J Popul Econ (2012) 25:1077–1105
DOI 10.1007/s00148-011-0356-x

ORIGINAL PAPER

Cultural barriers in migration between OECD countries

Michèle Belot · Sjef Ederveen

Received: 7 October 2008 / Accepted: 17 January 2011 /
Published online: 12 February 2011

Abstract This paper uses a unique set of new indicators enabling us to test the effects of cultural barriers on migration between OECD countries. Using data on migration flows between 22 OECD countries over the period 1990–2003, we find strong evidence for the negative effect of cultural differences on international migration flows. Cultural barriers do a much better job in explaining the pattern of migration flows between developed countries than traditional economic variables such as income and unemployment differentials.

Keywords International migration · Culture · OECD countries

JEL Classification J61 · F22 · O15

1 Introduction

One of the basic principles of the European Union is the freedom of movement of factors of production and, in particular, of workers. In practice, there is

Responsible editor: Klaus F. Zimmermann

Electronic supplementary material The online version of this article (doi:10.1007/s00148-011-0356-x) contains supplementary material, which is available to authorized users.

M. Belot (✉)
Centre for Experimental Social Sciences, Nuffield College, Oxford University, New Road, Oxford, UK
e-mail: michele.belot@nuffield.ox.ac.uk

S. Ederveen
Ministry of Economic Affairs, P.O. Box 20101, 2500 EC The Hague, The Netherlands

little movement, both between and within countries, despite the presence of large economic differentials.[1] Numerous empirical studies establish the low migratory responses to unemployment and wage differentials (see for example Barro and Sala-i-Martin 1995; Gros 1996). Obviously, there are gains from migration and the challenge is to explain why people fail to exploit them. The focus of the literature has therefore shifted towards finding explanations for the "European immobility puzzle".[2]

Low migratory responses to unemployment and wage differentials can explain why European mobility is low, but they fail to explain why migration flows between some countries are almost non-existent, while between other countries there are substantial flows in both directions (see Table A.2 in the Appendix). For example, there are large migration flows between Belgium and France, in both directions; and there is very little migration between Greece and Finland in any direction. Obviously, economic differentials between countries will not help much in explaining this pattern. Cultural proximity, on the other hand, is a strong candidate as it is pair-specific and varies across pairs of countries.

The main contribution of this paper is to investigate the role of cultural barriers by using a wide range of refined indicators of these cultural barriers to migration. Until now, appropriate measures lacked, which explains why the empirical literature has drawn little attention to the role of these costs. The research into the determinants of migration hardly ever goes beyond the inclusion of a simple dummy for sharing a common language. This paper provides unique empirical evidence showing that cultural barriers play a crucial role in migration, far beyond the effect of speaking a different language, and do a better job in explaining migration patterns between developed countries than differentials in economic variables.

Our empirical analysis focuses on a panel of 22 OECD countries over the period 1990–2003. We do not include any developing countries, as our interest is primarily in the determinants of migration between developed countries. We suspect that the mechanisms driving migration between developed countries may be different than those driving migration between developing and developed countries. Economic differentials are expected to be much smaller between OECD countries for example, although they do remain substantial. For example, the GDP per capita in Luxembourg is three times as large as in Greece. The inclusion of both economic and cultural variables allows us to compare the relative effects of both factors on migration. The fact that our sample includes countries outside the European Economic Area also enables us to evaluate the effect of the law allowing free mobility of people across borders.

[1] See for example Decressin and Fatas (1995).

[2] Braunerhjelm et al. (2000) present a detailed overview of this literature.

To estimate the migration model we use a population-averaged negative binomial model. This is a methodological contribution to the empirical literature. With this estimation technique we are able to cope with several econometric problems that the empirical migration literature is frequently plagued with. These problems arise because migration flows are not distributed in the standard way. Migration flows are always discrete and non-negative, and small numbers are overrepresented. This is often dealt with by log-linearizing the equation. However, this technique may not be appropriate in the presence of heteroskedasticity. Our estimation technique therefore seems preferable to what has been used before in the literature.

The empirical analysis presented in this paper yields important insights in the effects of economic and cultural variables on migration between OECD countries. It shows that especially the cultural and linguistic links between countries are very important in explaining migration flows. The impact of a one standard deviation increase in linguistic distance on migration flows between two countries is for example much stronger than the effect of a one standard deviation increase in unemployment.

Before discussing the econometric results, we will give an overview of the related literature in the next Section. Section 3 presents a simple theoretical model of migration that justifies the empirical strategy. The construction of the different cultural distance measures is discussed in Section 4. Section 5 is devoted to the estimation of the migration equation. It discusses the data set and other variables included in the analysis, the econometric specification and the results. Section 6 concludes.

2 Related literature

There is an extensive literature on the theoretical determinants of international migration (see Ghatak and Levine (1996), Borjas (1999a) and Hatton and Williamson (2005) for surveys). The determinants that explain international migration can be classified into a few broad categories: economic incentives, demographic explanations, distance and network effects. We briefly describe these categories and discuss the empirical literature.

The importance of economic incentives is stressed by the classical model of Harris and Todaro (1970). It assumes that individuals base their migration decision on the differential between the expected income at destination and the expected income at home. The model predicts that economic differentials should lead to compensating migration flows. Economic differentials should then decrease over time.

Given the large persistent economic differentials observed across OECD countries, this basic model does not seem to fit with the picture of the developed world, since we observe persistent and large economic differentials between countries, and little migration between them. The existence of economic

differentials suggests that there are economic incentives to migrate between countries, but people do not seem to exploit them. Note that differentials in expected income are not the only relevant economic incentives we should take into account, especially when focusing on the developed world. Countries differ in their social security and tax systems. These differences could also influence migration decisions.

It could be that levels of income matter more than differentials (Braunerhjelm et al. 2000). Developed countries provide a socially acceptable income to almost everyone. The generosity of the welfare system could decrease the incentives to migrate, precisely for those who should migrate (see Chorny et al. 2007, for a comprehensive discussion). For example, Antolin and Bover (1997) find for Spain that registered unemployed (entitled to benefits) are less likely to move than non-registered unemployed (not entitled to benefits). Welfare recipients have also fewer incentives to migrate, as they would probably lose their rights to welfare benefits. Similarly, some argue that families may offer income substitutes in case of negative shocks and reduce the need to migrate.

Next, we find a series of explanations based on the composition of the population, and on the idea that some segments of the population are more inclined to migrate than others (demographic explanations). The first is the *age structure* of the population: young cohorts are more mobile than old cohorts (see Fertig and Schmidt 2002). All else equal, the aging of the population in the developed world would lead to a fall in the average migration rate. There could be several reasons why older workers are less likely to migrate (Tassinopoulos and Werner 1999): older workers have acquired more specific human capital, which could be lost in case of migration; they have fewer years to recoup their migration investment; or they may face higher migration costs (stronger social ties, higher costs to learn a foreign language, etc.).

The second demographic explanation has to do with the *education structure* of the population: higher skilled workers are more likely to migrate than lower skilled workers (e.g., Wildasin 2000 for the United States, Mauro and Spilimbergo 1999 for Spain, and Gianetti 2001 for Italy). There is also a large evidence based on micro data. Education increases significantly the probability of moving. The reason why this is the case could either be that high-skilled workers face lower migration costs or that they gain more from migrating. High-skilled workers are more likely to speak another language or have better qualities of adaptation, which makes it less costly for them to migrate. Similarly, if migration costs are to some extent fixed, the relative burden of migration costs will be higher for low-skilled than high-skilled workers. The former may be credit-constrained and not be able to pay the migration costs (see Pedersen et al. 2008). High-skilled workers may also gain more by migrating, if for example wage differentials are larger for high-skilled workers than low-skilled ones.

Finally, a third demographic factor that may play a role is *female participation*. Coordinating migration decisions of two-earner households may be more difficult than of one-earner households. We could therefore expect that

countries with high participation rates of women have a lower propensity to migrate.[3]

Another determinant of migration often mentioned in the literature is the distance between countries, defined in a broad way. First, the physical distance between the place of origin and the place of destination could discourage migration for two reasons: it is a psychic cost (and direct migration cost) and it reduces the quality of information about the destination. The further away the country, the less likely people will be informed about job opportunities, income differentials, etc. The argument was already mentioned by Sjastaad (1962). We could use the same argument for migration costs: the further away the country of destination, the worse the information people have about costs they will need to incur when migrating.

The importance of cultural links is also at the centre of the recent literature on networks. The main idea is that the presence of a national community in the country of destination could increase its attractiveness (Carrington et al. 1996). All over the world, we find "Little Italy's", "Chinatowns", etc. showing that ethnic groups tend to cluster in some geographical areas. There are many ways these communities could ease the immigration of their national counterparts. For example, they could provide information about the local customs and values, job opportunities, etc. They could also provide a substitute to the social network in the country of origin. It may indeed be easier to migrate to a geographical area with a high concentration of people sharing the same language and culture than one's own. A number of empirical studies find supporting evidence for the existence of network effects (Munshi 2003; Mayda 2010; Pedersen et al. 2008; Hatton and Williamson 2006; Clark et al. 2004; McKenzie and Rapoport 2007). All these studies include both developing and developed countries. It is not certain, however, that networks play an important role in international migration between developed countries. Gross and Schmitt (2003) find that cultural communities are more attractive for immigrants from non-OECD source countries than from OECD source countries.

Most empirical studies find a strong correlation between the size of the national community in the country of destination and the importance of migration flows. One should be careful, however, in the interpretation of this coefficient. Palloni et al. (2001) discuss the fact that the existence of networks is not the only theory that could explain the positive correlation. Other theories lead to exactly the same predictions. First, there is the "common characteristics and constraints" theory, saying that individuals living in the same region are likely to share common characteristics and constraints influencing the migration decision. Controlling for a wide set of characteristics of the country of origin and destination can help reducing the influence of this effect in the observed correlation. Second, migration decisions often involve households

[3]Changes in female labor participation in the destination country may matter as well. Ederveen et al. (2007) show that the sustenance of low labor mobility despite regional disparities can partly be explained through the adjustment in female labor participation.

and families, instead of individuals. Migration decisions of the same family will be correlated, even in the absence of network effects. We should keep this in mind when turning to the empirical analysis.

The lack of data on aggregate migration flows between countries compromised for a long time the empirical testing of migration theories. The only aggregate data available were on net immigration rates, i.e., simply the difference between the growth rate of the population and the natural growth rate. The problem is that the net immigration rate confuses information about inflows and outflows, and is a unilateral measure. Since migration inflows and outflows rarely involve two countries only, this type of measure makes it impossible to test the Harris–Todaro theory that assumes that migration decisions are based on economic differentials between pairs of countries. Gross inflows and outflows between two countries provide much better information than net immigration rates. It is hard, however, to obtain consistent data on gross inflows and outflows between pair of countries. Countries do not agree on the definition of a migrant, they differ in the way they measure inflows, etc. For these reasons, migration data has been incomplete and unsatisfactory for many years.

Van Wissen and Visser (1998) is one of the rare studies on migration flows between a large number of developed countries. They use data on gross migration flows between the fifteen European countries of the EU (before May 2004) for the year 1994. They find strong effects of the variables measuring the stocks of foreigners in the country of origin and of destination. On the other hand, differentials in GDP do not have a significant effect, neither do physical distance and language proximity (the author introduces dummies for four language groups and classify the countries accordingly).

Two recent papers (Mayda 2010; Pedersen et al. 2008) analyze the determinants of gross migration flows into OECD countries, testing for a series of migration theories. Mayda (2010) uses OECD data on 14 OECD countries, over the period 1980–1996. She finds that the earnings differentials stimulate migration, and that this effect is dominated by the pulling effect of the GDP per worker at destination. The GDP per worker at origin does not have a strong effect, which could be justified by the combination of fixed migration costs and binding poverty constraints. Physical distance matters as well, but sharing a common language does not. Finally, sharing a common colonial past has a surprising negative effect on emigration rates.

One drawback of the OECD data is that they do not report all flows between countries. Small flows in particular are likely to be underreported. Pedersen et al. (2008) have constructed an impressive data set including 27 OECD destination countries and 129 source countries, for the period 1990–2000. They show that the determinants of emigration differ across countries. They grouped countries in various ways (according to the income level and the type of welfare state) and identified clear patterns in migration determinants. They find strong evidence of network effects in all countries, but these effects are stronger when destination countries offer a limited social protection to immigrants. Also, they find no support for the welfare magnet hypothesis

(i.e., immigrants being attracted by generous welfare systems in developed countries). Their indicator of generosity of welfare system, the degree of tax pressure, does not have a different effect on poor or rich countries. Finally, they find a different result from Mayda (2010) with respect to the effect of sharing a common language or a colonial past. In their specification, both have a positive effect on migration flows.

The contribution of this paper in the context of the existing literature is threefold: first, we focus on migration flows between *developed* countries, because we suspect that the mechanisms driving migration between these countries might be different than those driving migration between developing and developed countries. Second, by focusing on developed countries only, we can investigate specifically the role of cultural barriers to migration by including a number of refined indicators capturing cultural distance between countries. Third, we present results based on an econometric specification which specifically addresses important econometric issues associated with this type of analysis.

3 Theoretical background

3.1 Theoretical background

Our ambition is not to provide a new theory of migration. Nevertheless, it is helpful to present a simple model of migration to structure our thoughts. Suppose a world with N countries ($N = 1, \ldots j, \ldots N$). An individual i in a given country considers migration opportunities in $(N-1)$ other countries. Let us denote the income of individual i in country j at time t by $Y_{i,j,t}$. Income consists not only of labor income but could also include social welfare benefits, etc.

Let us assume that $Y_{i,j,t}$ can be split into a "country-specific" component $\overline{Y}_{j,t}$ and an "individual-specific component" μ_i, the latter being randomly distributed across the population:

$$Y_{i,j,t} = \overline{Y}_{j,t} + \mu_i \tag{1}$$

The individual-specific component captures differences in education level, ability, etc.

Denote h the home country and n the destination country, such that the income of individual i in his home country at time t is denoted by $Y_{i,h,t}$ and her income in the destination country at time t is denoted by $Y_{i,n,t}$. Each migration opportunity is associated with migration costs $C_{i,h,n,t}$ (costs of migration for individual i to migrate from country h to country n at time t). Obviously, we have $C_{i,h,h,t} = 0, \forall i$. Migration costs can be split in three parts: Costs independent of the country of destination (e.g. loss of local social network, loss of social security rights, etc.), costs independent of the country of origin (local regulations) and costs specific to the bilateral combination of origin and destination (physical and cultural distance). Each component can be further

split in a "non-individual-specific" component and an "individual-specific" component, the latter modelled as random component.

The net gain from migration is then:

$$NetG_{i,n,t} = Y_{i,n,t} - C_{i,h,n,t} \tag{2}$$

Suppose individuals have perfect information about their utilities in each location and about the migration costs associated with each move. The optimal location decision for individual i living in country h (at time $t-1$) is:

$$n^*_{i,h,t-1} = \arg\max_n NetG_{i,n,h,t} = \arg\max_n \left[Y_{i,n,t} - C_{i,h,n,t}\right] \tag{3}$$

Suppose that for each pair of countries, we can order the individuals according to their net utility $NetG_{i,n,h}$. We denote the density function we obtain by $f_{h,n}(i)$. The dependent variable in our model is the "gross migration flow" between two countries in a given year, that we denote $M_{j,k,t}$, as the flow from country j to country k in year t. The aggregate gross migration flow from country h to country n is then:

$$M_{h,n,t} = \sum_i \left(f_{h,n,t}(i) / n^*_{i,h,t-1} = n\right) \tag{4}$$

Most theoretical models come down to a specification of the following form:[4]

$$M_{h,s,t} = g\left(\overline{Y}_{h,t}, \overline{Y}_{n,t}, \overline{C}_{h,n}, \overline{W}_{h,t}\right) \tag{5}$$

where $M_{h,n,t}$ is the migration flow from country h to country n, $\overline{Y}_{h,t}$ and $\overline{Y}_{n,t}$ are country-specific elements (GDP per capita, unemployment rate, population), $\overline{C}_{h,n}$ are the costs of migration from country h to country n and $\overline{W}_{h,t}$ is an aggregate measure of the individual-component in the migration costs and income: share of young people in the total population, participation rate of women, etc. Note that not all variables vary over time.

This model is a static model and one may rightly argue that migration decisions are based on difference in lifetime expected values rather than on differences in current values. This is relevant since costs are likely to be incurred immediately, while the benefits may accrue over a longer period of time. We should bear this in mind when interpreting the results. Secondly, the migration process may be dynamic in essence. In particular, migration decisions are likely to be correlated over time through the formation of networks in the destination which may reduce the costs of migration for future migrants. The presence of network effects would imply that migration flows in period t are determined by past migration flows. The worry in terms of

[4]Some studies have the migration rate as a dependent variable instead of the number of migrants. The migration rate can potentially be problematic as a dependent variable if the size of the population in the denominator (origin or destination) is correlated with economic variables (see Young (1975) for an extensive discussion on the choice of the dependent variable in migration studies). We chose to enter the relative sizes of the population of the countries of origin and destination as regressors to avoid this problem.

our empirical analysis is that these dynamic effects may bias the estimates of the other variables. For example, suppose that migration decisions are mainly driven by network effects. Then, if the size of these networks is correlated with for example our distance variables, then our estimates of the effects of these variables on current migration flows will be biased. This is a concern if, for example, these variables may have mattered in the past and matter less now. If we do not control for network effects, we may wrongly conclude that these distance variables have affected migration flows in the most recent decades. We will address this issue in two ways: (1) we will control for the size of local communities (networks) explicitly, (2) we will investigate to what extent these networks are themselves correlated with our independent variables.

4 Cultural barriers to migration

Cultural distance between countries seems a priori appealing to explain mobility patterns between European countries, since there are indeed large linguistic and cultural differences between them. Differences in culture, language, values and norms translate into migration costs that reduce the attractiveness of migration. The role of cultural proximity has received some attention in the migration literature, particularly because of the empirical observation of geographical clustering of immigrants. It is well-known that immigrants tend to locate in areas populated by people with similar ethnic backgrounds. Edin et al. (2003) show that this geographical clustering can to some extent help immigrants settling in the country and enhances their economic success. We can extrapolate the argument at a higher level. If there are benefits to living in a cultural environment that is close to one's own cultural background, one would expect to observe larger migration flows between countries that are culturally close to each other.

The concept of culture is hard to define, let alone to measure in a meaningful way. Broadly defined, culture is the set of communication habits, norms, values which are shared by a community. There are many potential problems associated with measuring subjective aspects of culture. For this reason, we concentrate first on the more objective characteristics of culture: language and religion. Then, we introduce two indicators based on measures of norms and values from the sociological literature.

A major challenge in the construction of these indicators comes from the fact that the cultural and linguistic profile of countries reflects, at least partly, past migration flows. Take, for example Luxembourg, which has seen large immigration flows from Portugal in the seventies and now counts 14.4% of Portuguese-speaking people in the population.[5] Because of this large migration movement, one could argue that Portugal and Luxembourg are now culturally closer to each other. The challenge is then to say whether it

[5]Data from ethnologue (www.ethnologue.com).

is the cultural proximity that triggers the current migration flows between Portugal and Luxembourg, or whether it is simply the case that the tastes of Portuguese migrants are correlated, such that they will keep migrating to similar destinations over time. We build our indicators in a way that should minimize the chance that they will capture simple correlations of tastes.

The indicators of cultural and linguistic differences that we have constructed are much more refined than what has been used in the literature so far. Empirical studies have often used very rough and approximate indicators, like a dummy variable for common language. Our measures should provide a significant improvement on these existing indicators. We use four different variables: linguistic distance, religious distance and cultural distance measures based on two different sources. In the empirical analysis, we will show that these indicators have a substantial added value.

Linguistic distance We first propose an indicator of the linguistic distance between two countries. We concentrate on official languages instead of spoken languages, as the latter are more likely to reflect recent migration trends. Official languages have long been in place. It is harder to believe in a correlation between the current migration flows and those which triggered the establishment of official languages. The indicator of linguistic distance is defined as follows:

$$1 - \max_{\forall i \in A, \forall j \in B} \{proximity\{i, j\}\}$$

where i and j are indices for official languages in country A and B, respectively, and proximity is the measure constructed by Dyen et al. (1992), a group of linguists who built a measure of distance between Indo-European languages based on the proximity between a sample of words from each language. For countries that have more than one official language, we choose the highest value of proximity measured on all possible pairs of official languages.

Our indicator ranges from 0 for no linguistic distance to nearly 1 for high linguistic distance. For example, Belgium and France have a distance value of 0 because both have French as one of their official languages. For Switzerland and Portugal, which do not share a common language, we use the distance measure of Portuguese versus Italian, as they are linguistically closest according to the proximity measure. We refer to the appendix for more details. We use the same distance measures for Icelandic and Norwegian, as Norwegian was not in the data but belongs to the same sub-group in the linguistic classification tree (see Appendix). Finnish was not in the data either, because it is not an Indo-European language. Since it belongs to another language group, we set the distance equal to 1 for all Indo-European languages. The data are presented in the appendix (Table A.3).

Religious distance Most countries do not have an official "religion", so we use survey statistics to build an indicator of religious proximity. The large majority of the population in OECD countries belongs to one of the three

Christian groups (Catholicism, Protestantism, and Orthodoxy). The measure we constructed corresponds to the probability of drawing two individuals, one in each country, who would have a different religion. The corresponding formula reads as:

$$\text{DISTREL}_{A,B} = 1 - \sum_i s_{i,A} s_{i,B}.$$

The religious distance is equal to 1 minus the probability of drawing two people with the same religion. $s_{i,A}$ and $s_{i,B}$ are the respective shares with religion i in countries A and B.

To minimize the influence of recent migration flows, we only counted religious groups that are larger than 1% of the population. Note that countries are always strongly dominated by one specific religion, such that the role of migration movements in determining the current religious beliefs has probably been small anyway. The data are reported in the appendix (Table A.4).

Another way of measuring cultural differences is by measuring norms and values directly. Sociologists have built a number of measures of "norms and values", and have collected data on these measures across a large number of countries. We will use two large sets of measures widely used in the sociological literature.

Cultural distance (Hofstede) The first comprehensive set of measures is provided by (Hofstede 1991, 2001). He proposed a measure of cultural orientation of countries, based on a survey of 117,000 IBM employees across 50 countries and 3 multi-country regions. The original data were collected at two different points in time: 1968 and 1972. Although these data are collected more than 40 years ago, they are assumed to reflect values that are strongly embedded in centuries-old cultures. To further ensure the timeliness of the identified cultural concepts, only data that remained stable across both surveys were kept. According to Hofstede (2001), recent replications show no loss of validity.

Hofstede distinguished at first four and later five dimensions on which country cultures differ. The original four cultural dimensions with the corresponding definitions from (Hofstede 2001, pp.xix-xx) between brackets are:

1. Power distance ("the extent to which the less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally")
2. Uncertainty avoidance ("the extent to which a culture programs its members to feel either uncomfortable or comfortable in unstructured situations")
3. Individualism versus collectivism ("the degree to which individuals are supposed to look after themselves or remain integrated into groups")
4. Masculinity versus femininity ("the distribution of emotional roles between the genders")

Later, Hofstede added a fifth dimension of national culture. This new dimension grasped the long-term versus short-term orientation (Confucian dynamism) in a society and was independent of the IBM survey, but instead based on a Chinese values survey. As the source is different and the data are not available for all countries, we exclude this dimension in our calculation of a measure of cultural distance. So we used the original four dimensions to compute a composite index of cultural distance between countries, following the strategy proposed by Kogut and Singh (1988):

$$\mathrm{CD}_{i,j} = \frac{1}{4}\sum_{k=1}^{4}\frac{(I_{i,k} - I_{j,k})^2}{V_k}$$

where $\mathrm{CD}_{i,j}$ is the cultural distance between country *i*and country *j*, $I_{i,k}$ is the Hofstede's score for country *i* with respect to the cultural dimension *k*. Finally, V_k is the variance of the indicator of dimension *k* for all countries included in the sample of Hofstede. Data are reported in the appendix (Table A.5).

Cultural distance (Inglehart and Baker) The second comprehensive set of cultural measures is the World Value Survey. The Survey covers 65 countries in the world, including 18 of the countries included in our data set. Inglehart and Baker (2000) did a factor analysis based on the various waves of the survey and summarized the data around two major dimensions (*dimension1* and *dimension2* here after): (1) traditional versus secular-rational and (2) survival versus self-expression values. Traditional societies are defined with respect to a series of variables such as the level of tolerance for abortion, divorce and homosexuality, the emphasis of male dominance in economic and political life, the importance of family life and parental authority and the emphasis on religion. The survival/self-expression dimension corresponds to the level of trust, tolerance, subjective well-being, political activism, and self-expression.

On the basis of these two dimensions, we computed an indicator of cultural distance between 2 countries (*i* and *j*) as follows:

$$\mathrm{DistInglehart}_{i,j} = \sqrt{(\mathrm{Dimension1}_i - \mathrm{Dimension1}_j)^2 + (\mathrm{Dimension2}_i - \mathrm{Dimension2}_j)^2}$$

The data are reported in the appendix (Table A.6).

Interdependence between the various cultural measures One may wonder whether these indicators are strongly correlated with each other and therefore not add very much independently. We report the correlation coefficients between these different measures, including a measure of physical distance in Table A.7 (Appendix).

The correlation coefficients between the different distance measures are surprisingly low and often even negative. This suggests that these measures capture different dimensions of culture. Linguistic distance has a clear interpretation. The costs of communication directly increase with linguistic distance. For example, it is often not very hard to understand a different language that is linguistically close to your own and therefore helps to reduce migration

costs. The measure of religious distance broadly groups pairs of countries into two groups: countries that are either very close to each other (when they belong to the same religious tradition) or far away from each other (when they do not belong to the same religious tradition). Both the Hofstede and the Inglehart and Baker measures seek to find an overarching pattern of norms and values and show much more variation across pairs of countries. Both measures summarize culture along different dimensions and the underlying data sources differ substantially. Therefore, it could be expected that they also grasp different dimensions of culture. A good example illustrating the differences between these measures is Greece. Greece is far away from the other countries both in terms of linguistic and religious distance; but there is much more variation in the sociological measures of cultural distance. Thus, in summary, we would expect that these various measures complement each other in providing a richer picture of cultural proximity between countries. A comprehensive discussion of the different cultural dimensions can be found in Vinken et al. (2004).

5 Estimation of the migration equation

In this section, we present our estimation strategy. Before turning to the estimation results, we first give a brief description of the data we use and discuss the econometric specification in some detail.

5.1 Data

We will now describe successively the variables and data used in the empirical analysis. Details about coverage and sources for all variables can be found in the appendix.

5.1.1 Migration flows

We use different sources to get a complete overview of the relevant migration flows. We start with collecting data from the OECD, based on the Continuous Reporting System on Migration (SOPEMI). The drawback of these data is that they include flows from a selected number of countries of origin only. Small inflows will be grouped by region or under the label "other countries". In order to constitute a more detailed data set, we use information provided by the Migration Policy Institute, using the same sources as the OECD (national statistical offices), but reporting more detailed information. These data were available for Australia, Austria, Denmark, Finland, Germany, the Netherlands, Norway, Sweden, the United Kingdom and the USA.[6] For the other countries, we use information provided by the National Statistical

[6]These data are on-line on the website www.migrationinformation.org.

Offices. Table A.2 in the appendix reports the average migration flows between any pair of countries in our data set.

Although we believe that the data we use are the most appropriate for this empirical research, they have their limitations as well. We should be aware of the special features of our data set. First, the data measures movements in *population* rather than in labor. It is an aggregate measure composed of different types of migrants with different motives. Second, the data is not perfectly homogenous across countries. Countries register migration flows in different ways. The most common way of registering foreigners is by citizenship. Some countries, however, register the foreign population according to their country of birth or country of previous residence. Third, another important difference between countries is the timing of the registration (duration of stay). The European Economic Area agreement has modified the registration requirements for citizens of these countries. This is mainly a problem for the UK data, since the UK stopped requiring a grant settlement. The number of people registered from European Economic Area (EEA) countries therefore underestimates the actual flows. Moreover, the UK stopped registering migrants from EEA countries since 1998.

5.1.2 Independent variables

Population sizes Since our dependent variable is the flow of migration between two countries over a given year, we need to control for population sizes at origin and destination. The inclusion of the population size at origin is intuitive, the inclusion of the population size at destination allows us to control for possible gravity effects. This specification is the most flexible specification to take account of possible gravity or scale effects, in particular it does not impose the elasticity of outflows to population to be equal to 1.

Economic variables We will control for the lagged GDP per capita (in constant $ prices and PPP adjusted), as well as for the unemployment rates at origin and destination. These should capture the economic push and pull factors.

Demographic variables We will control for a number of variables describing the demographic distribution of the population of the country of origin in dimensions that may matter for migration: share of tertiary educated, participation rate of women and the share of young people (20–39) as percentage of the total population. Also, we control for the population at origin and at destination.

Cultural barriers We will control for the various measures of cultural distance described above. In addition, we will control for the share of population from the country of origin in the country of destination to capture possible network effects.

Migration policy Countries differ in their immigration policies and regulations. Developed countries, and in particular the European Union, often have a dual system, i.e., imposing different regulations according to the country of origin. The most striking example is provided by the European Economic Area, where the movement of workers across borders is free. We directly measure for the effect of "open borders" by introducing a dummy variable equal to 1 if both countries allow free movement of workers between them (European Economic Area or New Zealand–Australia).

Additional variables Other variables will be considered in the robustness analysis, such as measures of the generosity of the unemployment benefit system (gross replacements rates at origin and destination), as well as generosity of the social security system (share of social expenditures in the GDP), to capture possible welfare magnet effects.

5.2 Econometric specification

The independent variable in our analysis is the total inflow from country j to country k. Figure 1 shows a histogram of this variable, where inflows have been grouped by intervals of ten people. The distribution is extremely skewed to the left, with a very high frequency of small numbers. A second important characteristic of our dependent variable is its discrete and non-negative nature. For these reasons, standard linear regression techniques may not be the most appropriate method to analyze these data.

There are two alternative methodologies one could use to deal with this type of data in a better and more efficient way. First, we could simply transform the dependent variable in its logarithm, which would come down to estimating a general linearized model. This method has been used previously in the literature. However, a recent paper of Santos Silva and Tenreyro (2006) shows that the estimated elasticities then can be highly misleading as they neglect the presence of heteroskedasticity. They illustrate their argument for a particular

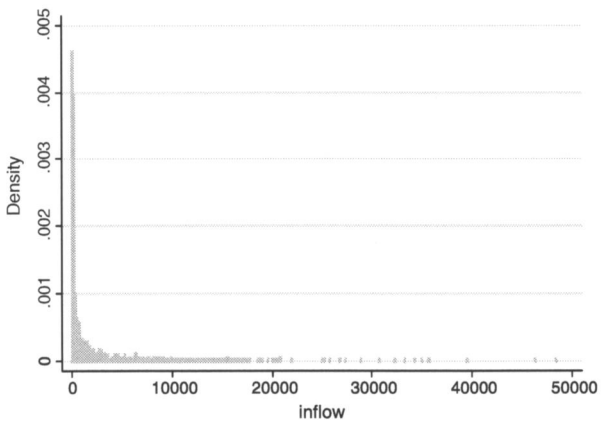

Fig. 1 Histogram of inflows (width = 10)

application, the gravity equation for trade, but the same argument holds for the topic under consideration here.

Therefore, we opt for a different strategy. We estimate a negative binomial model (Cameron and Trivedi 1986), which better fits the distribution of our dependent variable and allows for overdispersion. This model is based on a Poisson model, introducing a random component in the Poisson parameter. Formally, it assumes the dependent variable to be a random draw from a Poisson distribution, with mean and variance $\lambda_{n,h}$ with

$$\lambda_{h,n} = \exp(Z'_{h,n}\beta),$$

where $Z'_{h,n}$ is a vector of all characteristics mentioned here above.

The probability function is then:

$$P(M_{h,n} = k) = \frac{\lambda_{h,n}^{k} \exp(-\lambda_{h,n})}{k!} \tag{6}$$

$$\log \lambda_{h,n} = \beta' \overline{Z}_{h,n} + \log u_{h,n},$$

where u follows a gamma distribution with unit mean and variance θ.

The marginal distribution for the dependent variable is then:

$$f(M_{h,n}/\lambda_{h,n}, u_{h,n}) = \frac{(\lambda_{h,n}u_{h,n})^{k} \exp(-\lambda_{h,n}u_{h,n})}{k!} \tag{7}$$

and therefore:

$$f(M_{h,n}/\lambda_{h,n}) = \int \frac{(\lambda_{h,n}u_{h,n})^{k} \exp(-\lambda_{h,n}u_{h,n})}{k!} g(u_{h,n}) du_{h,n} \tag{8}$$

The expected mean of the dependent variable $E(M_{h,n}) = \lambda_{h,n}$ and the variance is $var(M_{h,n}) = \lambda_{h,n}\left(1 + \frac{1}{\theta}\lambda_{h,n}\right)$, which implies overdispersion (θ is a fixed parameter). Both specifications, the linear model with the log-transformed dependent variable and the negative binomial model, are examples of generalized linear models. These generalized models may require additional adjustments in order to take account of the structure of the data. First, our data has a panel structure, i.e. we have repeated observations for each panel. The assumption of independency of observations across time (*within* panels) is unlikely to be satisfied. To test this, we analyze the residuals of a linear regression of the log-transformed dependent variable including all cultural variables as covariates.[7]

A Wooldridge (2002) test confirms the presence of autocorrelation in the error terms of the same panels: The F statistic $F(1,309) = 8.957$, which rejects the hypothesis of no first-order correlation at the 1% level. One way of correcting for the within-panel correlation is to estimate a population-averaged negative binomial model which specifies the within-correlation structure of the panel directly. We will follow this approach in this paper.

[7]The exact results are not presented here, but are available from the authors upon request.

A second characteristic of our data is that we observe for each country a series of inflows and outflows. Flows involving the same country are likely to be correlated if we do not control for country-specific factors. For example, there might be reasons why the United States attracts migrants of all countries which are not directly observed. We could therefore face a problem of correlation *between* the error terms of different panels. It appears that the problem of cross-correlation is mainly present between groups involving the same country of destination. We therefore introduce fixed effects for the country of destination to correct for the correlation between panels.

5.3 Estimation results

We present here the results of the negative binomial specification. We will present a robustness analysis with alternative econometric specifications in Section 5.3.6.

5.3.1 Traditional variables

Table 1 (col. (1)) presents estimation results corresponding to a "traditional" specification, including controls for economic and demographic variables, as well as measures of physical proximity (physical distance and border sharing).

Table 1 Determinants of migrant flows

Dependent variable	(1) Inflow	(2) Inflow	(3) Inflow	(4) Inflow
Lagged GDP/cap dest. (×$1,000)	0.0339 (0.0085)**	0.0289 (0.0059)**	0.0406 (0.0092)**	0.0500 (0.0094)**
Lagged GDP/cap origin. (×$1,000)	−0.0309 (0.0086)**	−0.0128 (0.0063)*	−0.0520 (0.0092)**	−0.0611 (0.0093)**
Lagged unempl. rate dest.	−0.0273 (0.0046)**	−0.0284 (0.0031)**	−0.0295 (0.0050)**	−0.0299 (0.0051)**
Lagged unempl. rate origin	0.0031 (0.0045)	0.0004 (0.0030)	−0.0059 (0.0048)	−0.0088 (0.0049)
Population dest.	−0.0052 (0.0042)	−0.0056 (0.0029)	−0.0060 (0.0045)	−0.0061 (0.0046)
Population origin	0.0084 (0.0010)**	0.0073 (0.0010)**	0.0094 (0.0010)**	0.0100 (0.0010)**
Share tertiary educated origin	−0.0054 (0.0070)	−0.0087 (0.0073)	−0.0165 (0.0074)*	−0.0091 (0.0074)
Share young origin	0.0314 (0.0436)	0.0112 (0.0444)	0.0554 (0.0441)	0.0615 (0.0440)
Partic. rate women origin	0.0200 (0.0039)**	0.0109 (0.0029)**	0.0208 (0.0041)**	0.0175 (0.0042)**
Distance (1000 km)	−0.0584 (0.0133)**	−0.0649 (0.0135)**	−0.0660 (0.0132)**	−0.0595 (0.0132)**
Border sharing	1.1490 (0.1853)**	0.8491 (0.2005)**	0.6056 (0.2060)**	0.4125 (0.2090)*
Open borders	0.0672 (0.0218)**	0.0655 (0.0147)**	0.0814 (0.0237)**	0.1137 (0.0241)**
Common language		0.8338 (0.2096)**	0.5588 (0.2190)*	0.7658 (0.2202)**
Linguistic distance			−1.2395 (0.2636)**	−0.8045 (0.2702)**
Religious distance				−1.3956 (0.2333)**
Constant	5.6475 (1.3456)**	6.4156 (1.3518)**	5.3886 (1.3458)**	6.1121 (1.3528)**
Observations	2698	2698	2698	2698
Number of panels	314	314	314	314

Standard errors in parentheses
* $p = 0.05$; ** $p = 0.01$

Let us first comment on the effects of the "traditional" economic variables. The signs of the GDP-variables correspond to what we would expect, i.e. migration flows tend to go from poorer countries to richer countries. The effects of a 1% higher GDP per capita in the destination country or a 1% lower GDP per capita in the origin country are similar; they both raise the flow of migrants between both countries with about 0.6%. This is a bit higher than what is found on average in the empirical literature, but falls well within the range of reported results. Ederveen and Bardsley (2004) present a meta-analysis of 26 empirical studies and conclude that the average wage elasticity is 0.43%. These numbers may look rather small, but note that they capture the effects of differences in income on migration in one year. Over a number of years, a permanent difference of 1% in income has large implications for migration.

Second, we find that the unemployment rate at destination decreases immigration flows. The implied elasticity of our estimate for the destination country is that a 1% higher unemployment rate (i.e., 8.08% instead of 8.00%) lowers migration with 0.21%. This estimate is well in line with earlier research. Ederveen and Bardsley (2004) report an average elasticity of −0.15. The effect of the unemployment rate in the country of origin on the other hand is not statistically significant. This is in line with the findings in a number of earlier studies, e.g., Faini and Venturini (1994) and Hatton and Williamson (2005), and could be due to the fact that unemployed people are credit-constrained and are therefore not able to migrate.[8] One extension we will look at in one of the following sections is whether the generosity of the unemployment benefit system has any influence on migration flows.

Turning to the role of demographic characteristics, we find a positive, but not significant correlation between the share of young people in the country of origin and migration flows. The other two demographic variables have an effect opposite to what we would expect. Indeed, our estimates suggest that the share of tertiary workers has a *negative*, although insignificant, impact on immigration and that the participation rate of women *increases* immigration flows. These effects are opposite to what we would expect. The story may be that these indicators capture some attributes of more developed countries, often characterized by a higher level of human capital and a higher participation rate of women. Female participation may also have a positive effect on migration flows because it increase the stock of potential labor migrants.

Finally, we find that the physical distance, measured in kilometers between capital cities has a negative effect on migration flows. Every 100 extra kilometers of distance lower migration with 0.6%. Sharing a border, on the other hand, significantly increases the flows. Migration flows between neighbouring

[8]In addition, there might be a trade-off with respect to GDP per capita. The meta-analysis of Ederveen and Bardsley (2004) concludes that studies that include only unemployment and not GDP per capita report significantly higher coefficients.

countries are more than twice as large as flows between countries without a common border. Finally, we find that there are significantly larger migration flows between countries that allow free movement of people.

5.3.2 Cultural barriers

We now turn to the role of cultural distance between countries. We first start by introducing the simple dummy for sharing a common language, which is the most common way of controlling for linguistic proximity in the literature so far (Table 1, column (2)). We find that, indeed, sharing a common language significantly increases migration flows. In fact, the effect is comparable in magnitude to the effect of sharing a border. We then introduce our additional indicators of linguistic distance (col. (3)) and religious distance (col. (4)). We find that these two indicators also have a significant negative effect, in addition to the common language dummy. We also notice that the estimated importance of a common border decreases with the inclusion of cultural distance measures. This means that not only the fact that countries share a common border and

Table 2 Determinants of migration inflows—the role of culture

Dependent variable	(1) Inflow	(2) Inflow	(3) Inflow	(4) Inflow
Lagged GDP/cap dest.	0.1238 (0.0228)**	0.1203 (0.0227)**	0.0829 (0.0178)**	0.0917 (0.0192)**
Lagged GDP/cap origin	−0.0877 (0.0185)**	−0.0821 (0.0184)**	−0.0507 (0.0162)**	−0.0607 (0.0174)**
Lagged unempl. rate dest.	−0.0211 (0.0085)*	−0.0208 (0.0085)*	0.0015 (0.0072)	0.0013 (0.0079)
Lagged unempl. rate origin	−0.0046 (0.0079)	−0.0033 (0.0079)	−0.0098 (0.0068)	−0.0072 (0.0074)
Population dest.	−0.0148 (0.0067)*	−0.0140 (0.0067)*	−0.0051 (0.0053)	−0.0053 (0.0057)
Population origin	0.0102 (0.0011)**	0.0104 (0.0011)**	0.0102 (0.0011)**	0.0103 (0.0011)**
Share tertiary educated origin	−0.0065 (0.0088)	−0.0046 (0.0087)	−0.0195 (0.0087)*	−0.0193 (0.0087)*
Share young origin	0.0866 (0.0459)	0.0875 (0.0459)	0.1164 (0.0633)	0.1412 (0.0628)*
Partic. rate women origin	0.0150 (0.0062)*	0.0136 (0.0062)*	0.0121 (0.0059)*	0.0144 (0.0063)*
Distance (1000 km)	−0.0556 (0.0142)**	−0.0523 (0.0141)**	−0.0502 (0.0146)**	−0.0538 (0.0145)**
Border sharing	0.5628 (0.2304)*	0.5625 (0.2300)*	0.4880 (0.2489)*	0.4952 (0.2470)*
Open borders	0.1265 (0.0426)**	0.1287 (0.0426)**	0.0447 (0.0311)	0.0673 (0.0340)*
Linguistic distance	−1.6350 (0.2874)**	−1.7170 (0.2839)**	−1.7963 (0.3261)**	−1.9243 (0.3193)**
Religious distance	−0.4437 (0.2903)	−0.5940 (0.2753)*	−1.3086 (0.3270)**	−1.4255 (0.3204)**
Cultural distance (Hofstede)	−0.6188 (0.3664)			
Cultural distance (Inglehart)			−0.2482 (0.1058)*	
Constant	4.4211 (1.4692)**	4.3726 (1.4670)**	4.7474 (1.9417)*	3.7669 (1.9345)
Observations	1997	1997	1846	1846
Number of panels	246	246	219	219

Standard errors in parentheses
*$p = 0.05$; **$p = 0.01$

a common official language is important, but the distance between different languages matters as well.

In Table 2, we introduce the other measures of cultural distances. Because we miss information for several countries for these variables, the size of our sample drops significantly. For this reason, we introduce each of the culture variables separately, and report regression results excluding these variables but involving the same sample (columns (2) and (4) show results based on the same samples as the previous column and excluding the relevant culture variable). In all cases, the coefficients for the other variables seem to be affected by the sample size rather than by the specific culture variables. In column (1), we look at the effect of cultural distance based on the Hofstede dimensions. This variable has a negative but not significant effect. Column (3) presents the results when including the cultural distance variable based on Inglehart and Baker factor scores. This variable seems to have more explanatory power than the Hofstede variable: Cultural distance, based on measures of norms and values, matters next to the other measures of cultural distance (linguistic and religious distance variables).

To get some idea of the magnitude of these effects and the importance for migration flows in our sample, we calculated the effects of a one standard deviation increase of the different cultural variables and compared the results with the effects for GDP per capita and unemployment. Our regression results imply that an increase in linguistic distance with one standard deviation lowers the migration flow with 56% (calculated using the estimates in Table 2). This effect is about 50% higher than the effect of raising GDP per capita in the destination country with one standard deviation and much more than a change of one standard deviation in unemployment rates. To illustrate further, take two countries which share a common language, such as for example Belgium and France and compare them with two countries which do not have a common language, such as Spain and France (our linguistic distance indicator is equal to 0.266 for this pair of countries). All else equal, we calculated that the difference between the French and Spanish GDP per capita must be more than 40% larger than the difference between the French and Belgian GDP to generate the same immigration flow. Hence, if the French and Belgian GDP per capita were equal, the Spanish GDP per capita should be 40% lower than the French GDP per capita to generate the same emigration flows, solely because France and Spain speak a different language.

In a similar vein, we could calculate the estimated effects of a change in the other measures of cultural distance. An increase of 1% of the religious distance between countries, evaluated at the mean, lowers migration with 0.9%. This implies that an increase of the religious distance between countries with one standard deviation lowers migration with 35%. Greece is on average the religiously most distant country in our sample. Compared with the average religious distance migration flows involving Greece are about 40% lower.

Similarly, an increase of 1% in the Inglehart measure of cultural distance corresponds with a fall of migration of 0.33% (and one standard deviation increase implies a 17% lower migration flow). For the developed countries

Table 3 Determinants of migration flows—the role of networks

Dependent variable	(1) Inflow	(2) Inflow	(3) Inflow	(4) Inflow	(5) Inflow	(6) Inflow
Lagged GDP/cap dest.	0.1037 (0.0182)**	0.0911 (0.0170)**	0.1427 (0.0258)**	0.1345 (0.0251)**	0.1602 (0.0277)**	0.1460 (0.0288)**
Lagged GDP/cap origin	−0.0812 (0.0145)**	−0.0727 (0.0137)**	−0.0865 (0.0221)**	−0.0905 (0.0214)**	−0.0844 (0.0244)**	−0.0793 (0.0252)**
Lagged u. rate dest.	−0.0276 (0.0086)**	−0.0338 (0.0079)**	−0.0216 (0.0103)*	−0.0284 (0.0100)**	−0.0033 (0.0109)	−0.0116 (0.0114)
Lagged u. rate origin	−0.0001 (0.0078)	−0.0014 (0.0072)	−0.0003 (0.0092)	−0.0027 (0.0089)	−0.0011 (0.0112)	−0.0020 (0.0117)
Population dest.	−0.0117 (0.0073)	−0.0078 (0.0068)	−0.0174 (0.0081)*	−0.0135 (0.0079)	−0.0155 (0.0082)	−0.0106 (0.0085)
Population origin	0.0083 (0.0013)**	0.0092 (0.0013)**	0.0082 (0.0014)**	0.0088 (0.0014)**	0.0090 (0.0015)**	0.0101 (0.0014)**
Share tertiary educated origin	0.0097 (0.0102)	0.0000 (0.0102)	0.0079 (0.0119)	0.0019 (0.0119)	−0.0027 (0.0117)	−0.0121 (0.0117)
Share young origin	0.0927 (0.0533)	0.0871 (0.0536)	0.0878 (0.0550)	0.0858 (0.0551)	0.1267 (0.0755)	0.1180 (0.0753)
Part. rate women origin	0.0121 (0.0062)	0.0089 (0.0060)	0.0050 (0.0075)	0.0050 (0.0074)	0.0031 (0.0085)	−0.0018 (0.0087)
Distance (1000 km)	−0.0592 (0.0191)**	−0.0634 (0.0191)**	−0.0612 (0.0194)**	−0.0677 (0.0193)**	−0.0550 (0.0199)**	−0.0566 (0.0198)**
Border sharing	0.4628 (0.2386)	0.3661 (0.2402)	0.3705 (0.2672)	0.2997 (0.2680)	0.4044 (0.2768)	0.3175 (0.2756)
Open borders	0.1421 (0.0398)**	0.1499 (0.0367)**	0.1906 (0.0485)**	0.2024 (0.0469)**	0.1137 (0.0525)*	0.1205 (0.0548)*
Linguistic distance	−1.1420 (0.3312)**	−1.6562 (0.3289)**	−1.5322 (0.3745)**	−1.9956 (0.3702)**	−1.6786 (0.4214)**	−2.1735 (0.4150)**
Religious distance	−0.6135 (0.3468)	−0.4828 (0.3484)	−0.5192 (0.3947)	−0.2634 (0.3939)	−1.0169 (0.4620)*	−0.8980 (0.4604)
Cultural distance (Hofstede)			−0.2216 (0.4585)	−0.5680 (0.4574)		
Cultural distance (Inglehart)					−0.1144 (0.1290)	−0.2151 (0.1280)
Share of population of origin in dest.	0.0019 (0.0004)**		0.0018 (0.0005)**		0.0018 (0.0004)**	
Constant	3.5264 (1.6547)*	4.9414 (1.6564)**	3.7698 (1.7344)*	4.8734 (1.7335)**	2.9586 (2.3551)	4.6729 (2.3602)*
Observations	1504	1504	1302	1302	1111	1111
Number of panels	196	196	170	170	145	145

Standard errors in parentheses
*$p = 0.05$; **$p = 0.01$

in our sample, a one standard deviation change in the different measures of cultural distance therefore has a stronger effect than a one standard deviation change in the economic variables.

5.3.3 Network effects

The results so far suggest that culture matters a lot in explaining migration flows. However, we need to be careful with interpreting these results. It could be, for example, that culture has mattered in the past and has led to the implantation of communities which were culturally close in other countries. It could be the presence of these communities that is now the main driving factor of migration flows rather than cultural proximity. To investigate the possible role of network effects, we introduce a variable measuring the size of local ethnic groups (i.e., the size of the population of the same nationality as the immigrants in the country of destination) and investigate how it influences our estimates.

The results are reported in Table 3. We find a positive and significant effect of this variable. The magnitude of the other cultural variables seems somewhat reduced. However, again, part of the reason why the effects of cultural variables change seems to be because of the smaller sample size. Nevertheless, we observe a reduction in the magnitude of all coefficients of cultural proximity;

Table 4 Determinants of networks (averages over the period 1990–2003

	Log (share of population of origin in destination)		
Dependent variable	(1)	(2)	(3)
GDP/capita destination	0.0005 (0.0001)**	0.0000 (0.0001)	0.0004 (0.0001)**
GDP/capita origin	−0.0002 (0.0000)**	−0.0001 (0.0001)*	−0.0001 (0.0000)**
Unempl. rate destination	0.0996 (0.0469)*	1.0735 (0.0951)**	0.0939 (0.0487)
Unempl. rate origin	0.0380 (0.0335)	0.0958 (0.0546)	0.0094 (0.0335)
Population destination	−0.0105 (0.0029)**	0.0077 (0.0027)**	−0.0086 (0.0027)**
Population origin	0.0128 (0.0014)**	0.0117 (0.0017)**	0.0105 (0.0014)**
Share tertiary educated origin	−0.0072 (0.0127)	−0.0454 (0.0135)**	0.0029 (0.0131)
Share young origin	0.1247 (0.0560)*	0.1093 (0.0741)	0.0690 (0.0539)
Part. rate women origin	0.0224 (0.0105)*	0.0139 (0.0126)	0.0142 (0.0121)
Distance (1000 km)	−0.1356 (0.0304)**	−0.1400 (0.0297)**	−0.1137 (0.0289)**
Border sharing	0.6195 (0.2464)*	0.7169 (0.2502)**	0.6771 (0.2563)**
Open borders	−0.4249 (0.5591)	−0.6643 (0.5517)	0.0020 (0.5309)
Common language	0.4532 (0.2590)	0.1122 (0.2525)	0.3527 (0.2619)
Linguistic distance	−1.0612 (0.3861)**	−1.8128 (0.4541)**	−1.0675 (0.3959)**
Religious distance	−0.6884 (0.3991)	−1.1195 (0.4450)*	−0.4003 (0.4185)
Cultural distance (Ingleheart)		−0.0565 (0.1149)	
Cultural distance (Hofstede)			−0.0329 (0.4190)
Constant	−5.8736 (2.5201)*	−3.7468 (3.3235)	−4.1318 (2.4811)
Observations	244	166	213
R squared	0.80	0.85	0.80

Standard errors in parentheses
*$p = 0.05$; **$p = 0.01$

and the Inglehart cultural distance coefficient is no longer significant. The effect of linguistic distance, however, remains large and significant. As an additional exercise, we investigate the relationship between the sizes of these local communities (networks) and all our independent variables, in the long run, averaging these over the period 1990–2003. The results are reported in Table 4. The correlation between these cultural distance variables and the size of networks is, except for linguistic distance, low or insignificant. Thus, we fail to find strong evidence that the sizes of existing communities are correlated with cultural distance. These results suggest that migration in the past may not have been driven as much by cultural proximity than recent migration. This is in line with the historical description of migration patterns between European countries for example which in the past seem to have been driven by economic factors.

5.3.4 Migration between countries with open borders

The next question we ask is whether cultural distance matters as much (or maybe more) between countries which have open borders. One motivation

Table 5 Determinants of migration flows between countries with open borders

	(1) Inflow	(2) Inflow	(3) Inflow	(4) Inflow
Lagged GDP/cap dest.	0.0502 (0.0149)**	0.1833 (0.0537)**	0.1892 (0.0636)**	0.1919 (0.0553)**
Lagged GDP/cap origin	−0.0795 (0.0141)**	−0.1712 (0.0345)**	−0.1400 (0.0454)**	−0.1437 (0.0425)**
Lagged unempl. rate dest.	−0.0396 (0.0085)**	−0.0188 (0.0208)	−0.0231 (0.0290)	−0.0240 (0.0240)
Lagged unempl. rate origin	−0.0220 (0.0087)*	−0.0320 (0.0171)	−0.0137 (0.0211)	−0.0143 (0.0187)
Population dest.	−0.2205 (0.0885)*	−0.3742 (0.1787)*	−0.3872 (0.2619)	−0.3503 (0.2224)
Population origin	0.0266 (0.0029)**	0.0240 (0.0032)**	0.0187 (0.0039)**	0.0188 (0.0040)**
Share tertiary educated origin	0.0336 (0.0135)*	0.0505 (0.0169)**	0.0430 (0.0251)	0.0431 (0.0259)
Share young origin	0.0525 (0.0523)	0.0493 (0.0529)	0.0543 (0.0665)	0.0534 (0.0684)
Partic. rate women origin	0.0266 (0.0061)**	0.0371 (0.0093)**	0.0303 (0.0126)*	0.0273 (0.0122)*
Distance (1000 km)	−0.1409 (0.1756)	−0.4466 (0.1961)*	−0.3400 (0.2639)	−0.3687 (0.2699)
Border sharing	0.3505 (0.2498)	0.2850 (0.2817)	0.2626 (0.3412)	0.2895 (0.3537)
Common language	0.8328 (0.3663)*	0.9062 (0.3537)*	0.9653 (0.4636)*	1.0543 (0.4763)*
Linguistic distance	0.3176 (0.3646)	0.1189 (0.4314)	0.0773 (0.5808)	0.1129 (0.5994)
Religious distance	−1.9191 (0.2616)**	−1.3169 (0.3178)**	−1.2007 (0.4417)**	−1.2038 (0.4566)**
Cultural distance (Hofstede)		0.2223 (0.4238)	0.4035 (0.5587)	0.3827 (0.5789)
Share of population of origin in dest.			0.0028 (0.0015)	
Constant	11.5299 (2.5919)**	12.2582 (3.6878)**	11.0902 (5.2398)*	11.5779 (4.7303)*
Observations	1411	967	608	608
Number of panels	202	147	96	96

Standard errors in parentheses
*$p = 0.05$; **$p = 0.01$

for the analysis of cultural barriers in this paper is the "European immobility puzzle", arguing that cultural distance may explain why mobility is relatively low in Europe, despite the principle of free mobility. In Table 5, we present results including country pairs which have open borders only, that is, that are members of the European Union or European Economic Area. We find that culture does matter, but only the coefficients of "common language" and "religious distance" remain large and significant. On the other hand, the effects of linguistic distance and cultural distance (Hofstede) are not significant. We also fail to find evidence for network effects within the European Economic Area. The other variables seem to matter less as well, for example, sharing a border or the physical distance between capital cities have a lower effect than when we consider the larger sample of countries. Maybe contrary to our expectation, the economic variables do play a large and significant role, the estimated coefficients of the GDP/capita variables for example are larger than for the larger sample. Thus, we conclude that mobility between European countries does respond both to cultural and economic differentials.

5.3.5 Extensions

We extend our empirical analysis by including information about the generosity of the unemployment benefit systems in the country of origin and of destination. We use data on the gross replacement rates, provided by the OECD. Since we did not have information for all countries, our sample was substantially reduced when including these two variables. The results are shown in Table 6. We find here that the gross replacement rate in the country of origin has a significant negative effect on emigration. This confirms the hypothesis that unemployed workers may not be willing to move because they have a sufficient income in their home country. Note that the unemployment rate variables are not significant anymore, but this is due to the reduction in sample. The second column shows the results based on the same small sample, excluding the replacement rates variables. The coefficients for all other variables remain essentially the same.

The second extended specification we investigate is one controlling for the level of social expenditures in the country of origin and destination. This is to test whether migration flows are driven by differences in social benefits. Migrants could be attracted by countries with generous social security systems, a phenomenon that is usually referred to as the "welfare magnet effect" (see Borjas 1999b). The number of observations drops to a large extent. The estimates we find are exactly opposite to what one would expect (note that including the gross replacement rate or not does not affect the results). There is less immigration in countries with more generous social security systems than in others. The welfare magnet hypothesis does not find support here, when considering developed countries only. Preferably, we would differentiate by skill level, as generous welfare states are presumably particularly attractive for low-skilled migrants, but our data do not allow a more elaborate strategy.

Table 6 Determinants of migration flows—the role of social security

	(1) Inflow	(2) Inflow	(3) Inflow	(4) Inflow
Lagged GDP/cap dest.	0.1583 (0.0211)**	0.1536 (0.0217)**	0.0430 (0.0179)*	0.0495 (0.0170)**
Lagged GDP/cap origin	−0.0682 (0.0173)**	−0.0741 (0.0176)**	−0.0952 (0.0146)**	−0.0946 (0.0145)**
Lagged unempl. rate dest.	−0.0014 (0.0076)	−0.0034 (0.0078)	−0.0176 (0.0089)*	−0.0223 (0.0085)**
Lagged unempl. rate origin	−0.0082 (0.0065)	−0.0083 (0.0067)	0.0036 (0.0081)	−0.0021 (0.0076)
Population dest.	−0.0364 (0.0061)**	−0.0369 (0.0063)**	−0.0235 (0.0073)**	−0.0266 (0.0070)**
Population origin	0.0086 (0.0012)**	0.0101 (0.0011)**	0.0113 (0.0011)**	0.0114 (0.0011)**
Share tertiary educated origin	−0.0016 (0.0083)	−0.0043 (0.0083)	−0.0118 (0.0077)	−0.0108 (0.0077)
Share young origin	0.1085 (0.0477)*	0.0894 (0.0475)	0.1214 (0.0453)**	0.1193 (0.0454)**
Partic. rate women origin	0.0080 (0.0057)	0.0092 (0.0058)	0.0283 (0.0056)**	0.0271 (0.0054)**
Distance (1000 km)	−0.0599 (0.0152)**	−0.0535 (0.0150)**	−0.0567 (0.0146)**	−0.0530 (0.0143)**
Border sharing	0.4889 (0.2290)*	0.4912 (0.2283)*	0.5194 (0.2074)*	0.5002 (0.2076)*
Open borders	0.1149 (0.0321)**	0.1153 (0.0334)**	0.1058 (0.0377)**	0.0997 (0.0368)**
Linguistic distance	−1.6766 (0.2849)**	−1.6947 (0.2842)**	−1.5237 (0.2577)**	−1.4867 (0.2578)**
Religious distance	−0.8265 (0.2782)**	−0.7053 (0.2771)*	−0.9633 (0.2426)**	−0.9744 (0.2428)**
Gross repl. rate dest.	−0.0028 (0.0037)			
Gross repl. rate origin	−0.0127 (0.0032)**			
Social expenditures as percentage destination (%)			−0.0199 (0.0095)*	
Social expenditures as percentage origin (%)			−0.0101 (0.0074)	
Constant	4.0237 (1.5010)**	4.1607 (1.5018)**	5.1542 (1.4567)**	4.6725 (1.4364)**
Observations	1699	1699	1797	1797
Number of panels	260	260	289	289

Standard errors in parentheses
*$p = 0.05$; **$p = 0.01$

5.3.6 Robustness analysis

The last exercise we do is a robustness check of our results across econometric specifications. As we mentioned in the introduction, the econometric specification is potentially a matter of concern in this type of analyses, given the nature of the data (in particular the large presence of zeros in the dependent variable) and the presence of heteroskedasticity in the error term.

It is useful to compare our baseline specification (including cultural variables), to alternative econometric specifications. We present results based on a population-average model with the dependent variable being the logarithm of the inflow, and a specification based on the inflow rate. In essence, the results are quite similar across these specifications. But the effects tend to be more precisely estimated with the negative binomial specification. There are a number of differences worth documenting. First, the coefficients of the lagged

Table 7 Determinants of migration flows—robustness across econometric specifications

	(1) Inflow	(2) Lninflow	(3) Lninflowrate
Lagged GDP/cap dest.	0.1238 (0.0228)**	0.0974 (0.0297)**	0.0653 (0.0267)*
Lagged GDP/cap origin	−0.0877 (0.0185)**	−0.0930 (0.0212)**	−0.1035 (0.0196)**
Lagged unempl. rate dest.	−0.0211 (0.0085)*	−0.0236 (0.0123)	−0.0266 (0.0117)*
Lagged unempl. rate origin	−0.0046 (0.0079)	0.0004 (0.0109)	−0.0132 (0.0106)
Population dest.	−0.0148 (0.0067)*	−0.0137 (0.0088)	
Population origin	0.0102 (0.0011)**	0.0104 (0.0010)**	
Share tertiary educated origin	−0.0065 (0.0088)	0.0035 (0.0083)	−0.0113 (0.0084)
Share young origin	0.0866 (0.0459)	0.0942 (0.0420)*	0.1527 (0.0424)**
Partic. rate women origin	0.0150 (0.0062)*	0.0176 (0.0069)*	0.0302 (0.0069)**
Distance (1000 km)	−0.0556 (0.0142)**	−0.0635 (0.0133)**	−0.0576 (0.0136)**
Border sharing	0.5628 (0.2304)*	0.6265 (0.2078)**	0.2317 (0.2140)
Linguistic distance	−1.6350 (0.2874)**	−1.2905 (0.2661)**	−1.5728 (0.2702)**
Religious distance	−0.4437 (0.2903)	−0.5548 (0.2663)*	−0.9934 (0.2747)**
Cultural distance (Hofstede)	−0.6188 (0.3664)	−0.6044 (0.3321)	−0.1016 (0.3429)
Open borders	0.1265 (0.0426)**	0.0782 (0.0625)	0.1029 (0.0609)
Constant	4.4211 (1.4692)**	3.6531 (1.4356)*	−6.8383 (1.4422)**
Observations	1997	1965	1965
Number of panels	246	242	242

Standard errors in parentheses
*$p = 0.05$; **$p = 0.01$

unemployment rate (both for the country of origin and destination) are not significant in the specification based on the log of the inflow, while they are in the baseline specification, Second, there is a difference for the estimated coefficient of the effect of open borders, which is significant in our baseline specification and is less precisely estimated in the two other specifications.

Thus, overall, in this particular case, it seems that our econometric specification does not make a crucial difference in identifying the effects we have found, in particular the effects of cultural variables. But of course, we cannot claim that this conclusion will hold in all contexts (Table 7).

6 Conclusions

This paper provides unique evidence for the role of cultural barriers in migration between developed countries. We propose a series of new indicators measuring these barriers in a more precise way than has ever been done in the literature. More specifically, we introduce more refined measures of the cultural distance between countries, correcting for the proximity between languages and religions.

We present an empirical analysis of migration flows between 22 OECD countries over the period 1990–2003. We find strong evidence of the importance of cultural links between countries, going well beyond the simple sharing of a common language. Migration flows between countries with closely related

languages are likely to be much larger than between countries with unrelated languages. Similarly, the proximity in religions and culture also stimulates migration.

As culture is hard to change, this implies that migration flows between culturally very distinct developed countries will remain low for many years to come. Well-designed policies can take away many obstacles for migration, like administrative procedures, but this can hardly compensate for cultural differences. Europe probably has to accept that it is very difficult to substantially increase labor mobility between a number of countries. Judging from the importance of linguistic and cultural distances, policies to raise labor mobility could be targeted at reducing cultural distance, for example by encouraging foreign language learning, in particular since what seems to matter for mobility between European countries is sharing a common language.

Adjusting to a different culture is just one example of the costs people face when migrating to another country. Migration costs include a wide range of obstacles to movement, from the obvious costs of physically moving to the costs of settling in a new country. Even if in principle workers are free to move, like in the EEA, they are in practice confronted with a series of obstacles hampering their movement, such as the lack of recognition of foreign qualifications, the lack of transferability of pension rights, etc. It may be valuable to investigate in further research the role of these other—institutional—obstacles.

In conclusion, our results show that cultural barriers play a crucial role in migration, and do a better job in explaining migration patterns between developed countries than differentials in economic variables. This paper shows that migration costs have a large potential in explaining migration patterns, leaving the door open to more research on identifying more of these costs.

Acknowledgements The authors would like to thank the editor Klaus Zimmermann, as well as three anonymous referees for their constructive comments. They also thank Roel Jennissen, Rob Euwals, Henri de Groot, Pierre Koning, Guido van Steen, Jeroen van de Ven, seminar participants at the University of Essex and at the CPB Netherlands Bureau for Economic Policy Analysis for useful comments on previous versions of the paper. The authors are also very grateful to Twan Verschaeren for his support in the data collection.

References

Antolin P, Bover O (1997) Regional migration in Spain: the effect of personal characteristics and of unemployment, wage and house price differentials using pooled cross-sections. Oxf Bull Econ Stat 59(2):215–235

Barro R, Sala-i-Martin X (1995) Economic growth. McGraw-Hill, London

Borjas GJ (1999a) The economic analysis of immigration. In: Ashenfelter OC, Card D (eds) Handbook of labor economics, 1st edn, vol 3, chap. 28. Elsevier, pp 1697–1760

Borjas GJ (1999b) Immigration and welfare magnets. J Labor Econ 17(4):607–637

Braunerhjelm P, Faini F, Norman V, Ruane F, Seabright P (2000) Integration and the regions of Europe: how the right policies can prevent polarization, Monitoring European Integration 10. CEPR, London

Cameron A, Trivedi P (1986) Econometric models based on count data: comparisons and applications of some estimators and tests. J Appl Econ 1(1):29–53
Carrington WJ, Detragiache E, Vishwanath T (1996) Migration with endogenous moving costs. Am Econ Rev 86(4):909–930
Chorny V, Euwals R, Folmer K (2007) Immigration policy and welfare state design; a qualitative approach to explore the interaction. CPB document 153. Netherlands Bureau for Economic Policy Analysis, The Hague
Clark X, Hatton TJ, Williamson JG (2004) What explains emigration out of Latin America? World Dev 32(11):1871–1890
Decressin J, Fatas A (1995) Regional labor market dynamics in Europe. Eur Econ Rev 39:1627–1655
Dyen I, Kruskal JB, Black P (1992) An Indoeuropean classification: a lexicostatistical experiment. Trans Am Philos Soc 82(5):1–132
Ederveen S, Bardsley N (2004) The influence of wage and unemployment differentials on labour mobility in the EU: a meta-analysis, ACCESSLAB Research Report, Workpackage 3, January 2004
Ederveen S, Nahuis R, Parikh A (2007) Labour mobility and regional disparities: the role of female labour participation. J Popul Econ 20(4):895–913
Edin P-A, Fredriksson P, Åslund O (2003) Ethnic enclaves and the economic success of immigrants: evidence from a natural experiment. Q J Econ 118(1):329–357
Faini R, Venturini A (1994) Migration and growth: the experience of Southern Europe. CEPR Discussion Papers 964
Fertig M, Schmidt CM (2002) Mobility within Europe, the attitudes of European youngsters. RWI Discussion Papers 1
Ghatak S, Levine P (1996) Migration theories and evidence: an assessment. J Econ Surv 10(2):159–98
Gianetti M (2001) Migration and skill complementarities. Labour 15(1):1–31
Gros D (1996) A reconsideration of the optimum currency area approach: the role of external shocks and labour mobility. Natl Inst Econ Rev 158:108–117
Gross D, Schmitt N (2003) The role of cultural clustering in attracting new immigrants. J Reg Sci 43(2):295–318
Harris J, Todaro M (1970) Migration, unemployment & development: a two-sector analysis. Am Econ Rev 60(1):126–42
Hatton T, Williamson JG (2005) What fundamentals drive world migration? In: Borjas GJ, Crisp J (eds) Poverty International Migration and Asylum, Palgrave-Macmillan
Hatton TJ, Williamson JG (2006) Refugees, asylum seekers and policy in Europe. In: Foders F, Langhammer RF (eds) Labour mobility and the world economy. Kiel, Springer
Hofstede G (1991) Cultures and organizations: software of the mind. McGraw-Hill, London
Hofstede G (2001) Culture's consequences: comparing values, behaviors, institutions, and organizations across nations, 2nd edn. Sage, London
Inglehart R, Baker WE (2000) Modernization, cultural change, and the persistence of traditional values. Am Sociol Rev 65(1):19–51
Kogut B, Singh H (1988) The effect of national culture on the choice of entry mode. J Int Bus Stud 22(2):209–24
Mauro P, Spilimbergo A (1999) How do the skilled and the unskilled respond to regional shocks? The case of Spain, IMF Staff Papers 46 (1)
Mayda AM (2010) International migration: a panel data analysis of the determinants of bilateral flows. J Popul Econ 23(4):1249–1274
McKenzie D, Rapoport H (2007) Self-selection patterns in Mexico–U.S. migration: the role of migration networks, Policy Research Working Paper Series 4118, The World Bank
Munshi K (2003) Networks in the modern economy: Mexican migrants in the U.S. labor market. Q J Econ 118(2):549–597
Palloni A, Massey D, Ceballos M, Espinosa K, Spittel M (2001) Social capital and international migration: a test using information on family networks. Am J Sociol 106(5):1262–98
Pedersen P, Pytlikova M, Smith N (2008) Selection and network effects—migration flows into OECD countries, 1990–2000. Eur Econ Rev 52(6):1160–1186
Santos Silva JMC, Tenreyro S (2006) The log of gravity. Rev Econ Stat 88(4):641–658

Sjastaad L (1962) The costs and returns of human migration. J Polit Econ 70:80–93
Tassinopoulos A, Werner H (1999) To move or not to move—migration of labour in the European Union, IAB Labour Market Research Topics 35
Van Wissen L, Visser H (1998) An explanatory analysis of international migration flows within the European Economic Area. Chapter 14. In: Gorter C, Nijkamp J, Poot J (eds) Crossing borders: regional and urban perspectives on international migration. Ashgate, Aldershot
Vinken H, Soeters J, Ester P (2004) Comparing cultures. Dimensions of culture in a comparative perspective. Brill, Leiden
Wildasin D (2000) Labor market integration, investment in risky human capital and fiscal competition. Am Econ Rev 90(1):73–95
Wooldridge JM (2002) Econometric analysis of cross-section and panel data. MIT Press, Cambridge
Young G (1975) The choice of dependent variable for cross-section studies of migration. Can J Econ 8(1):93–100