NOTA DI LAVORO

Which Factors Explain the Rising Ethnic Heterogeneity in Italy? An Empirical Analysis at Province Level

By **Cristina Cattaneo**, Fondazione Eni Enrico Mattei (FEEM)

Which Factors Explain the Rising Ethnic Heterogeneity in Italy? An Empirical Analysis at Province Level

By Cristina Cattaneo, Fondazione Eni Enrico Mattei (FEEM)

Summary

The paper investigates the determinants of ethnic heterogeneity of the Italian provinces. Among other factors, the paper tests empirically whether gradual improvements in distant communication boost the generation of ethnically heterogeneous provinces. Consequently to easier communication, movers increasingly rely on an enlarged community for identity transmission, rather than on localized peer effects of the ethnic enclaves. The empirical estimation provides support to this hypothesis. Improvements in internet communications are found to increase the ethnic diversity of the Italian provinces.

Address for correspondence:

Cristina Cattaneo
Fondazione Eni Enrico Mattei
Corso Magenta 63
20123 Milano
Italy
Phone: +39 0252036983
E-mail: cristina.cattaneo@feem.it

Which factors explain the rising ethnic heterogeneity in Italy? An empirical analysis at province level

Cristina Cattaneo*

Fondazione Eni Enrico Mattei (FEEM)

Abstract

The paper investigates the determinants of ethnic heterogeneity of the Italian provinces. Among other factors, the paper tests empirically whether gradual improvements in distant communication boost the generation of ethnically heterogeneous provinces. Consequently to easier communication, movers increasingly rely on an enlarged community for identity transmission, rather than on localized peer effects of the ethnic enclaves. The empirical estimation provides support to this hypothesis. Improvements in internet communications are found to increase the ethnic diversity of the Italian provinces.

1. Introduction

The upward surge of immigration has introduced a growing interest in the economic implications of ethnic heterogeneity. The intensification of the emigration flows from less-developed to economic developed countries is not questionable and the current trend is likely to continue in the future. The enlargement of the community of foreigners is a characteristic of our cities, and will bring important economic implications.

There is a large and growing body of empirical literature that studies the economic benefits of cultural diversity in destination countries, where cultural diversity is proxied by ethnic or linguistic diversity. Given that ethnicity and linguistic groups are carriers of cultural identity and, therefore, ethnicity is an important component of diversity, the economics literature often treats the two as synonymous. Moreover, each ethnic group owns specific skills and abilities, which could serve as valuable factors in the overall production. While in cross-countries analysis the relationship between cultural diversity and economic performance is found to be quite mixed (EASTERLY and LEVINE, 1997; COLLIER, 2001; EASTERLY, 2001; ALESINA and LA FERRARA, 2005) in finer spatial units, such as cities and provinces, a positive link is reported (GLAESER et al., 1992; OTTAVIANO and PERI, 2005; 2006; BELLINI et al., 2011; SPARBER, 2010). In a dense environment, in fact, where the interactions among diverse people are more likely to happen, the potential benefits related to diversity are larger.

Individuals belonging to different cultures have different ways of addressing the same problem, possess complementary pieces of information, and by means of informal communication, available in a dense environment such as cities, reach better and quicker solutions. JACOBS (1969) emphasizes the powerful link between industrial diversity and innovation. The author believes that important knowledge transfers arise from outside the core industry. Areas with highly diversified industries as opposed to geographically concentrated industries should display greater growth. ASHRAF and GALOR (2007) place the fortune of Europe in its heterogeneity, developed in many years of foreign people invasion. This cross-fertilization could have been responsible for the shift

from an agriculture based regime to a production regime characterised by new manufacturing technologies. This may have contributed to the "reverse of fortunes" between Europe and China. The latter was historically richer than Europe, but its supremacy has been challenged by its high degree of cultural homogeneity.

Nonetheless the established positive consequences, the factors that drive ethnic diversity at the local level have received limited inspection. The investigation of the determinants of ethnic heterogeneity is the scope of the present paper. In the first part of the paper, the rising trend in ethnic heterogeneity at the province level is documented. In the second part, it is assessed which factors explain such sharp increase. Standard economic and socio-demographic pull factors are introduced in the estimated specification, along with an additional determinant suggested by a formal model developed by OTTAVIANO and PRAROLO (2009), O-P hereafter. In this model, the ethnical composition and the variety of cities are influenced by communication and transportation improvements.

Cultural diversity fosters the exchange of ideas and knowledge, allowing steady growth to rise, and this produces a greater pay off to foreigners in terms of higher wages. At the same time however, the cultural identities of movers are likely to be preserved within the own ethnic enclave in the destination countries.[1] Migrants face the trade-off between an environment which maximizes their economic utility and an environment which maximizes their social utility. In other words, the migrants face the dilemma of living either in an ethnically heterogeneous setting or in an ethnic enclave. A solution to this dilemma can be reached by means of communication improvements. O-P formalize this hypothesis through a simple model with two mobile cultural groups and two open cities. They predict that two ethnic groups coexist within the same city when communication is easy, whereas they chose different cities when diaspora members find it hard to communicate at distance. Progress in transportation and communication helps offsetting the dilution of cultural identities, which occurs when one moves from an ethnic enclave to a more ethnically diverse environment. Improvements of this type reduce the social cost of relocating within the destination

country, as help migrants maintaining contacts with friends and family of their own ethnic group, while living in different cities. This in turn should increase the city heterogeneity, as individuals of different ethnicities chose to coexist in the same location in order to respond to the positive externalities generated by a culturally heterogeneous environment.

The remainder of the paper is organized as follows. Section 2 presents a brief review of the literature. Section 3 briefly sketches the O-P model and describes the methodological structure. In Section 4 a description of the data is presented. Section 5 presents the empirical exercise. Section 6 adds some robustness checks. Section 7 provides a summary and the conclusions.

2. Literature

This paper is related to two strands of literature, the first being the empirical studies that analyse the determinants of location choices of the foreign population. These studies combine both individual as well as location specific characteristics to explain settlement behaviour (BARTEL, 1989; ZAVODNY, 1997, 1998; JAEGER, 2000; BORJAS, 2001; BAUER et al. 2002, 2005; DAMM, 2009). Residential settlement by race influences the ethnic mix of cities, and therefore its ethnic heterogeneity. In particular, there are studies that provide a detailed analysis of the complete location pattern of foreigners. They investigate not only the settlement behaviour upon arrival but they also analyse subsequent internal mobility in destination countries. ASLUND (2005) is an example of this type. Exploiting a natural experiment, he highlights three types of determinant factors: the presence of earlier migrants from own country of birth, the size of the location in terms of population and labour market opportunities. ZORLU and MULDER (2008) document a substantial relocation of the foreign population after one year upon arrival. Moreover, they confirm ASLUND's findings regarding the importance of socio-economic factors and of the presence of co-ethnics in selecting the location of subsequent moves.

The second related literature is about the measurement of diversity. The economics literature in this field is still at an early stage, as only few diversity indices have been proposed and discussed on a theoretical background. The most widely used measure of diversity is the index of ethno-

linguistic fractionalization (ELF), first proposed in a statistical context under the name of Gini-Simpson index. The index is a decreasing transformation of the Herfindahl index of concentration and is simply a function of the shares of the different ethnic groups in the population. This simplicity however, represents both its advantage and disadvantage. On the one hand, it is easy to compute and interpret, but on the other, given the limited information included in the index, it seems inadequate for a wide range of applications. Population shares alone, for examples, are not able to capture preferences or complementarities among different ethnic types. To overcome this limitation, different measures have been proposed. BOSSERT et al. (2011) suggest the Generalized Index of Fractionalization (GELF), which measures the expected dissimilarity between two randomly drawn individuals. GREENBERG (1956) incorporates the degree of resemblance between different languages to produce a generalization of the ELF index. RAO (1982) produces a very similar indicator as the GREENBERG's (1956) index, which is called the quadratic entropy index. Overall, these indexes try to embody the effective distance of the groups in terms of a rich array of characteristics, such as socio-economic factors and their interactions, which are widely responsible for the effect of diversity on the economic outcome. DESMET et al. (2009) compare the performance of different diversity and polarization indexes, some incorporating language distance between ethnic groups and some without distance, in explaining cross countries income redistribution. On an empirical ground, the indexes that take into account the linguistic distance are found to be superior to the commonly used fractionalization index, which ignores it.

3. Methodology

The O-P model analyses the emergence of ethnically diversified cities and predicts that better communication drives foreigners to choose ethnically heterogeneous settings. On the one hand, foreigners face positive economic externalities generated by a culturally heterogeneous environment. On the other, an ethnically diversified environment generates social costs as people face the psychological strain of living with people having a different culture. Communication improvements lessen these costs as they offset the dilution of cultural identities, when foreigners

5

move from an ethnic enclave to a more ethnically diverse environment. The main empirical test of the paper is whether gradual improvements in distant communication and better transportations increase the ethnic diversity of local settings, as foreigners increasingly rely on an enlarged community for identity transmission, rather than on localized peer effects of an ethnic enclave.

In order to measure the ethnic heterogeneity, the index of ethno-linguistic fractionalization (ELF) is used. The index is computed as:

$$\text{DIV}_r = 1 - \sum_s (x_s)^2 \qquad (1)$$

where x_s is the share of foreigners of the specific origin group s computed for each local unit r. The index measures the probability that two individuals, randomly drawn from the city population, have different ethnicities. This index allows the measurement of not only the richness of a local unit in terms of ethnic groups, namely the number of groups that live in the city, but also of the evenness of the groups abundance, captured by the relative population shares.

More sophisticated indexes exist, which do not only consider the population shares but also the distance of the groups in terms of specific characteristics. The ethno-linguistic fractionalization index can be modified, by incorporating linguistic distance between the ethnic groups (GREENBERG, 1956). This extension can be very informative as it takes into account the degree of distinctiveness between different ethnic groups. This linguistic distance may influence the degree of social interaction between the different groups, which is finally responsible for the aforementioned productivity gains.

Given a matrix T that assigns a distance τ_{js} between the language spoken by the ethnic group j and s, the index is given by:

$$\text{Greenberg}_r = \sum_j \sum_s x_j x_s \tau_{js} \qquad (2)$$

where the x denotes the shares of the different ethnic groups living in local unit r. The matrix T is a standardized matrix, with $\tau_{jj} = 0$ and $\tau_{js} = \tau_{sj}$. Following DESMET et al. (2009) and FEARON (2003), the distance between language group j and s is computed according to:

$$\tau_{js} = 1 - (\frac{l}{m})^{\delta} \qquad (3)$$

where l is the number of shared branches between j and s, m is the maximum number of shared branches of the languages in the sample, and δ captures the degree to which the distance declines as the number of shared branches increases.[2] Information about linguistic trees is taken from the Ethnologue project. In agreement with DESMET et al. (2009), the parameter δ is settled to 0.05.

The empirical analysis is conducted using Italian data. Italy represents an interesting case for a variety of reasons. First, from being traditionally an example of labour exporting country, recently it became an important destination of the immigration flows. This implies that the analysis of migration issues has been under researched for Italy.

Second, Italy has many gateways of entry which are ethnic specific as they reflect the proximity with the origin countries for migrants. Large communities of Albanian are found in Puglia; immigrants from Balkans are concentrated in the North-East provinces; Liguria hosts large communities from Ecuador and Peru, eventually entering Italy after an initial period in Spain; many migrants from Tunisia and North Africa are found in Sicily; Chinese are largely settled in Florence. These gateways are likely to represent important places where the ethnic enclaves historically formed and whose distribution in the territory is not driven by economic conditions. This makes Italy an interesting case to evaluate how the settlement of migrants changed in time, following eventually economic opportunities.

An additional advantage can be quoted. The O-P model focuses on cities, as localized externalities, generated by people interactions, are the crucial factor favouring the coexistence of different cultures within the same city. However, to test the model with data from European countries, including Italy, the NUTS 3 level is the maximum level of disaggregation. This is because the regressors included in the estimated specification are not available at the city level but only at NUTS 3 level, which in Italy corresponds to provinces (see Table A1 for a description of the variables and their source). This limitation is lessened by some specific features of the Italian

demographic structure. First Italy hosts a large population in a limited geographical area, with an extremely high population density.[3] This makes the social interactions reasonably high even outside the urban area. Second, the capital cities of each province host about a third of the population of the entire province and the allocation of the foreigners between province and the main city reflects the allocation of the whole population. All these factors suggest that the use of the whole province rather than the capital city should not alter the empirical exercise. In fact, the chance of interactions, which justifies the use of the cities as territorial units, is preserved as well as the distribution of foreigners between cities and provinces.

4. Data Description

Data on the foreign population are drawn from the annual collection of information of the foreign population conducted by ISTAT.[4] The data are taken from the register offices of each Italian commune. The registers contain the number of resident foreigners at the beginning of the solar year classified by country of origin, along with the information of their movements, in terms of the new births, deaths, and relocation from different Italian towns or different countries.[5] These data allow a detailed disaggregation of foreigners in terms of country origin, as information for each specific country of origin is available. This dataset forms a balanced panel, which contains 103 provinces between 2002 and 2010.[6]

Italy has been historically a country of large emigration. From 1861 it produced nearly 30 million of emigrants. Only after 1970, the net balance of migration reversed, with the number of inflows exceeding the outflow. At present, Italy is an important host country, receiving an increasing number of migrants.

TABLE 1

The percentage of foreigners substantially increased between 2002 and 2010, being 2.6 in 2002 and raising to 7.2 in 2010 (Table 1). The geographical distribution of foreigners varied greatly and, among other factors, it was eventually influenced by the economic opportunities that the Italian

regions offered. In 2002 North-Central Italy hosted 3.4 percent of immigrants, whereas in the South lived less than one percent of foreigners.

TABLE 2

The increase in the number of foreigners produced an increase in diversity in Italy, as indicated by the rising trend in the Fractionalization index (Table 2).[7] Moreover, the index of Fractionalization is highly correlated with space, being higher in the Northern and Central Italian provinces, and lower in Southern provinces. A similar increasing trend and geographical distinction are documented by MOCETTI and PORELLO (2010) and by ACCETTURO et al. (2012), using data drawn from residence permits.

The increase in diversity is also confirmed by the Greenberg index, which incorporates the distance between the languages spoken by the different foreigners' groups (Table 3).

TABLE 3

In Figure 1 the index of fractionalization of the different provinces is grouped in quartiles. Both in 2002 and in 2010 all provinces in the South were disproportionately represented in the bottom two quartiles and the distribution. Some variation occurred between 2002 and 2010, although this tends to be limited to only one step change between the quartiles. For example there was no shift from the bottom to the third or fourth quartiles, and so on.

FIGURE 1

The analysis can be further conducted in terms of the area of origin of foreigners (Table 4). While it is documented a rising trend in immigration from all macro-areas of origin, the ethnic composition of the foreign population varied during the period considered. The contribution to the total of immigration from EU countries halved and became marginal compared to other origin areas. On the contrary, countries of Central Eastern Europe not only were the largest source of migrants in the entire period, but their contribution increased largely. In 2002 they represented 38 percent of the total stock, whereas in 2010 they picked to more than half of the total. In absolute term, the number of foreigners from this area more than quadrupled. Northern African countries

were the second largest origin region even if their contribution to the total shifted from being 24 percent in 2002, to 16 percent in 2010. Asia and Latin America were two other important sources of immigration for Italy, both in 2002 and in 2010.

TABLE 4

To test the O-P theoretical model, the key explanatory variable should be a measure of the transportation and telecommunication improvements occurred in the Italian provinces, as a proxy for mobility and connection costs. Connections of individuals living in different provinces became easier, both because the time to travel from one province to the other declined, and because new technologies made it simpler to communicate at distance. The wide penetration of computers and internet connections is responsible for the drastic improvements in the communication.

Different variables have been considered here. Regarding the transportation variables, the first option is the physical measure of the infrastructures, such as the kilometres of railroad and motorway.[8] Regarding the telecommunication variables, one measure is used here and it is the number of internet domains registered, available for 2002, 2004, 2007 and 2010. The increasing number of internet domains reflects the expansion in the number of people using broadband access technologies and therefore the growth in the use of internet. Table A1 provides a detailed description of the variables and their source.

TABLE 5

Surprisingly, between 2002 and 2010 the investments in infrastructures in railroad and motorway in Italy were minimal as far as the physical measure is considered. As reported in Table 5, the overall kilometres of railroad and motorway increased annually by only 0.6 and 0.4 percent, respectively. Conversely, the variable internet domain adequately captures the large improvements in telecommunication means, which occurred recently. Between 2002 and 2010, the number of registered internet domain increased annually by 44 percent.

5. Empirical Estimation

To understand the determinants of the level of diversity across provinces the index of fractionalization is estimated for 2002, 2004, 2007 and 2010.[9] The units of observation are the Italian provinces. Standard explanatory variables are integrated with the controls suggested by the O-P model. The following equation is therefore estimated:

$$DIV_{rt} = \alpha + \gamma' \textbf{DEMO}_{rt} + \beta \ pop_{rt} + \zeta \ act_{rt} + \phi \ growth_{rt} + \eta \ technology_{rt} + \lambda' \textbf{ORIGIN}_{rt} + \varepsilon_{rt} \quad (4)$$

where $\varepsilon_{rt} = \mu_r + \lambda_t + \nu_{rt}$ $\qquad\qquad\qquad\qquad\qquad\qquad$ r = 1, ….R; t = 1, …, T

The dependent variable is the ethno-linguistic fractionalization index for the province *r* in year *t*. The inputs of the index are the number of individuals from each country of origin. Given that the thirty largest source countries of immigration account for 90 percent of total foreign population, for simplicity, only these top countries have been selected, while the remaining were aggregated in the "other" category. The thirty most representative origin countries are, in descending order, Romania, Albania, Morocco, China, Ukraine, Philippines, Tunisia, Poland, Macedonia, India, Ecuador, Peru, Egypt, Moldova, Serbia and Montenegro, Senegal, Sri Lanka, Bangladesh, Pakistan, Nigeria, Germany, Ghana, Brazil, Bulgaria, France, Bosnia-Herzegovina, UK, Algeria, Russia, Croatia. In the robustness checks section, the index is computed for all single country of origin.

The vector of socio economic variables (DEMO) includes the percentage of the population in different age and education categories. These variables should proxy for discrimination factors. Locations, with younger and more educated inhabitants, are more likely to host a population who is more favourable to interactions with migrants. This characteristic represents an attractive factor to foreigners.

The total population (pop), which captures the size of the province, should also influence the location preference of migrants. Big provinces, for example, offer a larger variety of goods and services. They provide larger access to rental, including social, housing as well as to institutions for higher education.

The variables activity rate (act) and GDP growth (growth) proxy for economic factors such as job opportunities, which represent pull factors in the destination choice of migrants. Economic prospects are an important determinant of immigrants' location choice. In the empirical exercise, the unemployment rate has been introduced in lieu of the activity rate. The coefficient was statistically non-significant. For this reason only the activity rate has been employed. Both variables proxy for pull factors and none is preferable to the other.

Technology improvements, related to transportation and telecommunication, are added. Three variables have been selected in this analysis to capture how easily the foreign community could reach or communicate with the own-ethnic enclave, located in a different province. Ideal candidates would be measures of transportation and communication costs. However, these measures do not vary at province level and cannot be used in the estimation. Therefore, the selected variables proxy for variation in the speed and ease of transport and communication. The first two candidates are the actual levels of infrastructure in rail-line and motorway. Foreigners are unlikely to settle in provinces that are isolated and that are not well connected through a good system of transport-lines, as they will find it hard to physically reach their ethnic enclave. High levels of infrastructure of this kind can reduce the time to travel to reach the community. The third candidate is the number of internet domains registered in the different provinces. This is a proxy of the demand in communication infrastructure rather than the supply of the infrastructure. Unfortunately, no other measure of the effective infrastructure is available at such territorial disaggregation for the period of the analysis. However, this variable represents the quantifiable result of the rapid expansion in communication infrastructures, that occurred recently and that make it easier to communicate. Therefore, it should be highly correlated with measures of communication development.

Finally the ethnic group variables (ORIGIN) include the total number of foreigners from the different areas of origin. This is an attempt to control for ethnic specific preferences, which influence the importance of materialism and of parental as opposed to peer-driven cultural transmission in the O-P model. The paper finds that a multicultural environment is a stable solution

of the model when individuals place greater importance on the consumption of "materialistic" goods than on cultural goods, and when parental effects are more important than peer effects for cultural transmission. The disturbance contains the province fixed effect, the time effect and the conventional stochastic disturbance term (BALTAGI, 2005).

To take account for any within-group dependence, all models are estimated correcting standard errors for clustering (MOULTON, 1986). The use of variables in the specification that vary at regional level would suggest a clustering by region. However, CAMERON et al. (2008) report that when the number of clusters is limited, the t-statistics tend to be biased upward and therefore the tests over-reject. Given that Italy has 20 regions, to avoid the above problems, the standard errors are clustered by province.[10]

The results of the two-way error component model in 2002, 2004, 2007 and 2010 are reported in Table 6. In column (1), a parsimonious specification with only the standard covariates suggested by the literature is estimated. The table reports a tendency toward a rising diversity, experienced by the Italian provinces. The coefficients of the year dummies are positive and statistically significant. The average diversity index is two, five and eight percent higher in 2004, 2007 and 2010, respectively, than it was in 2002. The size of the province has a non-statistically significant effect on diversity. This finding contrasts the results of ASLUND (2005), ZORLU and MULDRE (2008) and DAMM (2009), who document a shift from locations with small population towards settings with large population.

TABLE 6

Economic opportunities exert a well determined effect on the location choice of foreigners. Provinces with a higher rate of activity and a higher GDP growth display, on average and ceteris paribus, a higher Fragmentation index. Labour market features represent important pull factors, as far as foreigners effectively respond to the advantages offered by the economically vibrant provinces, which offer greater job opportunities. This finding is in agreement with ASLUND (2005) and ZORLU and MULDRE (2008), who report that internal mobility of foreigners is driven by

13

economic prospects. Diversity is higher in provinces with a younger and more educated population. This may indicate that migrants chose to locate in provinces with a population who is more willing to interact with foreigners. Moreover, provinces which offer larger access to rental housing and to institutions for higher education may be attractive for both foreigners as well as for young and well educated natives.

Columns (2) to (5) add the variables suggested by the O-P model. In column (2) the communication variable is introduced. The communication variable exerts a positive and well determined effect on ethnic heterogeneity. The number of internet domain is positively correlated with the level of diversity, providing a robust support to the theoretical prediction of the model. Increasing the number of domains by ten units over 10'000 inhabitants, the fractionalization index augments by 0.1 percent. The second and third variables are the length of the railroad and motorway, expressed in kilometres. None of the variables exerts a significant effect on the level of diversity. In the last column all transportation variables enter jointly. The coefficient of the domain variable is robust to this specification, whereas the coefficients of the transportation variables remain statistically non-significant. The empirical results suggest that communication rather than transportation improvements increase the degree of ethnic heterogeneity of the provinces. Gradual improvements in distant communication, through services provided by internet connections, allow migrants to respond to positive externalities generated by a culturally heterogeneous environment. Communication development made it easy to maintain the ethnic identity of foreigners while living outside the ethnic enclave.

It should be noted that Italy lags behind the other European countries in terms of endowments of transport infrastructure and this gap has widened remarkably in the last 20 years (GOBBO, 2007). For example the endowment in infrastructure is 15 percent lower than in Germany and United Kingdom (VISCO, 2012). Large investments in motorway occurred till the '80, when Italy owned the most modern and developed road network in Europe (ANCE, 2009), but these investments nearly stopped hereafter, with drastic implications in terms of congestion. This feature

is confirmed by the low rate of growth in the motorway and railroad infrastructures, documented in Table 5. This feature could be one of the explanation for the null significant effect of the transportation coefficients. A second motivation could simply be related to the superiority of communication compared to transportation improvements. Cell phones and internet allow for inexpensive and easy instant communication, such that migrants may prefer to communicate at distance rather than commute in order to maintain contacts with the own ethnic group.

A criticism can be raised on the validity of the transportation and communication variables as they may not be exogenous. More economically advanced provinces attract a larger number of migrants and at the same time demand larger investments in transportation and telecommunication to support the business. This would result in an upward bias in the estimated coefficients of these variables. However, one should note that the specification controls for variables proxing for economic opportunities, and the coefficients of these variables resulted positive and statistically significant, as expected. Second, large scale investments, such as those in motorway and railroad, reflect national strategies, which may respond to the need to sustain depressed areas.

For example, a big part of the improvements in the two measures of transportation were conducted in the South of Italy. Sicilia, Calabria and Campania, for example, register one the largest growth in kilometres of rail-line and motorway. This occurred despite the South of Italy is less economically developed. The same is true regarding the internet domains registered. Provinces in the South of Italy report a smaller number of registrations compared to Centre and North, but this area catches up quite rapidly as the largest increase in internet domains occurred in the South. This seems to suggest that the economic unobservables governing the distribution of foreigners in the provinces should be uncorrelated with the economic unobservables influencing the transportation and communication improvements.

6. Robustness checks

Up to now, the Fragmentation index was computed selecting the most important source countries, which account for nearly 90 percent of the total migration stock. To test whether the analysis is

robust to the way the inputs of the Fragmentation index have been selected, all countries are now distinguished. The empirical results for the alternative aggregation are reported in Table 7. The standard covariates as well as the domain variable are robust to this alternative aggregation. An exception is the coefficient of the activity rate, which turned statistically non-significant.

Diversity indexes that incorporate some measures of distance, such as the linguistic distance, are becoming popular in the empirical literature, as these indexes add a further dimension to capture the degree of distinctiveness between different groups. For this reason, the equation (4) is estimated using the Greenberg index (1959), as described in (2), as dependent variable.[11] Small changes are displayed in the new estimations. The coefficient of the domain variable is still positive and statistically significant, albeit it decreases in magnitude (Table 7).

As an additional robustness check, different values for the parameter δ in (3) are used. So far, in agreement with DESMET et al. (2009), the parameter is settled to 0.05, but it would be interesting to assess if the previous result is robust to the use of alternative values. The parameter δ discounts the relative distance of languages if they pertain to completely different families or if they belong to the same family. DESMET et al. (2009) find that high values of δ are not reasonable, as they tend to produce higher increase in distance between pairs of languages in the same family, compared to pairs in completely different branches. However, there is no *a priori* value for the parameter, and therefore different possibilities are tested. Two additional values are used, namely $\delta= 0.5$ and $\delta= 1$, the first in agreement with FEARON (2003) and the second with ESTEBAN et al. (2012). The significance of the coefficients of the domain variable as well as of the other covariates is robust to these different values (Table 7).

In an additional specification, provincial specific trends are introduced, to control for a possible trend in the diversity index experienced by the single province independently. The model is estimated in first differences augmented with province fixed effects. As displayed in Table 8, the key communication variable exerts a positive and statistically significant effect on the fractionalization index.

The dependent variable in equation (4) is an index that is bounded between 0 and 1, being the variable a proportion. The use of techniques that do not take into consideration this specific form can be problematic, as far as there is no guarantee that the fitted values of the regression lie within the admissible interval, and predictions larger than one or negative can result. To solve this problem, an alternative estimation is conducted, applying a logistic transformation to the dependent variable. The resulting variable, while mapping the original one, is no more constrained to lie between 0 and 1, and the fixed effect estimation can be applied.[12] This procedure has the pitfall that cannot produce a mapping if the original variable is exactly zero or one. This is not a problem here, as these extreme cases do not occur. The results of this alternative procedure are reported in Table 9. No relevant changes emerge in terms of sign of the coefficients, whereas in some cases the significance of the coefficients is affected. The transformation of the dependent variable affects the significance of the communication variable, which turns statistically insignificant.

6. Conclusion

The objective of this paper is to identify possible factors explaining the increasing ethnic heterogeneity of Italian provinces. Along with standard covariates, the paper introduces an additional determinant, suggested by the O-P theoretical model, on the development of ethnically heterogeneous cities. In the model, communication improvements are found to enhance the ethnic diversity of cities, as they enable migrants to respond to the positive externalities generated by a culturally heterogeneous environment, while contrasting the dilution of cultural identities, when one moves from an ethnic enclave to a more racially diversified environment.

The degree of heterogeneity of provinces is measured by the well-known index of ethno-linguistic fractionalization (ELF), which captures both the richness in terms of number of ethnic groups, and the evenness of the groups' abundance. An ethnically diverse setting is maximized by means of a rich mix and an even distribution of the different ethnic groups in the province. The preliminary description of the data reveals that Italy is increasingly becoming an important destination for foreigners. From being historically a country of large emigration, from the 1970 the

net balance of migration reversed and nowadays the presence of foreigners continues to increase. Immigration and geography are highly correlated, as far as migrants are more likely to settle in North-Central Italy, and less in Southern Italy. The distribution of migrants along the Italian territory influences the index of fractionalization. The index is higher in the Northern and Central Italian provinces, whereas Southern provinces show lower values of diversity. This implies that ethnically diverse provinces are disproportionately located in Northern and Central Italy.

The empirical estimations provide support to the prediction of the theoretical model and they reveal that communication improvements represent a valid explanation for the birth of ethnically heterogeneous provinces. The two-way error component models for the fractionalization index report that easier communication, proxied by the number of internet domains registered, increases the degree of heterogeneity of the provinces. The positive and significant coefficient of the internet domain is robust to a larger disaggregation of the ethnic groups for migrants, to the incorporation of a measure of linguistic distance between the different groups, computed for different values of the discount factor and to the incorporation of province specific trends.

A final robustness check is performed applying a logistic transformation to the index of ethnic fractionalization. This is done to limit the implications connected to estimations where the dependent variable is bounded. The results regarding the communication variable are however not robust to this alternative specification.

The empirical findings for Italy suggest that foreigners respond to economic opportunities available in the host countries. This mechanism is more likely to happen when the external background facilitates them in pursuing such a strategy. Foreigners chose to locate outside the ethnic enclave, as long as the available technologies help them to maintain contacts with the own ethnic group. Given the established positive effects of diversity on economic outcomes, this exercise suggests that an additional instrument – other than ethnic specific policies- is available to host countries to pursue an ethnic heterogeneous setting. It should be noted however that for data constraints this exercise considers only legal migrants. Illegal migrants are less able to benefit from

the opportunities available at destination countries, given their status. They may need more extensively to live inside the ethnic enclave, as the ethnic network happens to be crucial in assisting illegal migrants in looking for a job or in other activities in destination countries. This result may not hold if the analysis includes those migrants who do not possess the permit to stay.

Notes
1. BISIN et al. (2006) on the contrary report that ethnic identities are more intense in mixed rather than ethnically homogeneous neighbourhoods. They find that minority groups enhance their ethnic lifestyle, as a response to an hostile racial or ethnic environment, which characterise the mixed neighbourhood.
2. See FEARON (2003) for a clarification of the concept of language distance.
3. The population density is around 200 inhabitants per squared kilometres.
4. Rilevazione sulla "Popolazione residente comunale straniera per sesso ed anno di nascita"
5. The official statistics used in this analysis only collect the stock of regular migrants. Unfortunately no official data are available regarding the number and the distribution of irregular migrants over time.
6. Given that in Italy the number of administrative provinces is not constant in time, as new provinces have been added during the years, it is considered the list of provinces as it was in 2002.
7. The index is computed by selecting the thirty most representative origin countries and by grouping the remaining migrants in the "other" group. See section 5 for details.
8. An additional way to measure the level of infrastructures considers the monetary value of the capital, as suggested by GOLDSMITH (1951). Capital endowment is defined as the total sum of each annual investment in the specific capital good, with the number of years included in the computation capturing the average useful life of the good. For Italy a detailed estimation of the monetary value of the infrastructures has been computed by PICCI (2002) at NUTS 3 level, considering the public spending of the provinces in the specific goods. For the purpose of this study however, these estimates suffer a major drawback, as the expenses for important infrastructures such as the speed rail, are sustained at national level and are not imputed in the budget of the single provinces. For this reason, these measures of infrastructure are not considered in this study.
9. Only selected years have been considered in the estimations as the key communication variable is available only in 2002, 2004, 2007 and 2010.
10. For robustness checks, the standard errors are also adjusted for regional clusters. These estimations have not been reported for space constraints but are available upon request. No major changes in the tests of the parameters of the key explanatory variables result.
11. For each group of migrants, a single language has been selected, namely the one spoken by the majority of the population. This is a strong assumption, in particular for countries such as India, where the number of official languages is extremely vast. However, the data set does not allow the identification of the linguistic groups the migrants belong to.
12. In a cross section, two additional solutions could have been used. The first one assumes that proportion follows a beta distribution and estimations are performed accordingly, while the second applies the fractional logit model, proposed by PAPKE and WOOLDRIDGE (1996). However, to my knowledge, the properties of these estimations in a panel context have not been studied.

7. References

ACCETTURO A, BUGAMELLI M. and LAMORGESE A. (2012) Welcome to the machine: firms' reaction to low-skilled immigration, Economic working papers n. 846, Bank of Italy, Economic Research and International Relations Area

ALESINA A. and LA FERRARA E. (2005) Ethnic Diversity and Economic Performance, Journal of Economic Literature, 43, 762-800

ANCE (2009) Secondo Rapporto sulle Infrastrutture in Italia, Edilstampa, Rome

ASHRAF Q. and GALOR O. (2007) Cultural Assimilation, Cultural Diffusion and the Origin of the Wealth of Nations, CEPR Discussion Papers 6444

ASLUND O. (2005) Now and forever? Initial and subsequent location choices of immigrants, Regional Science and Urban Economics, 35(2), 141-165

BALTAGI B. H. (2005) Econometric Analysis of Panel Data, Wiley, Chichester

BARTEL A (1989) Where do the new immigrants live?, Journal of Labor Economics, 7(4), 371–391

BAUER T, EPSTEIN G. and GANG I. (2002) Herd effects or migration networks? The location choice of Mexican immigrants in the US, IZA Discussion Paper n. 551

BAUER T, EPSTEIN G. and GANG I. (2005) "Enclaves, language and the location choice of immigrants, Journal of Population Economics, 18(4), 649–662

BELLINI E., OTTAVIANO G.I.P., PINELLI D. and PRAROLO G. (2011) Cultural Diversity and Economic Performance: Evidence from European Regions, in CRESCENZI, R. and PERCOCO M. (Eds) Geography, Institutions and Regional Economic Performance, Springer Eds

BISIN A., PATACCHINI E., VERDIER T. and ZENOU Y. (2006) Bend It Like Beckham: Identity, Socialization and Assimilation, CEPR Discussion Papers 5662

BORJAS G. (2001) Does Immigration Grease the Wheels of the Labor Market?, Brookings Papers on Economic Activity, 32(1), 69-134.

BOSSERT W., D'AMBROSIO C. and LA FERRARA E. (2011) A Generalized Index of Fractionalization, Economica, 78(312), 723-750

CAMERON C., GELBACH J. and MILLER D. (2008) Bootstrap-Based Improvements for Inference with Clustered Errors, Review of Economics and Statistics, 90, 414-427

COLLIER P. (2001) Implications of ethnic diversity, Economic Policy, 16(32), 127-166

DAMM A. (2009) Determinants of recent immigrants' location choices: quasi-experimental evidence, Journal of Population Economics, 22(1), 145-174

DESMET K., ORTUNO-ORTIN I. and WEBER S. (2009) Linguistic Diversity and Redistribution, Journal of the European Economic Association, 7(6), 1291-1318

EASTERLY W. (2001) Can Institutions Resolve Ethnic Conflict? Economic Development and Cultural Change, 49(4), 687-706

EASTERLY W. and LEVINE R. (1997) Africa's Growth Tragedy: Policies and Ethnic Divisions, The Quarterly Journal of Economics, 112(4), 1203-50

ESTEBAN J., MAYORAL L. and RAY D. (2012) Ethnicity and Conflict: an Empirical Study, American Economic Review, 102(4), 1310-42

FEARON J. (2003) Ethnic and Cultural Diversity by Country, Journal of Economic Growth, 8, 195-222

GLAESER E. L., KALLAL H. D., SCHEINKMAN J. A. and SHLEIFER A. (1992) Growth in Cities, Journal of Political Economy, 100(6), 1126-52

GOBBO F. (2007) Infrastrutture e sviluppo: alcune riflessioni sulla situazione italiana, Economia Pubblica, 5-6, 5-28

GOLDSMITH R. W. (1951) A Perpetual Inventory Method of National Wealth, in Studies in Income and Wealth, 14, 5-74, National Bureau of Economic Research

GREENBERG J. H. (1956) The Measurement of linguistic diversity, Language,.32, 109-115

JACOBS J. (1969) The Economy of cities, New York, Vintage

JAEGER D. A. (2000) Local labor markets, admission categories and immigrant location choice, Working Paper, Department of Economics, Hunter College, NY

MOCETTI S. and PORELLO C. (2010) How does immigration affect native internal mobility? New evidence from Italy, Regional Science and Urban Economics, 40(6), 427-43.

MOULTON B. R. (1986) Random group effects and the precision of regression estimates, Journal of Econometrics, XXXII(1), 385-397

OTTAVIANO G. I. P. and PERI G. (2005) Cities and cultures, Journal of Urban Economics, 58(2), 304-337

OTTAVIANO G. I. P. and PERI G. (2006) The economic value of cultural diversity: evidence from US cities, Journal of Economic Geography, 6(1), 9-44

OTTAVIANO G. I. P. and PRAROLO G. (2009) Cultural Identity and Knowledge Creation In Cosmopolitan Cities, Journal of Regional Science, 49(4), 647-662

PAPKE L. E. and WOOLDRIDGE J. M. (1996) Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates, Journal of Applied Econometrics, 11(6), 619–632

PICCI L. (2002) Le infrastrutture in Italia. Le differenze territoriali e l'efficienza della spesa, in BALDASSARRI M., GALLI G. and PIGA G. (Eds) L'Italia nella competizione globale - Regole per il mercato, Edizioni il Sole 24 Ore, Milan

RAO R. C. (1982) Diversity: its measurement, decomposition, apportionment and analysis, Sankhyā, Series A, 44, 1-22

SPARBER C. (2010) Racial Diversity and Macroeconomic Productivity across US States and Cities, Regional Studies, 44(1), 71-85

VISCO I. (2012) L'efficienza della spesa per le infrastrutture, presented at the meeting: Infrastrutture e competitività 2012. 10 opere per la ripresa, Rome

ZAVODNY M. (1997) "Welfare and the locational choices of new immigrants" Economic and Financial Policy Review, issue Q II, pp. 2-10

ZAVODNY M. (1999) Determinants of recent immigrants' locational choices, International Migration Review, 33, 1014–1030

ZORLU A. and MULDER C. H. (2008) Initial and subsequent location choices of immigrants to the Netherlands, Regional Studies, 41, 1–20

Appendix

Table A1: Description of variables and source

Variable	2002-2010	Description
Origin	Source: register office. NUTS 3	Thousands of foreigners from countries of origin
Pop	Source: register office. NUTS 3	Thousands of inhabitants
Less 14	Source: register office. NUTS 3	Percentage of people below 14 years of age.
15-64	Source: register office. NUTS 3	Percentage of people between 15 and 64 years of age
Over 65	Source: register office. NUTS 3	Percentage of people over 65 years of age
Primary Education	Source: EUROSTAT. NUTS 2	Economically active population with pre-primary, primary and lower secondary education
Secondary Education	Source: EUROSTAT. NUTS 2	Economically active population with upper secondary and post-secondary non-tertiary education
Tertiary Education	Source: EUROSTAT. NUTS 2	Economically active population with tertiary education
Act	Source: EUROSTAT. NUTS 2	Economic activity rates, age 15 and over
GDP growth	Source: OECD. NUTS 3	Growth of GDP per capita, at constant price
Railroad	Source: EUROSTAT, Regional statistics (REGIO). NUTS2	Total length of railroad (Km).
Motorway	Source: EUROSTAT, Regional statistics (REGIO). NUTS2	Total length of motorways (Km).
Domain	Source: Institute for Informatics and Telematics, Italian National Research Council. NUTS 2	Number of domains registered over 10'000 inhabitants.

Table 1: The average share of foreigners (%)

	2002	2003	2004	2005	2006	2007	2008	2009	2010
TOTAL	2.55	3.29	3.87	4.29	4.67	5.48	6.21	6.72	7.19
North-Centre	3.40	4.36	5.15	5.73	6.25	7.24	8.16	8.78	9.35
South	0.95	1.28	1.49	1.62	1.74	2.22	2.57	2.87	3.18

Table 2: The average index of Fractionalization

	2002	2003	2004	2005	2006	2007	2008	2009	2010
TOTAL	0.050	0.064	0.075	0.083	0.090	0.105	0.118	0.127	0.136
North-Centre	0.067	0.085	0.100	0.111	0.120	0.139	0.154	0.165	0.175
South	0.019	0.025	0.029	0.032	0.034	0.044	0.050	0.056	0.062

Table 3: The average index of Greenberg

	2002	2003	2004	2005	2006	2007	2008	2009	2010
TOTAL	0.019	0.023	0.027	0.029	0.031	0.034	0.037	0.040	0.042
North-Centre	0.025	0.030	0.035	0.038	0.041	0.044	0.048	0.052	0.055
South	0.008	0.010	0.011	0.012	0.013	0.014	0.016	0.017	0.019

Table 4: Distribution of foreigners by macro area of origin- % and total

	2002		2010	
	%	**total**	**%**	**total**
EU	9.01	124'920	4.18	171'351
Central Eastern Europe	37.65	522'039	54.39	2'256'737
other Europe	1.07	12'762	0.37	13'379
Northern Africa	23.99	323'154	15.77	678'929
Western Africa	5.42	106'866	4.46	240'241
Eastern Africa	1.29	25'741	0.83	46'218
Central-Southern Africa	0.46	8'822	0.39	21'083
Latin America	6.82	128'046	5.59	354'186
North America	1.14	15'545	0.43	18'199
Central-Southern Asia	5.43	114'107	6.27	368'332
Western Asia	0.86	16'897	0.60	31'874
Eastern Asia	6.64	147'745	6.65	366'306
Oceania	0.23	2'295	0.07	2'642

Table 5: Improvements in transportation

Annual percentage change (%) 2002-2010	
Railroad (Km)	0.6
Motorway (Km)	0.4
Domain	43.6

Table 6: Two-way Error Component Model

	Dependent variable: Index of Fractionalization*100				
	(1)	(2)	(3)	(4)	(5)
Year 2004	2.0192***	1.6436***	2.0703***	2.0266***	1.6913***
	(0.2647)	(0.3226)	(0.2662)	(0.2670)	(0.3246)
Year 2007	4.8193***	3.8809***	4.8326***	4.8353***	3.9548***
	(0.4567)	(0.6496)	(0.4571)	(0.4650)	(0.6586)
Year 2010	7.5539***	5.8813***	7.5624***	7.5718***	5.9948***
	(0.5795)	(1.0070)	(0.5811)	(0.5871)	(1.0139)
Population	0.0052	0.0022	0.0041	0.0053	0.0020
	(0.0060)	(0.0062)	(0.0060)	(0.0061)	(0.0062)
Less 14	2.4643***	2.2466***	2.4035***	2.4609***	2.2316***
	(0.2223)	(0.2526)	(0.2276)	(0.2225)	(0.2571)
15_64	0.9681***	0.9226***	0.9325***	0.9672***	0.9107***
	(0.2111)	(0.1933)	(0.2043)	(0.2108)	(0.1907)
Secondary Educat.	0.0010	0.0247	-0.0041	0.0025	0.0234
	(0.0440)	(0.0438)	(0.0446)	(0.0444)	(0.0433)
Tertiary Education	0.2674***	0.2183***	0.2475***	0.2646***	0.2101***
	(0.0652)	(0.0596)	(0.0648)	(0.0656)	(0.0610)
Activity Rate	0.1790**	0.1291	0.1807**	0.1757**	0.1285
	(0.0815)	(0.0811)	(0.0827)	(0.0827)	(0.0849)
GDP growth	0.0183**	0.0209***	0.0204**	0.0179**	0.0210**
	(0.0083)	(0.0079)	(0.0083)	(0.0087)	(0.0082)
Domain		0.0074**			0.0070**
		(0.0035)			(0.0035)
Rail			0.0017		0.0007
			(0.0014)		(0.0013)
Motorway				-0.0010	-0.0012
				(0.0031)	(0.0029)
Albania	0.0001	0.0001	0.0001	0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Serbia and Montenegro	0.0002***	0.0002***	0.0002***	0.0002***	0.0002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Bosnia and Herzegovina	-0.0004	-0.0004	-0.0004	-0.0004	-0.0004
	(0.0008)	(0.0007)	(0.0007)	(0.0008)	(0.0007)
Macedonia, FYR	0.0004*	0.0004*	0.0004*	0.0004*	0.0004*
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Croatia	0.0022	0.0024*	0.0021	0.0021	0.0023*
	(0.0014)	(0.0013)	(0.0014)	(0.0014)	(0.0013)
Bulgaria	0.0013***	0.0013***	0.0013***	0.0013***	0.0012***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0004)
Romania	-0.00002	-0.00001	-0.00001	-0.00002	-0.000003
	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
Poland	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Ukraine	0.00005	0.00004	0.00002	0.00004	0.00003
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Moldova	0.0001	0.00002	0.0001	0.0001	0.00003
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Russia	0.0004	0.0004	0.0004	0.0003	0.0004
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
France	-0.0006	0.0000	-0.0007	-0.0006	-0.0001
	(0.0013)	(0.0014)	(0.0013)	(0.0013)	(0.0014)

Cont.

	(1)	(2)	(3)	(4)	(5)
			Cont.		
Germany	0.0005	0.0002	0.0006	0.0005	0.0002
	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0006)
UK	0.0025**	0.0023*	0.0027**	0.0025**	0.0024**
	(0.0011)	(0.0012)	(0.0011)	(0.0011)	(0.0012)
Philippines	-0.0007**	-0.0006*	-0.0007**	-0.0007**	-0.0006*
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
China	0.0002	0.0002	0.0002	0.0002	0.0002
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Bangladesh	-0.0002	-0.0002	-0.0002	-0.0002	-0.0002
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Sri Lanka	-0.0005**	-0.0005**	-0.0005**	-0.0005**	-0.0004*
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
India	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Pakistan	-0.0003**	-0.0004**	-0.0003**	-0.0003**	-0.0004**
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Algeria	-0.0026**	-0.0026**	-0.0025**	-0.0026**	-0.0025**
	(0.0011)	(0.0012)	(0.0011)	(0.0011)	(0.0012)
Egypt, Arab Rep.	0.0005***	0.0005***	0.0006***	0.0006***	0.0005***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Morocco	0.00004	0.0001	0.00002	0.00004	0.00004
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Tunisia	0.0010***	0.0010***	0.0009***	0.0009***	0.0010***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Ghana	-0.0005	-0.0004	-0.0005	-0.0005	-0.0004
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Nigeria	0.0010*	0.0011*	0.0010*	0.0010	0.0010*
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Senegal	-0.0007**	-0.0006**	-0.0006**	-0.0007**	-0.0006**
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Brazil	-0.0003	-0.0004	-0.0003	-0.0004	-0.0004
	(0.0007)	(0.0008)	(0.0007)	(0.0007)	(0.0007)
Ecuador	-0.0001***	-0.0001*	-0.0001**	-0.0001***	-0.0001*
	(0.00004)	(0.00004)	(0.00004)	(0.00004)	(0.00004)
Peru	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Observations	412	412	412	412	412
Number of codes	103	103	103	103	103
R-squared	0.9809	0.9816	0.9811	0.9809	0.9816

Notes: robust standard errors in parenthesis. Standard Errors adjusted for 103 clusters. *,**,*** significant at 10%, 5% and 1%, respectively. The least recent year dummy is always the one removed from regression. The other excluded variables are the percentage of people over 65 years of age and the percentage of people with primary education.

Table 7: Two-way Error Component Model. Alternative aggregation and alternative index

	Fractionalization*100	Greenberg *100	Greenberg *100	Greenberg *100
	Alternative aggregation	δ=0.05	δ=0. 5	δ=1
	(1)	(2)	(3)	(4)
Year 2004	1.6494***	0.3549***	0.9798***	1.2726***
	(0.3075)	(0.1174)	(0.1820)	(0.2343)
Year 2007	3.9021***	0.7363***	2.1662***	2.8926***
	(0.6227)	(0.2207)	(0.3520)	(0.4608)
Year 2010	5.9123***	1.1636***	3.2996***	4.3791***
	(0.9685)	(0.3506)	(0.5422)	(0.7083)
Population	0.0029	0.0014	0.0026	0.0030
	(0.0062)	(0.0022)	(0.0037)	(0.0047)
Less 14	2.3155***	0.6873***	1.4062***	1.7648***
	(0.2515)	(0.0880)	(0.1368)	(0.1785)
15_64	0.9354***	0.1483**	0.4730***	0.6556***
	(0.1921)	(0.0643)	(0.0988)	(0.1320)
Secondary Educat.	0.0294	0.0355*	0.0264	0.0229
	(0.0420)	(0.0190)	(0.0274)	(0.0337)
Tertiary Education	0.2190***	0.0584***	0.1208***	0.1561***
	(0.0596)	(0.0193)	(0.0353)	(0.0453)
Activity Rate	0.1198	-0.0133	0.0379	0.0701
	(0.0806)	(0.0236)	(0.0434)	(0.0575)
GDP growth	0.0218***	0.0046	0.0129***	0.0165***
	(0.0080)	(0.0031)	(0.0045)	(0.0058)
Domain	0.0077**	0.0024*	0.0041**	0.0052**
	(0.0034)	(0.0012)	(0.0018)	(0.0024)
Observations	412	412	412	412
Number of codes	103	103	103	103
R-squared	0.9822	0.9691	0.9816	0.9820

Notes: robust standard errors in parenthesis. Standard Errors adjusted for 103 clusters. *,**,*** significant at 10%, 5% and 1%, respectively. The least recent year dummy is always the one removed from regression. The other excluded variables are the percentage of people over 65 years of age and the percentage of people with primary education. Variables with the number of foreigners from the most important origin countries are included. See Table 7 for the list of main origin countries.

Table 8: Error Component Model. Province-specific trends

Dependent Variable: Delta Fractionalization Index	
Delta Population	-0.0156**
	(0.0069)
Delta Less 14	0.3129
	(0.2033)
Delta 15_64	0.1064
	(0.1276)
Delta Secondary Education	-0.0350**
	(0.0174)
Delta Tertiary Education	-0.0379
	(0.0259)
Delta Activity Rate	0.0045
	(0.0452)
Delta GDP growth	0.0136***
	(0.0041)
Delta Domain	0.0060***
	(0.0020)
Observations	309
Number of code	103
R-squared	0.7939

Notes: robust standard errors in parenthesis. Standard Errors adjusted for 103 clusters. *,**,*** significant at 10%, 5% and 1%, respectively. The excluded variables are the percentage of people over 65 years of age and the percentage of people with primary education. Variables with the number of foreigners from the most important origin countries are included. See Table 7 for the list of main origin countries.

Table 9: Alternative estimation. Logit Transformation

	Fractionalization Index	Greenberg Index δ=0.05
	(1)	(2)
Year 2004	0.4003***	0.3523***
	(0.0454)	(0.0389)
Year 2007	0.7827***	0.6283***
	(0.0868)	(0.0790)
Year 2010	1.1104***	0.9196***
	(0.1255)	(0.1225)
Population	-0.0016*	-0.0001
	(0.0008)	(0.0008)
Less 14	-0.0263	-0.0451
	(0.0287)	(0.0292)
15_64	0.0437*	0.0029
	(0.0239)	(0.0240)
Secondary Education	0.0002	-0.0055
	(0.0064)	(0.0056)
Tertiary Education	0.0247**	-0.0019
	(0.0096)	(0.0091)
Activity Rate	0.0139	0.0335***
	(0.0100)	(0.0095)
GDP Growth	0.0007	0.0003
	(0.0011)	(0.0012)
Domain	-0.0002	0.00003
	(0.0004)	(0.0004)
Observations	412	412
Number of code	103	103
R-squared	0.9756	0.9629

Notes: robust standard errors in parenthesis. Standard Errors adjusted for 103 clusters. *,**,*** significant at 10%, 5% and 1%, respectively. The least recent year dummy is always the one removed from regression. The other excluded variables are the percentage of people over 65 years of age and the percentage of people with primary education. Variables with the number of foreigners from the most important origin countries are included. See Table 7 for the list of main origin countries.

NOTE DI LAVORO DELLA FONDAZIONE ENI ENRICO MATTEI

Fondazione Eni Enrico Mattei Working Paper Series

Our Note di Lavoro are available on the Internet at the following addresses:
http://www.feem.it/getpage.aspx?id=73&sez=Publications&padre=20&tab=1
http://papers.ssrn.com/sol3/JELJOUR_Results.cfm?form_name=journalbrowse&journal_id=266659
http://ideas.repec.org/s/fem/femwpa.html
http://www.econis.eu/LNG=EN/FAM?PPN=505954494
http://ageconsearch.umn.edu/handle/35978
http://www.bepress.com/feem/

NOTE DI LAVORO PUBLISHED IN 2014