those destination countries (Docquier and Marfouk, 2006; OECD, 2008). As a result, most macro-analyses of the determinants of international migration focus on migrants living in OECD countries and therefore only deal with South-North and North-North migration (see e.g. Beine et al., 2011; Grogger

[3]On the origin side, the 10 largest countries of the continent are considered: Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay, and Venezuela. Only the two smallest South American countries, Guyana and Suriname, are not included in our analysis because of the lack of available data. The set of destination countries considered is discussed below in section 3.

and Hanson, 2011; Belot and Hatton, 2012; Hanson and McIntosh, 2012; Ortega and Peri, 2013; Bertoli and Fernandez-Huertas Moraga, 2013).

This focus on migration towards OECD countries implies that we know comparatively little about the size and characteristics of South-South migration in general, and regional migration in particular. This is problematic at the macro level because of the demographic and economic significance of many South-South migration corridors.[4] It may also lead to unwarranted generalizations regarding the determinants of international migration: factors that are found to explain migration towards OECD countries might not have the same relevance or magnitude when it comes to South-South migration. In particular, the selective nature of migration, which has been emphasized in the literature on South-North flows (see e.g. Grogger and Hanson, 2011; Belot and Hatton, 2012), has not yet been analysed in the case of South-South corridors. Moreover, regional migration could be partly explained by specific factors that are mostly irrelevant for long-distance flows, such as weak border controls or transnational ethnic groups.

In addition to their restricted geographical scope, previous studies which look at the determinants of migration towards OECD countries also raise a methodological issue. In those papers, the empirical relationships are derived from some microeconomic model of location choice. It is usually assumed – more or less explicitly – that the independence of irrelevant alternatives (IIA) holds, i.e. that the probability of choosing one destination over another is independent of the choice set.[5] If this is indeed the case, omitting destinations from the choice set might not matter too much. The IIA assumption, however, can only be tested if the complete choice set is considered, which is not the case because the set of possible destinations is restricted to countries for which data on the number of immigrants is available. In addition, if one starts from a small group of relatively homogenous countries (such as OECD countries), expanding the set of destinations to less similar countries decreases the likelihood that the IIA holds. Since

[4] Fortunately, this knowledge gap is currently being resolved thanks to several new migration databases that have extended the set of destination countries (Dumont et al., 2010; Ozden et al., 2011; United Nations Population Division, 2013). We discuss the Dumont et al. (2010) data in more detail in section 3.

[5] An important exception is the paper by Bertoli and Fernandez-Huertas Moraga (2013) which proposes a random utility maximization model without the IIA hypothesis. However, it focuses on migration flows towards a single destination country (Spain) and therefore does not study the determinants of migrants' choices towards competing destinations.

we deal in this paper with the full set of possible destination choices, it does not seem reasonable to assume a priori that the IIA hypothesis holds.

We therefore estimate a location choice model in a nested logit framework, where destination countries are grouped in different subsets according to their geographical location. This nested logit framework enables us to test for the IIA hypothesis properly. We assess the robustness of our results by applying various types of changes to the nesting structure.

Since wage differentials are a key determinant of migration, their accurate measurement is central to empirical studies on the determinants of international migration. Unfortunately, internationally comparable wage data are notoriously difficult to compile, all the more when developing or middle-income countries are concerned. Some studies, which do not look at selectivity, use GDP per capita to proxy for wages (e.g. Ortega and Peri, 2013). When disaggregating migration data by education level, as we do in this paper, it is however necessary to have wage measures by education level as well. In a similar setting, Grogger and Hanson (2011) estimate "education-specific" wages using country data on inequality (Gini index) and a parametric assumption on the income distribution. They assume that low-skilled individuals earn a wage equivalent to the 20th percentile of the income distribution, while high-skilled workers have earnings equivalent to the 80th percentile. The relevance of this assumption is unwarranted because it implies that the correspondence between the distribution of income and that of education is the same in all countries, which is unlikely.

In this paper, we adopt a more flexible approach where we match the quantiles of the income distribution to the share of each category of the education distribution. This procedure permits to better account for the cross-country differences in the educational structure of the population.

The first key result of this paper is methodological: we show that the IIA hypothesis is not compatible with migration choices made by South Americans. Wrongly assuming that it holds upwardly biases estimates of the marginal utility of income, especially for higher educated individuals. The second result is that the determinants of regional migration in South America are different from long-distance migration: regional migration is less selective and much more sensitive to international wage differentials.

The remainder of this paper is organized as follows. In section 2, we present our model of destination choice. In section 3, we present the data on South American emigrants in the world. The empirical framework is exposed in section 4, and the results are presented in section 5. Section 6 concludes.

2 A model of destination choice

We model individual destination choice as the outcome of a simple random utility discrete choice model (McFadden, 1977). An individual i born in country o may choose any destination d in a set $\mathcal{D}$ (including o). This choice is based on utility maximization: if individual i may obtain utility U_{id} in country d, he/she will choose country d if $U_{id} > U_{ic}$ ($d \neq c$ and $d, c \in \mathcal{D}$). The true utility U_{id} is however unknown and one can only observe its deterministic component $V_{id} = U_{id} - \varepsilon_{id}$. Therefore, individual i chooses destination d with probability P_{id}:

$$\begin{aligned} P_{id} &= Pr(U_{id} > U_{ic}), \forall d \neq c \\ &= Pr(V_{id} + \varepsilon_{id} > V_{ic} + \varepsilon_{ic}) \\ &= Pr(V_{id} - V_{ic} > \varepsilon_{ic} - \varepsilon_{id}) \end{aligned}$$

As is well known, different assumptions about the distribution of ε lead to different empirical models of destination choice (see e.g. Train, 2009). This broad framework, developed by Borjas (1987) on the basis of the Roy model (Roy, 1951), is ubiquitous in the migration literature on destination choice.[6]

More specifically, let us assume that the utility of an individual i born in o and living in $d \in \mathcal{D}$ is:

$$U_{iod} = \alpha W_{id} - \beta C_{iod} + \varepsilon_{iod} \tag{1}$$

where W_{id} is the expected wage of individual i in country d. C_{iod} represents the cost of migrating from o to d. This cost may vary according to some individual characteristics

[6] See for example Falaris (1987) and Dahl (2002) for microeconometric approaches (which also consider selectivity) and Pellegrini and Fotheringham (2002) for a survey of geography works on the subject.

and may have various components, some being destination-specific and others being variable by (o, d) pair. We assume that staying in the origin country is costless: $C_{ioo} = 0$. For now, for the sake of simplicity, we assume that the marginal utility of income, α, and the marginal disutility of migration costs, β, are the same for all individuals and country pairs. We will relax this assumption below. The variable ε_{iod} represents unobserved individual-specific characteristics. We assume that this last component is iid and follows a type-I extreme value (Gumbel) distribution: $F(\varepsilon) = \exp(-\exp(-\varepsilon))$.

We further assume that potential destinations can be grouped into $G+1$ exhaustive and mutually exclusive subsets $\mathcal{D}_g$ (with $g = 0, 1, \ldots, G$). The "destination" of non-migrants, i.e. their country of origin ($d = o$), is assumed to be the unique element of the subset $\mathcal{D}_0$.

Following Cardell (1997) and Berry (1994), let us now decompose the unobserved idiosyncratic component of utility ε_{iod} into two terms: ζ_{ig}, which is common to all destinations in the subset $\mathcal{D}_g$, and ν_{iod} which is iid and has a Gumbel distribution:

$$U_{iod} = \alpha W_{id} - \beta C_{iod} + \zeta_{ig} + (1-\sigma)\nu_{iod} \tag{2}$$

The term $\sigma \in [0, 1[$ is a measure of the within-group correlation of idiosyncratic utility.[7] As shown by Cardell, if $\nu \sim$ ext. value and $\zeta \sim \mathcal{C}(1-\sigma)$ (where $\mathcal{C}$ is the conjugate to the Type I extreme-value distribution), then $\varepsilon = \zeta + (1-\sigma)\nu \sim$ ext. value.

There are two polar cases to consider. First, if $\sigma = 0$, there is no within-group correlation among destinations and $\zeta = 0$. It is therefore unnecessary to nest destinations and the standard conditional logit model is obtained. Second, if $\sigma \to 1$, this means that destinations within a group are extremely similar. Choices between countries within a group are not discriminant and the analysis should rather be done at the group level. In between, for $0 < \sigma < 1$, the within-group correlation is positive but the across-groups correlation is null, which gives rise to the nested logit model.

As is well known, the conditional logit suffers from its reliance on the independence of irrelevant alternatives (IIA) hypothesis. If $\sigma = 0$ (and $\varepsilon \sim$ ext. value), the probability

[7] In theory, we could write σ_g instead of σ and have a specific within correlation for each group of destinations. However, as will become clear below, it is not possible to identify empirically group-specific correlations using aggregate data. We therefore assume that this correlation is the same across all groups. Alternatively, σ can be thought of as a weighted average of the different σ_g.

of choosing destination d is:

$$P_{iod} = \frac{e^{\alpha W_{id} - \beta C_{iod}}}{\sum_c e^{\alpha W_{ic} - \beta C_{ioc}}}.$$

As a result, the probability ratio of choosing between destinations d and c depends only on the attributes of d and c, and not on the attributes of other destinations:

$$\frac{P_{iod}}{P_{ioc}} = \frac{e^{\alpha W_{id} - \beta C_{iod}}}{e^{\alpha W_{ic} - \beta C_{ioc}}}.$$

This result implies that a new destination similar to d would not change the probability ratio of choosing c vs d, i.e. the IIA hypothesis, which is questionable.

In the nested logit model, the probability of choosing $d \in \mathcal{D}_g$ is:

$$P_{iod} = P_{iod|g} P_{iog} = \frac{e^{(\alpha W_{id} - \beta C_{iod})/(1-\sigma)}}{D_g^{\sigma} \left[\sum_g D_g^{(1-\sigma)} \right]} \tag{3}$$

where P_{iog} is the probability of choosing any country in $\mathcal{D}_g$, $P_{iod|g}$ is the probability of choosing country d conditionally on having chosen the group $\mathcal{D}_g$, and the term D_g is:

$$D_g = \sum_{d \in \mathcal{D}_g} e^{(\alpha W_{id} - \beta C_{iod})/(1-\sigma)}. \tag{4}$$

In this case, the probability ratio of choosing $c \in \mathcal{D}_h$ vs. $d \in \mathcal{D}_g$ depends on attributes of c and d, but also on those of the other choices:

$$\frac{P_{iod}}{P_{ioc}} = \frac{e^{(\alpha W_{id} - \beta C_{iod})/(1-\sigma)} D_h^{\sigma}}{e^{(\alpha W_{ic} - \beta C_{ioc})/(1-\sigma)} D_g^{\sigma}}.$$

If destinations c and d belong to the same nest, the above ratio depends only on attributes of c and d, while if they belong to different nests $\mathcal{D}_h$ and $\mathcal{D}_g$, it depends also on attributes of the other destinations in these nests, through the terms D_h and D_g. In other words, the IIA assumption holds within nests, but not across nests. The choice of the nesting pattern is therefore a crucial component of the empirical implementation of the nested logit. In section 4, starting from a baseline nesting structure, we will test the sensitivity of our empirical results to changes in the nesting pattern.

3 Data

Let us now turn to the data used in this analysis. We first present the global migration database DIOC-E, before providing some descriptive evidence on both the size and characteristics of South American emigrant populations circa 2000.

3.1 Migration data

As its name suggests, the *Database on Immigrants in OECD and non-OECD Countries* (DIOC-E) (Dumont et al., 2010) provides information on the number and characteristics of migrants living in various parts of the world. It consists in re-aggregated census microdata and includes demographic, economic and social variables. The database builds on 2000 census data for the 15+ population living in 28 OECD countries and 72 non-OECD countries. For most countries, the place of birth is used to identify migrants, although in some cases it was necessary to rely on criteria based on nationality. When census data were not available, population registers or large representative surveys were used.[8]

The main variables of the database are the country of origin, education, gender, age, labour force status and occupation, although not all cross-tabulations are available. In this paper, we mainly exploit the education dimension. One of the major challenges in compiling these data has been to harmonize the classification of variables which are not systematically collected based on international classifications. This is particularly the case for educational attainment in non-OECD countries of residence. Education mappings for the transformation of national education categories were built in order to be in line with the International Standard Classification of Education (ISCED).

No imputation or estimation is made to increase the coverage of the database in terms of destination countries. With 100 destination countries, the coverage of the world migrant stock in DIOC-E is about 72% (against 47% for the OECD-only database). The coverage is almost 100% for migrants living in South America since all countries

[8]The main data providers are the national statistical institutes of the different destination countries, as well as IPUMS, Afristat, Eurostat and CELADE. See Dumont et al. (2010) for a detailed description of the database and the sources. Note that DIOC-E is an extension of the *Database on Immigrants in OECD Countries* (DIOC) (OECD, 2008) to non-OECD destinations.

are included.[9] For migrants born in a given South American country, we have data on migrants living in 51 countries, including 29 of the 30 main destinations.[10] Using data from United Nations Population Division (2013) as a benchmark, we estimate that the global coverage of South American emigrants in our analysis is close to 98%.

As explained in section 4, we also use data on South American emigrants living in the United States, disaggregated by state of residence. Those data, which are not available in DIOC-E, were constructed directly from the 2000 census microdata of this country (Minnesota Population Center, 2014). The definition of the different variables, in particular education, is matched to that of DIOC-E.

3.2 South American emigrants around the world: descriptive evidence

Beyond the fact that we have an almost comprehensive origin-destination migration matrix, the interest to focus on South America lies in the intensive migration movements within the area, as well as towards other parts of the world (Organization of American States, 2012).

Based on DIOC-E, around the year 2000, there were about 5.4 million migrants in the world, aged 15+ and born in South America. Women represented 53% of the total. The main destination country for these migrants was obviously the United States, with almost 1.6 million, followed by Argentina (960 thousands), Venezuela (675 thousands) and Spain (615 thousands). Then, Japan and Italy hosted about the same number of South American migrants (190 thousands each), followed by Paraguay and Brazil (125 thousands each) and Chile and Canada (about 110 thousands each).

Overall, in 2000, about 40% of international migrants aged 15+ born in South America lived in another country of the region, 30% lived in the United States, 20% lived in Europe and the remaining 10% lived in other countries. Among "regional" migrants, close to 90% lived in a neighbouring country. The two main regional corridors were Colombian migrants in Venezuela (580 thousands) and Paraguayan migrants in

[9] Except Guyana and Suriname, which represent less than 0.4% of the total population of South America.

[10] The only significant destination absent from our data is China, because foreigners are not covered by the Chinese census. According to estimates by United Nations Population Division (2013), there were about 60 thousands South American migrants living in China in 2000, although this figure is difficult to ascertain.

Argentina (305 thousands). Outside South America, the three most important communities were the Ecuadorian and the Peruvian in the United States (resp. 280 and 260 thousands), and the Ecuadorian in Spain (190 thousands).

As shown in Figure 1, the distribution of education of migrants is very different between destinations. About 30% of South American migrants aged 25-64 living in the OECD countries had a tertiary education, while this was only the case for about 10% of those living in another South American country. Most of the latter (65%) had at most a lower secondary education, while this proportion was less than 30% in OECD countries. It is striking to note that the distribution of education of South American migrants living in South America is almost identical to that of South American non-migrants. There are obviously variations in this pattern across countries of origin, but it suggests that the selection process (resulting from self-selection or from migration policies) is widely different in the two groups of destinations.

Some descriptive statistics on the average number of migrants by education level are provided in Table 1. This table highlights the difference between contiguous and non-contiguous country pairs. The average number of low-educated migrants is 16 times higher if the origin and destination countries are contiguous than if they are not. For migrants with an intermediate level of education, this ratio is 3.5, while it is only 1.7 for highly-educated migrants. Again, this indicates that regional migration tends to be less selective than long-distance migration.

4 Empirical framework

4.1 Nested logit estimation using aggregate data

In the theoretical model (section 2), we have implicitly made the assumption that individual data are available to perform a microeconometric analysis based on the nested logit model. As noted above, however, our migration database is composed of *origin* $\times$ *destination* $\times$ *education* population cells.

As shown by Berry (1994), it is possible to recover the structural parameters of a microeconomic discrete choice model using aggregate data. The method developed by Berry has been used extensively in the literature devoted to the econometric estimation of supply and demand curves in differentiated products markets, where the researcher typically observes product characteristics, market characteristics, and market shares, but not individual choices (see e.g. Berry et al., 1995; Nevo, 2001; Petrin, 2002). As far as we know, this approach has been used only twice in the migration literature. In their analysis of internal migration in the United States, Sorensen et al. (2007) examine the extent to which New Deal spending affected domestic migration patterns in the second half of the 1930s. Ortega and Peri (2013) also use a nested logit model with aggregate data on international migration. Although they allow for unobserved individual heterogeneity between migrants and non-migrants, they assume that the IIA hypothesis holds for all migration destinations.

Basically, the approach developed by Berry (1994) involves the inversion of the market-share function to obtain the mean utility levels associated with each product. The first step is the equivalence between individual choice probability (equation (3)) and the population average, i.e. the market share of a given alternative. If we retain education as a potentially discriminant individual characteristic in destination choice, we get the following expression for the "market share" of destination country $d \in \mathcal{D}_g$ for individuals of education level E born in country o:

$$s_{Eod} = \frac{e^{\delta_{Eod}/(1-\sigma)}}{D_g^{\sigma}\left[\sum_g D_g^{(1-\sigma)}\right]}. \tag{5}$$

where the term δ_{Eod}, which is introduced to simplify the notation, represents the mean utility of choice d for individuals with education level E born in country o:

$$\delta_{Eod} = \alpha W_{Ed} - \beta C_{Eod}. \tag{6}$$

The index $E = \{1, 2, 3\}$ denotes the level of education (primary, secondary, tertiary). Here, D_g is equivalent to the expression in equation (4), but at the population level.

The group share, i.e. the probability of choosing one of the destination in group $\mathcal{D}_g$, is the sum of the market shares of all destinations of the group:

$$\bar{s}_{Eog} = \sum_{d \in \mathcal{D}_g} s_{Eod} = \frac{D_g^{(1-\sigma)}}{\left[\sum_g D_g^{(1-\sigma)}\right]}. \tag{7}$$

The market share of destination $d \in \mathcal{D}_g$ can then be written as the product of the group share by the market share of destination d as a fraction of the total group share:

$$s_{Eod} = \bar{s}_{Eod|g} \cdot \bar{s}_{Eog}. \tag{8}$$

In Berry (1994), only one market is considered and the mean utility of the "outside good", δ_{Eoo}, is set to zero, which simplifies somewhat the problem. In our case, the outside good is simply the choice to remain in the origin country ($o = d$). Since we will consider several origin countries in a same estimation, we cannot assume that the mean utility of staying is the same for all origin countries. The market share of staying in the origin country is therefore:

$$s_{Eoo} = \frac{e^{\delta_{Eoo}/(1-\sigma)}}{D_0^{\sigma}\left[\sum_g D_g^{(1-\sigma)}\right]}. \tag{9}$$

Taking the log difference of market shares (5) and (9), we get:

$$\ln(s_{Eod}) - \ln(s_{Eoo}) = \frac{\delta_{Eod}}{1-\sigma} - \sigma \ln D_g - \delta_{Eoo}. \tag{10}$$

This expression depends on D_g, which is unknown. Taking the log of the group share in equation (7) and combining with equation (9), we obtain:

$$\ln D_g = \frac{\ln(\bar{s}_{Eog}) - \ln(s_{Eoo}) + \delta_{Eoo}}{1-\sigma}. \tag{11}$$

We now substitute this expression in (10) and replace the group share $\bar{s}_{Eog}$ according to (8):

$$\ln(s_{Eod}) - \ln(s_{Eoo}) = \delta_{Eod} - \delta_{Eoo} + \sigma \ln(\bar{s}_{Eod|g}) \tag{12}$$

The relevant market shares in the above equation are:

$$s_{Eod} = \frac{n_{Eod}}{(n_{Eoo} + \sum_g \sum_{d \in g} n_{Eod})}$$
$$s_{Eoo} = \frac{n_{Eoo}}{(n_{Eoo} + \sum_g \sum_{d \in g} n_{Eod})}$$
$$\bar{s}_{Eod|g} = \frac{n_{Eod}}{\sum_{d \in g} n_{Eod}}$$

where n_{Eod} is the population of migrants in the corresponding *origin* $\times$ *destination* $\times$ *education* cell, and n_{Eoo} is the number of non-migrants born in country o with education E.

Substituting δ in equation (12) according to (6), and adding an error term u leads to the following empirical equation:

$$\ln(n_{Eod}) - \ln(n_{Eoo}) = \alpha(W_{Ed} - W_{Eo}) - \beta C_{Eod} + \sigma \ln(\bar{s}_{Eod|g}) + u_{Eod}. \qquad (13)$$

In this relationship, as in the theoretical model from which it is derived, one underlying assumption is that the marginal utility of income, α, and the marginal disutility of migration costs, β, are the same for all individuals and country pairs. This does not need to be the case. In particular, since we are looking at the role of education in destination choice, it makes sense to relax the assumption that low-educated and highly-educated individuals react identically to a similar income or distance differential. We therefore allow the coefficients α and β to vary with education. In addition, geographical proximity between the origin country o and the destination country d might also affect the perception of income differences or migration costs. For a country pair (o, d) let us define proximity as $P = 1$ if the two countries are close to each other, and 0 otherwise. In our case, we can use a narrow definition of proximity, such as contiguity among South American countries, or a broader one by assuming for instance that any South American destination is close, and any other destination is not. Thus α and β are also allowed to vary according to proximity between the origin and destination. This more flexible model is written:

$$\ln(n_{Eod}) - \ln(n_{Eoo}) = \alpha_{EP}(W_{Ed} - W_{Eo}) - \beta_{EP} C_{Eod} + \sigma \ln(\bar{s}_{Eod|g}) + u_{Eod}. \qquad (14)$$

There are several econometric and data issues to address in order to estimate equation (14) properly. We first discuss the measurement of wages and migration costs. Second, we turn to the issue of the endogeneity of the within-group share. Third, we deal with the presence of zeros in the data. Fourth, we discuss methods to assess the robustness of the results to the choice of the nesting pattern.

Before delving into these issues, note that, although we have data on migrants and non-migrants aged 15 and older, we restrict the analysis to individuals aged between 25 and 64. Indeed, it is best to exclude as much as possible individuals who have not finished school because their current level of education is not relevant. Moreover, since a significant proportion of younger and older individuals might not participate to the labour market, their location decision might be driven by other considerations than differences in expected wages, which are the main focus of our model.

4.2 Education-specific wages

The education-specific expected wages in the origin and destination countries, W_{Eo} and W_{Ed}, are key explanatory variables of migration in our model. Since comparable data on education-specific wages in a broad sample of countries are quite difficult to come by, we have to estimate those wages. To solve the same problem, Grogger and Hanson (2011) approximate the wages corresponding to primary and tertiary education by the 20th and 80th percentiles of the income distribution of each country. Those percentiles are estimated under the assumption that income has a lognormal distribution, and using Gini coefficients as an indicator of dispersion.

This approach requires using Gini coefficients G measured at the national level. Under the assumption that income X has a log normal distribution $\ln X \sim \mathcal{N}(\mu, \sigma^2)$, the standard deviation of the distribution can be computed as:

$$\sigma = \sqrt{2}\Phi^{-1}\left(\frac{G+1}{2}\right) \tag{15}$$

where Φ is the CDF of the standard normal distribution. Income quantiles are then given by:

$$x_a = E(X) \exp(\sigma z_a - \sigma^2/2)$$

where z_a is the a-quantile of the standard normal distribution and $E(X)$ is an estimate of the economy-wide average wage.

It is however unsatisfactory to assume, à la Grogger and Hanson (2011), that the mapping between the distributions of education and income is the same in all countries. Using the information on the distribution of education available in DIOC-E, we are able to relax this stringent assumption. We assume that the relevant wage for each education category (primary, secondary and tertiary) is the mean wage in this category. We proceed as follows. Gini coefficients for all origin and destination countries are first obtained from the World Income Inequality Database (WIID3.0b) (WIDER, 2014). We estimate the standard deviation of the income distribution, σ, as in equation (15). For this calculation, we use all the Gini coefficients based on income or earnings in the WIID3.0b database for the period 1990-2000. Since several Gini are available for some countries (for different years, or from different sources), we compute the standard deviation σ for each Gini estimate and then use the country average – obtained by taking the square root of the arithmetic mean of σ^2. Then, we compute the wages corresponding to the upper limits of the primary and secondary education categories:

$$x_1^U = E(X) \exp(\sigma \Phi^{-1}(p_1) - \sigma^2/2) \quad \text{and} \quad x_2^U = E(X) \exp(\sigma \Phi^{-1}(p_2) - \sigma^2/2),$$

where p_1 and p_2 are the cumulative shares of the primary and secondary education levels in the population. Here, $E(X)$ is approximated by the country's GDP per capita in constant PPP (averaged over the period 1990-2000, computed from World Bank (2013)). Next, the mean wage for each category is computed as the first moment of the truncated lognormal distribution (Johnson et al., 1994):

$$
\begin{aligned}
W_1 &\equiv \overline{x_1} = E(X)\frac{\Phi(x_1^0 - \sigma)}{\Phi(x_1^0)} \\
W_2 &\equiv \overline{x_2} = E(X)\frac{\Phi(\sigma - x_1^0) - \Phi(\sigma - x_2^0)}{\Phi(x_2^0) - \Phi(x_2^0)} \\
W_3 &\equiv \overline{x_3} = E(X)\frac{\Phi(\sigma - x_2^0)}{\Phi(-x_2^0)}
\end{aligned}
$$

where $x_1^0 = (\ln(x_1^U) - \mu)/\sigma$ and $x_2^0 = (\ln(x_2^U) - \mu)/\sigma$.

As discussed below, we also need to construct education-specific wages for all the US states. We apply the same procedure as the one used at the country level. State-level per capita incomes and Gini coefficients are taken from the Historical Income Tables published by the Census Bureau.[11]

Some descriptive statistics on wage differentials between destination and origin countries are provided in Table 2. The average wage differential between destination and origin for non-contiguous country pairs is quite large but decreases with the level of education: for the low-educated category, wages are between 1.5 and 6 times higher in non-contiguous destinations than in origin countries, while this ratio is at most 3 for the highly-educated. In the case of contiguous destinations, the picture is more mixed: some origin countries, such as Argentina, Brazil, Chile, or Venezuela, have higher wages than their neighbours, while others, such as Bolivia, Colombia, or Peru, have lower wages. The slope of the relationship between the destination/origin wage ratio and education reveals the relative strength of returns to education between each origin country and its neighbours. For example, low-educated individuals in Brazil earn on average 5% more than in the neighbouring countries, but highly-educated ones earn about 82% more (1/0.55=1.82). On the contrary, in Venezuela, the wage premium compared to neighbouring countries is very high for low-educated individuals while it is zero for highly-educated ones.

In the econometric analysis, we will mainly use the log-difference of the education-

[11] Per capita income data and Gini coefficients by state are available at www.census.gov/hhes/www/income/data/historical/state/index.html. For both measures, we take the average of the 1989 and 1999 figures. To make them comparable to the country-level data, US state-level per capita incomes are multiplied by the GDP per capita of the US and divided by the nationwide per capita income.

specific wages between the destination and origin countries, i. e. substituting $(W_{Ed} - W_{Eo})$ with $(\ln W_{Ed} - \ln W_{Eo})$ in equation (14). We will compare our results with those obtained using Grogger and Hanson (2011)'s approach, where the wage of individuals with primary, secondary and tertiary education is approximated respectively by the 20th percentile, the median and the 80th percentile of the income distribution. In addition, we will compare the results of the log-utility model, favoured by Belot and Hatton (2012), with those of the linear-utility one, advocated by Grogger and Hanson (2011), where the destination-origin wage difference is taken in levels and not in logs.

4.3 Migration costs

In our main analysis, migration costs C_{Eod} between the origin and destination countries are approximated by a single variable: the geographical distance between countries o and d (in log), obtained from the CEPII database (Mayer and Zignago, 2011), which proxies for the monetary cost of transportation and the non-monetary cost of absence. Other factors, such as linguistic proximity and migration policies, might also affect migration costs and are indeed often included as regressors in empirical analyses of the determinants of international migration.

In our case, since all origin countries are either Spanish speaking or Portuguese speaking, linguistic proximity between origin and destination countries is only discriminant among non-Latin American destinations. Among those destinations, we can for example distinguish those where a significant share of the population speaks one of the Romance languages (e.g. Spanish, Portuguese, French, Italian) from those where it is not the case. We will present some results using this additional variable, which we construct using data from Mayer and Zignago (2011).

Migration policies have a direct influence on the size and composition of migration flows. Those policies are however very heterogeneous across countries, and there is no single variable (or group of variables) that can record this complexity. Currently available indicators of migration policy are incomplete and of poor quality. Instead, we will include destination country fixed effects to capture some of the differences in migration policies.

Beine et al. (2011) have shown that the presence of a diaspora, proxied by lagged bilateral stock, can significantly foster later migration flows because it can reduce some of the costs of migration. We can of course expect that an existing group of migrants from a given South American country in a given destination should encourage further migration in this corridor. In our analysis, however, since we study migration stocks and not flows, it is not possible to identify such a diaspora effect.

4.4 Endogeneity of the within-group share

As noted by Berry (1994), the term $\ln(\bar{s}_{Eod|g})$ in equation (13) is clearly endogenous, because the market share s_{Eod} and the within-group share have the same numerator n_{Eod}. Estimating the model properly therefore requires (at least) one excluded instrument.

We follow the practice of the industrial organization literature and use the characteristics of other countries in the group as instruments for the within group share. More specifically, for each origin-destination pair, we compute the average (education-specific) wage differential and the average distance between the origin country and the other destinations in the same group. The average wage differential of the other destinations in the group should be correlated negatively with the within group share: for a given destination, if the wages offered by the other countries in the group increase, its market share within the group should decrease. On the contrary, the average distance of the other destinations should be correlated positively with the within group share. The averaged wage differentials and averaged distance of the other countries in the same group of destinations are plausibly uncorrelated with the destination market share itself.

Another potential relevant instrument is simply the number of destinations in the group: it should be negatively correlated with the within-group share while having no direct correlation with the destination market share. Note that, for this instrument, the rank condition can be satisfied only if there is enough heterogeneity in the number of destinations per group. In our empirical analysis, we use those three instruments jointly.

4.5 Dealing with zero cells

The presence in the database of cells with zero population, which are dropped when the dependent variable is logged, might threaten the validity of the estimates. In our case, this problem is relatively limited: for the education-specific bilateral data ($N = 1530$), we have 137 zero cells, about 9% of the data. This issue is common to all macro analyses of the determinants of migration and the literature has proposed various options to deal with it. Zero cells can simply be dropped, but this makes the dataset unbalanced and may lead to inconsistent estimates of the parameters. Another option is to add 1 to each cell, but this can also generate biased estimates.

Another approach to deal with zero cells would be to implement a two-step Heckman regression. The first step involves the estimation of a selection equation for the probability of a positive migration flow for a given country pair, and the second step is the original model modified by the inclusion of the inverse Mills ratio. This approach requires to use an instrument in the first step, that can explain the presence or absence of migration between two countries, but is uncorrelated to the level of migration between these two countries. It is unfortunately not easy to find such an instrument.

In this paper, we instead use Poisson regression models, as suggested first by Silva and Tenreyro (2006) in the context of gravity trade models (more precisely, Poisson pseudo-maximum likelihood - PPML). This approach ensures that our coefficients' estimates will be consistent. This requires to transform slightly the estimated equation (14):

$$n_{Eod} = \exp\left[\alpha_{EP}(W_{Ed} - W_{Eo}) - \beta_{EP}C_{Eod} + \sigma\ln(\bar{s}_{Eod|g}) + \theta\ln(n_{Eoo}) + u_{Eod}\right] \quad (16)$$

where θ is constrained to 1.

Combining the Poisson estimation with the treatment of endogeneity of the within group share can be done in a straightforward fashion with a two-stage residual inclusion (2SRI) approach (Cameron and Trivedi, 2013). We first run an OLS regression of $\ln(\bar{s}_{Eod|g})$ on the exogenous regressors and the instruments. We then compute the predicted residual $\hat{v}_{Eod}$. The second stage is a Poisson regression of n_{Eod} on the exogenous regressors, $\ln(\bar{s}_{Eod|g})$ and $\hat{v}_{Eod}$. Note that this approach requires to compute bootstrapped standard errors because $\hat{v}_{Eod}$ is a generated regressor in the second stage.

We will compare the results of the 2SRI Poisson model to that of the 2SLS approaches (dropping zero cells or adding 1 to each cell).

4.6 Choice of the nesting pattern

There are 10 South American origin countries in our analysis (Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Peru, Paraguay, Uruguay and Venezuela). For a migrant born in one of those countries, there are 51 possible destination countries: the 9 other South American countries, 10 Central American and Caribbean countries[12], 22 European OECD countries[13], 5 non-European OECD countries[14], and 5 other countries[15].

As noted above, the results from a nested logit estimation are inherently conditional on the nesting pattern that is chosen. Since there is no unique way to classify the destination countries in groups, we use three distinct approaches to test the sensitivity of the results to the choice of the nesting pattern.

First, we estimate our model for different nesting patterns, with an increasing disaggregation of destination countries:

- A: South America / all other destinations,
- B: South America / Central America and Caribbean / all other destinations,
- C: South America / Central America and Caribbean / North America and European OECD / all other destinations,

Second, we study the influence of the geographical unit of analysis, by splitting the United States into 51 distinct destinations, with the following nesting pattern:

- D: South America / Central America and Caribbean / North America (Canada + 51 US states) / European OECD / all other destinations.

Third, for each nesting pattern described above, we explore the robustness of our results to marginal changes in the composition of the groups, in the spirit of a permutation test. For each of the nesting patterns A, B and C, we re-estimate the model for all

[12]Belize, Costa Rica, Dominican Republic, Guatemala, Honduras, Mexico, Nicaragua, Panama, Puerto Rico and Salvador.

[13]Austria, Belgium, Czech Republic, Denmark, France, Finland, Germany, Greece, Hungary, Ireland, Italy, Luxembourg, Norway, Netherlands, Poland, Portugal, Spain, Slovenia, Slovakia, Switzerland, Sweden, and the United Kingdom.

[14]Australia, Canada, Japan, New Zealand, and the United States

[15]Israel, Philippines, Russia, South Africa, and Ukraine.

possible one-to-one swaps of countries between groups. For example, for the nesting pattern A, since there are 10 destinations in the South America group and 42 destinations in the "other destinations" group, this generates 420 alternative nesting pattern which are marginally different from A.[16] We proceed similarly for nesting patterns B and C, for which we obtain respectively 740 and 932 alternative nesting patterns. If the coefficients obtained from these estimations differ widely from those of the original nesting patterns, this will imply that these results are very sensitive to local misspecifications of the groups.

5 Results

5.1 Conditional vs nested logit

Our main results are presented in Tables 3 to 9. In Table 3, we first contrast results from a conditional logit applied to OECD destination countries only (column 1) to a similar model where all destination countries are considered (column 2). This model results from the Poisson pseudo-maximum likelihood (PPML) estimation of equation (16) with the constraint $\sigma = 0$. For both models, as expected, the estimated marginal utility of income is positive while distance has a negative effect on migration. In the first model, there is no significant difference in the wage differentials and distance coefficients across education levels, while there are significant differences in the second model, especially between the primary and tertiary levels. The magnitude of the coefficients is also different, which is a first indication that the determinants of South-South migration are different from that of South-North migration.

As discussed above, we believe that these conditional logit models are misspecified. In column 3, we therefore relax the IIA assumption and estimate a nested logit with two groups: South American destinations vs all other destinations (nesting pattern A). The coefficient of the within group share σ is significantly different from 0, indicating that the IIA hypothesis is indeed rejected, which constitutes our first key result. Compared to the conditional logit in column 2, the gradients of the wage differentials and distance

[16] If countries are ranked alphabetically in both groups, the first swap to consider would be to put Argentina in the "other destinations" group and Australia in the South American group.

coefficients are much steeper. This is our second important finding: wage differences between origin and destination countries have a much stronger impact on low-skilled migration than on highly-skilled migration. Low-skilled migration is also more strongly affected by distance. For low-skilled individuals, lower expected earnings imply that migration is only feasible for relatively low migration costs, that is for closer destinations.

In the estimation of the nested logit, the endogeneity of the within-group share is dealt with three excluded instruments: the average wage differentials and average distance of the other countries in the same group of destinations, and the number of countries in the group. The relevance of these instruments can be assessed by looking at the first-stage results, reported in Table 5.[17] For the nested logit model estimated in Table 3, the first-stage results are in column 1 of Table 5: all three instruments are significantly correlated with the within-group share, and the F-statistic of a joint significance test of all excluded instruments is well above 10. The signs of the instruments' coefficients are consistent with the arguments given in section 4.4: (i) for each destination country, higher wages in other countries of the same group of destinations will reduce its within-group share; (ii) among destinations in a given group, the countries that are closer to the origin country will have a higher within-group share; and (iii) the within-group share is lower in groups with more destination countries. In addition, we also report the coefficient of the first-stage residuals in the second-stage equation at the bottom of Table 3 (column 3), which is a direct Hausman test of the exogeneity of the within-group share. The coefficient of the first-stage residuals being significantly different from zero, the null hypothesis that the within-group share is exogenous can be rejected.

5.2 Role of proximity and different nesting patterns

Now that we have confirmed the irrelevance of the IIA hypothesis for our analysis, we estimate the model presented in equation (16) for different nesting patterns. The results are presented in Table 4. For all the nesting patterns, the IIA hypothesis is unambiguously rejected.

[17]This table only reports the coefficients of the excluded instruments. The full first-stage results are available upon request.

Contrary to the simple nested logit model of Table 3 (column 3), the models presented in Table 4 include interactions between the education-specific coefficients and the contiguity dummy. This additional flexibility reveals that the marginal utility of income is different, not only for different levels of education – with the negative gradient already discussed – but also depending on the proximity between the origin and destination countries. This is our third key result. We observe that (i) for both contiguous and non-contiguous country pairs, low educated individuals are more sensitive to wage differences than those who are better educated; (ii) the effect of wage differences is always higher for contiguous countries than for non-contiguous countries; and (iii) the gradient is steeper for contiguous country pairs. This result holds for the three different nesting patterns, although the point estimates are not exactly the same. For the distance coefficient, we also find differences across education and according to proximity between origin and destination. The effect of distance on migration is much higher for low educated individuals and for neighbouring countries. Overall, the values for individual coefficients are quite similar for the three nesting patterns, which indicates that results are not overly sensitive to this aspect of modelling choice.[18]

We can further investigate this issue with nesting pattern D (column 4 in Table 4). In this configuration, the US is now considered as a collection of 51 distinct destinations, each with a different set of education-specific wages and with a different distance to each South American origin country. Remarkably, the coefficient of the within group share is very close to those obtained with nesting patterns A, B and C. The coefficients for the wage differentials and distance are also very close, with similar gradients by education level and a comparable effect of contiguity.

For all the models in Table 4, the first-stage results are given in Table 5: the three instruments are jointly significant, with F-statistics comfortably above 10. The Hausman tests of exogeneity of the within-group share are reported for the different nesting patterns in the last row of Table 4 and indicate that the within-group share is indeed endogenous.

[18]The same models with a South America dummy instead of the contiguity dummy give qualitatively similar results, which are reported in the appendix (Table 10 for the nested logit estimates and Table 11 for the first-stage results).

5.3 Treatment of zero cells

In the results presented thus far, we account for the presence of zero cells in the dataset through the use of a Poisson regression. In Table 6, we focus on the simple nesting pattern A (South American destinations vs other destinations) to compare this approach with other methods to deal with zero cells. Column 1 reproduces the results of the Poisson regression with nesting pattern A (column 1 in Table 4), while columns 2 and 3 present two variants of the 2SLS estimation: cells with zero migrants are simply dropped (column 2), or each cell gets one additional migrant (column 3). These two different 2SLS models give very similar results, with only marginal differences in point estimates. They are however quite different from the Poisson regression. In particular, in the 2SLS regressions, there is no education gradient in the effect of wage differential on migration for non-contiguous country pairs, while a significant gradient is still found for contiguous country pairs – although it is much less pronounced than in the Poisson model. This result likely reflects the misspecification of the 2SLS approach.[19]

5.4 Education-specific wages

Another issue that we have highlighted above is the estimation of education-specific wages. With the method proposed by Grogger and Hanson (2011) as a starting point, we have matched the income distribution to the education distribution. In Table 7, we compare the results given by the two methods (columns 1 and 2) for the nesting pattern A. Although our main results are obtained assuming a log-utility model, we also estimate a linear-utility model for the same nesting pattern (column 3). Compared to our baseline results, the use of the Grogger-Hanson version of the education-specific wages mainly affects the coefficients of the wage differential for non-contiguous country pairs. In particular, the negative gradient discussed above disappears. For contiguous country pairs, the wage differential coefficients are very similar to those of the baseline model. The distance coefficients are not much affected by this change in the wage variable. When wages are assumed to enter the utility function linearly, the results are similar to those obtained with the log-utility model, although logically with different

[19]The first-stage results for the 2SLS regressions are given in columns 1 and 2 of Table 12.

point estimates.[20]

5.5 Migration costs

Regarding migration costs, we expand the baseline model by adding a variable reflecting linguistic proximity between the origin and destination countries, or destination countries fixed effects to take into account differences in immigration policies. The results are shown in Table 8. Column 1 reproduces our baseline model for nesting pattern A. Column 2 introduces in this model the "Romance language" variable, interacted with education levels. This variable captures linguistic proximity between origin countries, which are Spanish or Portuguese speaking, and destination countries. Language proximity between origin and destination has a positive effect on migration. Although this effect seems larger for lower-educated migrants, the confidence intervals for the three coefficients overlap and are thus not statistically significant from each other. The introduction of linguistic proximity in the model has only a minor impact on the coefficients of other variables, except distance for non contiguous country pairs which becomes insignificant. On the contrary, destination fixed effects (column 3) do not have any major impact on the distance variables, but lower the wage differential coefficients, especially for non contiguous pairs. A likelihood-ratio test indicates that the destination country dummies are jointly significant. In addition to proxying for some elements of migration policies, which may be education-specific, they also absorb part of the overall income differential between South American countries and OECD countries.[21]

5.6 Robustness to marginal changes in the nesting pattern

As noted above, the coefficients of a nested logit estimation are inherently sensitive to the definition of the nesting pattern. This is indeed what we observe for the different nesting patterns in Table 4: although the results are qualitatively similar, which is somewhat reassuring, the point estimates are different for each nesting pattern. The approach followed in Table 4 is to increase progressively the number of nests (nesting

[20] The first-stage results for the models in columns 2 and 3 of Table 7 are given in columns 3 and 4 of Table 12.

[21] The first-stage results for the models in columns 2 and 3 of Table 8 are given in columns 5 and 6 of Table 12.

patterns A, B and C), or to disaggregate one country into several distinct destinations (nesting pattern D). Another way of assessing the impact of the choice of the nesting pattern on estimated coefficients is to introduce marginal changes in the allocation of countries to nests. The results of these robustness checks for nesting patterns A, B and C are presented in Table 9. For each nesting pattern, we report the mean and the 5th and 95th percentiles of the distribution of the wage and distance coefficients[22] for the replications of our model (equation 16). We also look at the distribution of the within-share coefficient (σ). Each replication corresponds to a different marginal modification of the nesting pattern, where two countries from two different nests are swapped. There are 420 possible replications for the nesting pattern A, 740 for B and 932 for C. Comparing the coefficients obtained through these replications and the actual values in Table 4 reveals that the main results discussed above are quite robust to marginal misspecifications of the nesting pattern. For the nesting pattern A (resp. B and C), the coefficient of the within group share σ is in the interval 0.5–0.8 (resp. 0.34–0.92 and 0.42–0.79) for 90% of replications. The 5th and 95th percentiles for the wage and distance coefficients are also consistent with the results found previously, in particular the negative relationship between the wage differential coefficient and education, and the difference in coefficients between contiguous and non-contiguous country pairs.

6 Conclusion

In this paper, we explore the determinants of regional vs long-distance international migration, emphasizing the relative selectivity associated with these two types of mobility. More specifically, we analyse emigration from 10 South American countries towards (almost) the whole world. This represents a step forward compared to most macro-analyses of the determinants of international migration, which usually focus on migrants living in OECD countries. We also depart from the existing literature by relaxing the assumption of independence of irrelevant alternatives (IIA) in the destination choice model, through a nested logit approach. Finally, to estimate education-specific wages, which are a key determinants of migration, we match the income distribution to the education

[22] Interactions between education level, contiguity dummy and, respectively, the wage differential and the log of distance.

distribution.

From a methodological viewpoint, we show that the IIA hypothesis is not compatible with migration choices made by South Americans. Wrongly assuming that it holds upwardly biases estimates of the marginal utility of income, especially for higher educated individuals. Due to the dependence of nested logit estimates with respect to the nesting pattern of destinations, we assess the robustness of our results using two different strategies. Starting from a simple dichotomy between South American countries and all other destinations, we first incrementally disaggregate the latter group and estimate our model for the different nesting patterns obtained. We also re-estimate each model after applying systematic marginal changes to the nesting pattern. In both cases, our results are qualitatively unchanged.

The determinants of regional migration in South America are different from that of long-distance migration: regional migration is less selective and much more sensitive to international wage differentials. More precisely, we find that (i) for both contiguous and non-contiguous country pairs, low educated individuals are more sensitive to wage differences than those who are better educated; (ii) the effect of wage differences is always higher for contiguous countries than for non-contiguous countries; and (iii) the gradient is steeper for contiguous country pairs.

Results obtained for different variants of the model show that our estimates are not overly sensitive to the precise measurement of education-specific wages or to the manner in which those wages enter the utility function. They are however quite different according to the approach taken to account for bilateral corridors with zero migrant: compared to the Poisson regression, 2SLS models are clearly misspecified.

The upcoming availability of data from the 2010 census round will hopefully allow us to check whether the results obtained here remain valid for more recent migration flows. In addition, it would also be interesting to analyse the determinants of regional and long-distance emigration for other origin countries or regions, so as to compare them with the South American case.

References

Bakewell, O., 2009, South-south migration and human development: Reflections on African experiences, Human Development Research Paper 2009/07, UNDP.

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

Belot, M. V. K. and T. J. Hatton, 2012, Immigrant selection in the OECD, *The Scandinavian Journal of Economics*, 114, 1105–1128.

Berry, S., 1994, Estimating discrete-choice models of product differentiation, *The RAND Journal of Economics*, 25, 242–262.

Berry, S., J. Levinsohn and A. Pakes, 1995, Automobile prices in market equilibrium, *Econometrica*, 63, 841–890.

Bertoli, S. and J. Fernandez-Huertas Moraga, 2013, Multilateral resistance to migration, *Journal of Development Economics*, 102, 79–100.

Borjas, G. J., 1987, Self-selection and the earnings of immigrants, *American Economic Review*, 77, 531–53.

Cameron, A. C. and P. K. Trivedi, 2013, *Regression Analysis of Count Data*, Cambridge University Press.

Cardell, N. S., 1997, Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity, *Econometric Theory*, 13, 185–213.

Dahl, G. B., 2002, Mobility and the return to education: Testing a Roy model with multiple markets, *Econometrica*, 70, 2367–2420.

Docquier, F. and A. Marfouk, 2006, International migration by educational attainment (1990-2000) - release 1.1, in: C. Ozden and M. Schiff (eds.), *International Migration, Remittances and Development*, Palgrave Macmillan.

Dumont, J.-C., G. Spielvogel and S. Widmaier, 2010, International migrants in developed, emerging and developing countries: An extended profile, OECD Social, Employment and Migration Working Papers 114, OECD.

Durand, J. and D. S. Massey, 2010, New World orders: Continuities and changes in Latin American migration, *Annals of the American Academy of Political and Social Science*, 630, pp. 20–52.

Falaris, E. M., 1987, A nested logit migration model with selectivity, *International Economic Review*, 28, 429–443.

Gould, W. T. S., 1974, International migration in Tropical Africa: A bibliographical review, *International Migration Review*, 8, pp. 347–365.

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

Hanson, G. H. and C. McIntosh, 2012, Birth rates and border crossings: Latin American migration to the US, Canada, Spain and the UK, *The Economic Journal*, 122, 707–726.

Johnson, N. L., S. Kotz and N. Balakrishnan, 1994, *Continuous Univariate Distributions, Vol. 1*, John Wiley and Sons.

Kritz, M. M. and D. T. Gurak, 1979, International migration trends in Latin America: Research and data survey, *International Migration Review*, 13, pp. 407–427.

Mayer, T. and S. Zignago, 2011, Notes on CEPII's distances measures: The GeoDist database, Working Papers 2011-25, CEPII.

McFadden, D. L., 1977, Modelling the choice of residential location, Cowles Foundation Discussion Papers 477, Cowles Foundation for Research in Economics, Yale University.

Minnesota Population Center, 2014, *Integrated Public Use Microdata Series, International: Version 6.3*, University of Minnesota.

Nevo, A., 2001, Measuring market power in the ready-to-eat cereal industry, *Econometrica*, 69, 307–342.

OECD, 2008, *A Profile of Immigrant Populations in the 21st Century: Data from OECD Countries*, OECD Publishing.

Organization of American States, 2012, *International Migration in the Americas: Second Report of the Continuous Reporting System on International Migration in the Americas (SICREMI) 2012*, Organization of American States.

Ortega, F. and G. Peri, 2013, The effect of income and immigration policies on international migration, *Migration Studies*, 1, 47–74.

Ozden, C., C. R. Parsons, M. Schiff and T. L. Walmsley, 2011, Where on Earth is everybody? The evolution of global bilateral migration 1960-2000, *The World Bank Economic Review*, 25, 12–56.

Pellegrini, P. A. and A. S. Fotheringham, 2002, Modelling spatial choice: A review and synthesis in a migration context, *Progress in Human Geography*, 26, 487–510.

Petrin, A., 2002, Quantifying the benefits of new products: The case of the minivan, *Journal of Political Economy*, 110, 705–729.

Ratha, D. and W. Shaw, 2007, South-South migration and remittances, World Bank Working Paper 102, World Bank.

Roy, A. D., 1951, Some thoughts on the distribution of earnings, *Oxford Economic Papers*, 3, 135–146.

Seccombe, I. J., 1985, International labor migration in the Middle East: A review of literature and research, 1974-84, *International Migration Review*, 19, pp. 335–352.

Silva, J. S. and S. Tenreyro, 2006, The log of gravity, *The Review of Economics and Statistics*, 88, 641–658.

Sorensen, T., P. Fishback, S. K. Allen and S. Kantor, 2007, Migration creation, diversion, and retention: New Deal Grants and migration: 1935-1940, IZA Discussion Papers 3060, Institute for the Study of Labor (IZA).

Train, K., 2009, *Discrete Choice Methods with Simulation*, Cambridge University Press.

United Nations Population Division, 2013, *Trends in International Migrant Stock: Migrants by Destination and Origin*, United Nations.

WIDER, 2014, *UNU-WIDER World Income Inequality Database, Version 3.0b*, UNU-WIDER.

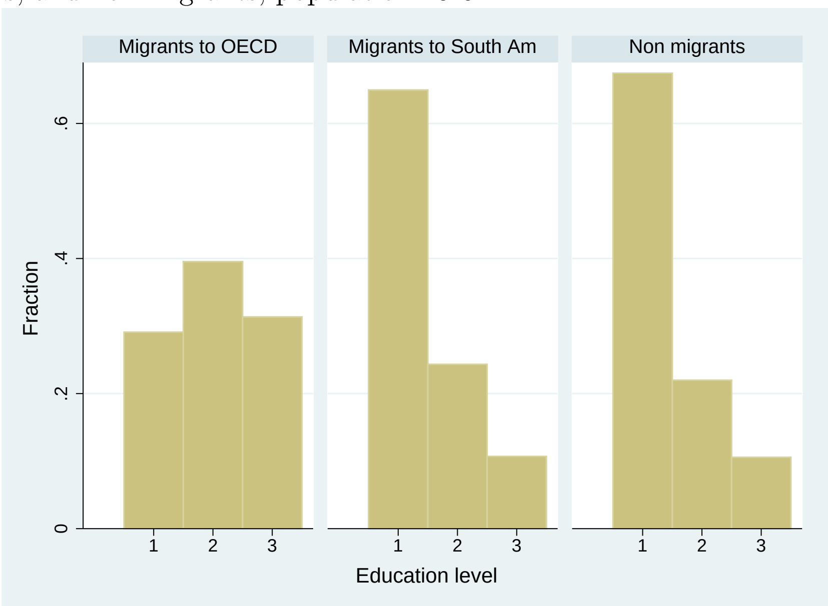
World Bank, 2013, *World Development Indicators*, World Bank.

Table 1: Average number of migrants by education level for contiguous and non-contiguous country pairs

	Non-contiguous country pairs Mean (sd)	Contiguous country pairs Mean (sd)
Migrants E1	1635	27087
	(8038)	(70268)
Migrants E2	2287	8200
	(11585)	(15825)
Migrants E3	1831	3177
	(7741)	(4008)

Source: *Database on Immigrants in OECD and non-OECD Countries (DIOC-E)*, OECD.

Figure 1: Distribution of education for South American migrants, according to their destinations, and non-migrants, population 25-64



Source: *Database on Immigrants in OECD and non-OECD Countries (DIOC-E)*, OECD.

Table 2: Destination/origin wage ratio by education level for each origin country and contiguous vs non-contiguous destinations

	Non-contiguous			Contiguous		
	E1	E2	E3	E1	E2	E3
Origin countries						
Argentina	1.82	1.38	1.20	0.67	0.77	0.87
Bolivia	5.93	3.73	2.83	2.04	2.03	2.14
Brazil	2.54	1.34	0.95	0.96	0.67	0.55
Chile	2.64	1.56	0.98	0.82	0.65	0.50
Colombia	3.44	2.26	1.63	1.31	1.28	1.16
Ecuador	3.28	2.13	1.71	0.82	0.74	0.81
Paraguay	4.55	2.63	1.87	1.74	1.56	1.43
Peru	5.10	4.10	3.06	1.54	2.09	2.22
Uruguay	1.41	1.31	1.55	0.70	1.01	1.52
Venezuela	1.57	1.25	1.16	0.57	0.78	1.00

Sources: *Database on Immigrants in OECD and non-OECD Countries (DIOC-E)*, OECD; WIDER (2014).

Note: The wage ratio for low educated individuals between Argentina and all non-contiguous destinations is on average 1.87, i.e. the wage is 87% higher in non-contiguous destinations than in Argentina; it is 0.67 for contiguous destinations, i.e. the wage is 33% lower in contiguous destinations than in Argentina.

Table 3: Conditional logit OECD vs Conditional logit all countries vs Nested logit

	(1) Cond. logit OECD	(2) Cond. logit All	(3) Nested logit A
$\ln(\bar{s}_{Eod\|g})$ (σ)			0.820^{a}
			(0.137)
E2	2.764	-4.325	-7.311^{b}
	(7.669)	(3.093)	(3.615)
E3	-1.090	-8.518^{a}	-11.276^{a}
	(6.467)	(2.850)	(2.998)
Δ ln wage $\times$ E1	0.980^{b}	1.856^{a}	1.767^{a}
	(0.403)	(0.324)	(0.479)
Δ ln wage $\times$ E2	0.925^{b}	1.571^{a}	0.899^{a}
	(0.360)	(0.269)	(0.249)
Δ ln wage $\times$ E3	0.654^{a}	1.116^{a}	0.381^{b}
	(0.236)	(0.234)	(0.172)
ln distance $\times$ E1	-3.741^{a}	-1.868^{a}	-1.708^{a}
	(0.660)	(0.327)	(0.386)
ln distance $\times$ E2	-3.838^{a}	-1.155^{a}	-0.562^{a}
	(0.588)	(0.218)	(0.168)
ln distance $\times$ E3	-3.313^{a}	-0.565^{a}	0.018
	(0.334)	(0.154)	(0.208)
First-stage residuals			0.314
(Hausman test)			(0.095)
Observations	810	1530	1530

Columns 1 and 2: Robust standard erros in parentheses.
Column 3: Bootstrapped standard errors in parentheses (1000 replications).
^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table 4: Heterogeneity of wage and distance coefficients by education, proximity measured by contiguity dummy, for different nesting patterns

	(1) Nested logit A	(2) Nested logit B	(3) Nested logit C	(4) Nested logit D
$\ln(\bar{s}_{Eod\|g})$ (σ)	0.511^{a}	0.500^{a}	0.464^{a}	0.610^{a}
	(0.052)	(0.083)	(0.068)	(0.042)
E2	-0.427	-0.527	-0.501	0.643
	(3.291)	(3.005)	(1.642)	(2.765)
E3	-1.517	-3.127	-1.216	0.618
	(2.096)	(2.476)	(1.878)	(2.118)
Contiguity	10.034^{b}	8.847^{b}	10.326^{a}	10.604^{a}
	(4.007)	(3.933)	(3.198)	(3.576)
Contiguity × E2	0.155	-0.255	-0.876	-1.145
	(4.461)	(4.228)	(4.819)	(3.874)
Contiguity × E3	-4.344	-3.526	-6.400^{c}	-6.966^{c}
	(4.495)	(4.185)	(3.641)	(3.653)
Δ ln wage × No Contig × E1	0.876^{a}	1.545^{a}	1.280^{a}	1.322^{a}
	(0.287)	(0.309)	(0.256)	(0.123)
Δ ln wage × No Contig × E2	0.609^{a}	1.091^{a}	0.946^{a}	0.754^{a}
	(0.176)	(0.232)	(0.217)	(0.148)
Δ ln wage × No Contig × E3	0.296^{c}	0.720^{a}	0.626^{a}	0.292^{a}
	(0.156)	(0.243)	(0.206)	(0.092)
Δ ln wage × Contig × E1	2.624^{a}	2.636^{a}	2.708^{a}	2.704^{a}
	(0.476)	(0.420)	(0.427)	(0.358)
Δ ln wage × Contig × E2	1.912^{a}	1.888^{a}	1.853^{a}	1.806^{a}
	(0.230)	(0.224)	(0.235)	(0.299)
Δ ln wage × Contig × E3	0.992^{a}	0.981^{a}	0.964^{a}	0.954^{a}
	(0.155)	(0.143)	(0.145)	(0.175)
ln distance × No Contig × E1	-0.815^{a}	-1.197^{a}	-1.179^{a}	-0.864^{a}
	(0.252)	(0.296)	(0.152)	(0.237)
ln distance × No Contig × E2	-0.522^{a}	-0.832^{a}	-0.828^{a}	-0.646^{a}
	(0.202)	(0.203)	(0.217)	(0.155)
ln distance × No Contig × E3	-0.278^{c}	-0.405^{c}	-0.628^{a}	-0.522^{a}
	(0.157)	(0.236)	(0.175)	(0.102)
ln distance × Contig × E1	-1.966^{a}	-2.122^{a}	-2.318^{a}	-1.995^{a}
	(0.476)	(0.457)	(0.458)	(0.356)
ln distance × Contig × E2	-1.835^{a}	-1.914^{a}	-2.016^{a}	-1.818^{a}
	(0.298)	(0.308)	(0.283)	(0.336)
ln distance × Contig × E3	-1.007^{a}	-1.036^{a}	-1.081^{a}	-0.957^{a}
	(0.169)	(0.158)	(0.147)	(0.136)
First-stage residuals	0.605	0.697	0.786	0.557
(Hausman test)	(0.047)	(0.049)	(0.048)	(0.050)
Observations	1530	1530	1530	3030

Bootstrapped standard errors in parentheses (1000 replications).
^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table 5: First-stage results for models in Tables 3 and 4

	Table 3	Table 4			
	(1) Nest. logit A	(2) Nest. logit A	(3) Nest. logit B	(4) Nest. logit C	(5) Nest. logit D
Average wage difference	-1.635^a	-1.638^a	-1.609^a	-1.850^a	-2.410^a
	(0.174)	(0.175)	(0.184)	(0.182)	(0.155)
Average distance	3.512^a	3.277^a	3.082^a	2.561^a	0.721^a
	(0.329)	(0.337)	(0.264)	(0.270)	(0.178)
Nb. countries by group	-0.249^a	-0.221^a	-0.248^a	-0.237^a	-0.053^a
	(0.020)	(0.021)	(0.016)	(0.019)	(0.004)
Adj. R-squared	0.343	0.351	0.395	0.346	0.283
F-test	186.2	124.5	149.9	127.4	242.4
F-test df	(3,1517)	(3,1508)	(3,1508)	(3,1508)	(3,3008)
Observations	1530	1530	1530	1530	3030

Robust standard errors in parentheses.
^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table 6: Dealing with zero cells: Poisson vs 2SLS

	(1) IV Poisson	(2) 2SLS $n > 0$	(3) 2SLS $n + 1$
$\ln(\bar{s}_{Eod\|g})$ (σ)	0.511^{a}	0.743^{a}	0.736^{a}
	(0.052)	(0.033)	(0.027)
E2	-0.427	-0.914	-0.808
	(3.291)	(1.035)	(0.938)
E3	-1.517	-2.036^{b}	-1.717^{b}
	(2.096)	(0.961)	(0.852)
Contiguity	10.034^{b}	3.380	3.344
	(4.007)	(3.157)	(3.124)
Contiguity × E2	0.155	1.875	1.741
	(4.461)	(4.226)	(4.215)
Contiguity × E3	-4.344	1.370	1.016
	(4.495)	(3.518)	(3.494)
Δ ln wage × No Contig × E1	0.876^{a}	0.396^{a}	0.309^{a}
	(0.287)	(0.070)	(0.060)
Δ ln wage × No Contig × E2	0.609^{a}	0.394^{a}	0.369^{a}
	(0.176)	(0.074)	(0.068)
Δ ln wage × No Contig × E3	0.296^{c}	0.340^{a}	0.309^{a}
	(0.156)	(0.065)	(0.057)
Δ ln wage × Contig × E1	2.624^{a}	1.311^{a}	1.314^{a}
	(0.476)	(0.338)	(0.338)
Δ ln wage × Contig × E2	1.912^{a}	1.298^{a}	1.299^{a}
	(0.230)	(0.235)	(0.235)
Δ ln wage × Contig × E3	0.992^{a}	0.798^{a}	0.799^{a}
	(0.155)	(0.250)	(0.250)
ln distance × No Contig × E1	-0.815^{a}	-0.491^{a}	-0.503^{a}
	(0.252)	(0.090)	(0.080)
ln distance × No Contig × E2	-0.522^{a}	-0.221^{a}	-0.244^{a}
	(0.202)	(0.075)	(0.071)
ln distance × No Contig × E3	-0.278^{c}	-0.028	-0.078
	(0.157)	(0.058)	(0.056)
ln distance × Contig × E1	-1.966^{a}	-0.880^{b}	-0.888^{b}
	(0.476)	(0.407)	(0.408)
ln distance × Contig × E2	-1.835^{a}	-0.955^{a}	-0.959^{a}
	(0.298)	(0.369)	(0.370)
ln distance × Contig × E3	-1.007^{a}	-0.726^{a}	-0.729^{a}
	(0.169)	(0.204)	(0.204)
First-stage residuals	0.605		
(Hausman test)	(0.047)		
Observations	1530	1393	1530

Column 1: Bootstrapped standard errors in parentheses (1000 replications).
Columns 2 and 3: Robust standard errors in parentheses.
^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table 7: Education-specific wages: Baseline vs. 20/50/80 wages vs. linear wages

	(1) W: E1/E2/E3	(2) W: 20/50/80	(3) W: linear
$\ln(\bar{s}_{Eod\|g})$ (σ)	0.511^{a}	0.469^{a}	0.408^{a}
	(0.052)	(0.053)	(0.067)
E2	-0.427	0.792	-0.203
	(3.291)	(3.234)	(3.311)
E3	-1.517	-0.738	-1.257
	(2.096)	(1.642)	(2.174)
Contiguity	10.034^{b}	10.294^{b}	9.134^{b}
	(4.007)	(4.098)	(4.396)
Contiguity × E2	0.155	-2.366	0.208
	(4.461)	(3.944)	(3.931)
Contiguity × E3	-4.344	-5.926	-3.774
	(4.495)	(4.686)	(4.924)
Δ [ln] wage × No Contig × E1	0.876^{a}	0.429^{a}	0.181^{a}
	(0.287)	(0.134)	(0.061)
Δ [ln] wage × No Contig × E2	0.609^{a}	0.481^{b}	0.063^{a}
	(0.176)	(0.189)	(0.015)
Δ [ln] wage × No Contig × E3	0.296^{c}	0.438^{a}	0.018^{a}
	(0.156)	(0.153)	(0.005)
Δ [ln] wage × Contig × E1	2.624^{a}	1.972^{a}	0.899^{a}
	(0.476)	(0.418)	(0.235)
Δ [ln] wage × Contig × E2	1.912^{a}	1.469^{a}	0.235^{a}
	(0.230)	(0.213)	(0.026)
Δ [ln] wage × Contig × E3	0.992^{a}	1.079^{a}	0.045^{a}
	(0.155)	(0.092)	(0.007)
ln distance × No Contig × E1	-0.815^{a}	-0.875^{a}	-0.933^{a}
	(0.252)	(0.243)	(0.278)
ln distance × No Contig × E2	-0.522^{a}	-0.764^{a}	-0.660^{a}
	(0.202)	(0.236)	(0.218)
ln distance × No Contig × E3	-0.278^{c}	-0.483^{a}	-0.411^{a}
	(0.157)	(0.182)	(0.148)
ln distance × Contig × E1	-1.966^{a}	-2.103^{a}	-1.903^{a}
	(0.476)	(0.489)	(0.560)
ln distance × Contig × E2	-1.835^{a}	-1.771^{a}	-1.804^{a}
	(0.298)	(0.316)	(0.308)
ln distance × Contig × E3	-1.007^{a}	-1.010^{a}	-1.047^{a}
	(0.169)	(0.188)	(0.161)
First-stage residuals	0.605	0.676	0.697
(Hausman test)	(0.047)	(0.052)	(0.070)
Observations	1530	1530	1530

Bootstrapped standard errors in parentheses (1000 replications).
The wage difference variable is Δ ln wage in columns 1 and 2, and Δ wage in column 3.
^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table 8: Migration costs: Baseline vs. linguistic proximity vs. destination fixed effects

	(1) Distance	(2) Language	(3) Dest. FE
$\ln(\bar{s}_{Eod\|g})$ (σ)	0.511^a	0.414^a	0.453^a
	(0.052)	(0.053)	(0.096)
E2	-0.427	1.531	-0.174
	(3.291)	(3.374)	(3.551)
E3	-1.517	0.235	-0.894
	(2.096)	(2.127)	(2.902)
Contiguity	10.034^b	16.499^a	15.336^a
	(4.007)	(4.258)	(5.687)
Contiguity × E2	0.155	-2.110	-1.648
	(4.461)	(4.546)	(3.035)
Contiguity × E3	-4.344	-6.542	-5.644
	(4.495)	(4.737)	(3.597)
Δ ln wage × No Contig × E1	0.876^a	0.849^a	0.556^a
	(0.287)	(0.179)	(0.179)
Δ ln wage × No Contig × E2	0.609^a	0.648^a	0.424^c
	(0.176)	(0.181)	(0.229)
Δ ln wage × No Contig × E3	0.296^c	0.307	0.003
	(0.156)	(0.198)	(0.133)
Δ ln wage × Contig × E1	2.624^a	2.686^a	1.943^a
	(0.476)	(0.478)	(0.476)
Δ ln wage × Contig × E2	1.912^a	1.905^a	1.423^a
	(0.230)	(0.235)	(0.394)
Δ ln wage × Contig × E3	0.992^a	0.999^a	0.453
	(0.155)	(0.151)	(0.326)
ln distance × No Contig × E1	-0.815^a	-0.219	-0.700^c
	(0.252)	(0.213)	(0.409)
ln distance × No Contig × E2	-0.522^a	-0.121	-0.455
	(0.202)	(0.247)	(0.316)
ln distance × No Contig × E3	-0.278^c	0.157	-0.256
	(0.157)	(0.112)	(0.212)
ln distance × Contig × E1	-1.966^a	-2.184^a	-2.556^a
	(0.476)	(0.494)	(0.449)
ln distance × Contig × E2	-1.835^a	-1.949^a	-2.222^a
	(0.298)	(0.291)	(0.373)
ln distance × Contig × E3	-1.007^a	-1.071^a	-1.515^a
	(0.169)	(0.165)	(0.265)
Romance × E1		1.579^a	
		(0.352)	
Romance × E2		1.190^a	
		(0.296)	
Romance × E3		1.034^a	
		(0.191)	
First-stage residuals	0.605	0.737	0.714
(Hausman test)	(0.047)	(0.063)	(0.135)
Observations	1530	1530	1530

Bootstrapped standard errors in parentheses (1000 replications).
^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table 9: Regression coefficients when nesting patterns are modified at the margin

	A Mean (p5) (p95)	B Mean (p5) (p95)	C Mean (p5) (p95)
$\ln(\bar{s}_{Eod\|g})$ (σ)	0.62	0.47	0.48
	(0.44)	(0.23)	(0.32)
	(0.80)	(0.80)	(0.74)
Δ ln wage $\times$ No Contig $\times$ E1	0.97	1.43	1.27
	(0.82)	(1.17)	(1.03)
	(1.22)	(1.78)	(1.64)
Δ ln wage $\times$ No Contig $\times$ E2	0.68	0.99	0.91
	(0.56)	(0.75)	(0.71)
	(0.82)	(1.25)	(1.15)
Δ ln wage $\times$ No Contig $\times$ E3	0.31	0.63	0.57
	(0.20)	(0.44)	(0.37)
	(0.42)	(0.79)	(0.74)
Δ ln wage $\times$ Contig $\times$ E1	2.64	2.75	2.77
	(2.33)	(2.37)	(2.38)
	(3.15)	(3.18)	(3.14)
Δ ln wage $\times$ Contig $\times$ E2	1.83	1.88	1.85
	(1.55)	(1.63)	(1.64)
	(2.10)	(2.10)	(2.10)
Δ ln wage $\times$ Contig $\times$ E3	0.95	0.97	0.96
	(0.71)	(0.77)	(0.79)
	(1.22)	(1.19)	(1.18)
ln distance $\times$ No Contig $\times$ E1	-0.95	-1.06	-1.04
	(-1.38)	(-1.45)	(-1.39)
	(-0.77)	(-0.64)	(-0.74)
ln distance $\times$ No Contig $\times$ E2	-0.52	-0.72	-0.74
	(-0.69)	(-1.08)	(-0.98)
	(-0.36)	(-0.22)	(-0.52)
ln distance $\times$ No Contig $\times$ E3	-0.23	-0.40	-0.59
	(-0.48)	(-0.84)	(-0.84)
	(0.01)	(0.20)	(-0.39)
ln distance $\times$ Contig $\times$ E1	-1.77	-2.12	-2.21
	(-2.14)	(-2.55)	(-2.52)
	(-1.37)	(-1.61)	(-1.73)
ln distance $\times$ Contig $\times$ E2	-1.68	-1.89	-1.94
	(-1.86)	(-2.15)	(-2.13)
	(-1.48)	(-1.56)	(-1.60)
ln distance $\times$ Contig $\times$ E3	-0.91	-1.03	-1.04
	(-1.05)	(-1.20)	(-1.17)
	(-0.72)	(-0.78)	(-0.81)
Number of replications	420	740	932

Mean coefficients, 5th and 95th percentiles of the distribution of coefficients.

Appendix A Supplementary tables

Table 10: Heterogeneity of wage and distance coefficients by education, proximity measured by South American dummy, for different nesting patterns

	(1) Nested logit A	(2) Nested logit B	(3) Nested logit C	(4) Nested logit D
$\ln(\bar{s}_{Eod\|g})$ (σ)	0.509^a	0.570^a	0.490^a	0.600^a
	(0.072)	(0.087)	(0.077)	(0.043)
E2	-1.661	-2.304	-2.090	0.185
	(3.487)	(2.953)	(3.071)	(3.702)
E3	-1.583	-5.817^b	-3.608^b	0.535
	(2.737)	(2.779)	(1.681)	(2.871)
South Am.	13.128^a	9.574^b	11.571^a	15.649^a
	(4.510)	(4.128)	(3.741)	(3.702)
South Am. × E2	-1.774	-1.424	-2.780	-4.411
	(4.640)	(4.615)	(6.232)	(5.355)
South Am. × E3	-8.316	-4.415	-8.327^b	-11.693^b
	(5.136)	(4.507)	(4.135)	(4.885)
Δ ln wage × No South Am. × E1	0.956^a	1.700^a	1.426^a	1.347^a
	(0.329)	(0.329)	(0.272)	(0.121)
Δ ln wage × No South Am. × E2	0.621^a	1.255^a	1.117^a	0.810^a
	(0.215)	(0.242)	(0.218)	(0.157)
Δ ln wage × No South Am. × E3	0.209	0.812^a	0.749^a	0.335^a
	(0.186)	(0.280)	(0.244)	(0.082)
Δ ln wage × South Am. × E1	2.340^a	2.325^a	2.376^a	2.482^a
	(0.506)	(0.473)	(0.455)	(0.309)
Δ ln wage × South Am. × E2	1.422^a	1.398^a	1.350^a	1.350^a
	(0.268)	(0.274)	(0.301)	(0.337)
Δ ln wage × South Am. × E3	0.675^a	0.645^a	0.616^a	0.570^a
	(0.180)	(0.191)	(0.212)	(0.121)
ln distance × No South Am. × E1	-0.921^a	-1.469^a	-1.563^a	-0.822^b
	(0.251)	(0.285)	(0.127)	(0.319)
ln distance × No South Am. × E2	-0.485^c	-0.906^a	-1.038^a	-0.555^a
	(0.248)	(0.227)	(0.363)	(0.188)
ln distance × No South Am. × E3	-0.364^a	-0.373^c	-0.747^a	-0.469^a
	(0.100)	(0.205)	(0.169)	(0.100)
ln distance × South Am. × E1	-2.598^a	-2.656^a	-3.063^a	-2.713^a
	(0.544)	(0.521)	(0.489)	(0.324)
ln distance × South Am. × E2	-1.970^a	-1.980^a	-2.210^a	-1.958^a
	(0.317)	(0.329)	(0.315)	(0.477)
ln distance × South Am. × E3	-1.018^a	-1.013^a	-1.155^a	-0.942^a
	(0.208)	(0.212)	(0.223)	(0.332)
First-stage residuals	0.603	0.635	0.773	0.570
(Hausman test)	(0.068)	(0.056)	(0.058)	(0.052)
Observations	1530	1530	1530	3030

Bootstrapped standard errors in parentheses (1000 replications).
^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table 11: First-stage results for models in Table 10

	(1) Nest. logit A	(2) Nest. logit B	(3) Nest. logit C	(4) Nest. logit D
Average wage difference	-1.684[a]	-1.516[a]	-1.757[a]	-2.411[a]
	(0.181)	(0.185)	(0.183)	(0.154)
Average distance	3.377[a]	3.366[a]	2.952[a]	0.337
	(0.336)	(0.383)	(0.405)	(0.219)
Nb. countries by group	-0.415[a]	-0.253[a]	-0.261[a]	-0.053[a]
	(0.101)	(0.018)	(0.023)	(0.004)
Adj. R-squared	0.350	0.399	0.354	0.288
F-test	77.6	141.1	124.2	262.9
F-test df	(3,1509)	(3,1508)	(3,1508)	(3,3008)
Observations	1530	1530	1530	3030

Robust standard errors in parentheses.
^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table 12: First-stage results for models in Tables 6, 7 and 8

	Table 6		Table 7		Table 8	
	(1) 2SLS $n > 0$	(2) 2SLS $n + 1$	(3) W: 20/50/80	(4) W: linear	(5) Language	(6) Dest. FE
Average wage difference	-1.592[a]	-1.682[a]	-1.164[a]	-0.960[a]	-1.608[a]	0.038
	(0.166)	(0.171)	(0.157)	(0.158)	(0.171)	(0.236)
Average distance	2.399[a]	3.213[a]	3.599[a]	3.350[a]	2.606[a]	3.746[a]
	(0.312)	(0.337)	(0.344)	(0.345)	(0.327)	(0.244)
Nb. countries by group	-0.168[a]	-0.216[a]	-0.247[a]	-0.238[a]	-0.179[a]	-0.344[a]
	(0.019)	(0.020)	(0.021)	(0.021)	(0.020)	(0.019)
Adj. R-squared	0.350	0.351	0.351	0.334	0.389	0.769
F-test	100.5	123.6	115.1	97.9	103.4	115.6
F-test df	(3,1372)	(3,1509)	(3,1508)	(3,1508)	(3,1505)	(3,1458)
Observations	1393	1530	1530	1530	1530	1530

Robust standard errors in parentheses.
^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$