Multicultural cities, communication and transportation improvements. An empirical analysis for Italy

Cristina Cattaneo[♣]

June 2010

Abstract

This paper tests empirically the predictions of the Ottaviano and Prarolo model on the development of multicultural cities. The model predicts that urban diversity increases as communication and transportation costs decrease. Multicultural cities emerge as foreigners respond to positive externalities generated by a culturally heterogeneous environment. 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. Conversely, an ethnically diversified environment generates negative externalities as people pay the psychological costs of living with people having a different culture. The hypothesis tested empirically is whether gradual improvement in distant communication boosts the generation of multicultural cities, as movers increasingly rely on an enlarged community for identity transmission, rather than on localized peer effects of segregated environment. The empirical estimation provides support to the prediction of the model. A better access to the airports, a reduction in the cost of flying as well as the improvements in internet communication are found to increase city diversity.

JEL code: R11, F22
Keywords: multicutural cities, ethnic diversity, productivity
Acknowledgments: This paper was funded within the EC project "Sustainable Development in a Diverse World" VI Framework Program. Usual disclaimers apply.

[♣] Cristina Cattaneo, Fondazione Eni Enrico Mattei (FFEM), cristiana.cattaneo@feem.it

1. Introduction

The issue of racial and ethnic segregation has been widely studied and the interest has been posed in the economic implications of marginalization, for both marginalised groups and for the society as a whole. Moreover the mechanisms that drive racial segregation have been investigated. A novel interest in the ethnic structure of a society has been recently introduced, where the crucial node has shifted from studying the costs of segregation to studying the benefits of integration. The emigration flows from less-developed to economic developed countries intensify rapidly and this trend is likely to continue in the future.

The enlargement of the community of foreigners is an unquestioned characteristic of our cities, and will bring important economic implications. Ottaviano and Peri (2006) and Bellini et al. (2009) report that multicultural cities generate positive externalities, because a diverse environment fosters productivity. 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. The existence of positive implications of multiculturalism has motivated the investigation of the determinants of cosmopolitan cities.

According to Ottaviano and Peri (2006) and Bellini et. al (2009), the positive externalities of the multiculturalism are maximized for increasing social interactions among the different ethnic groups. An environment, where all minority groups are segregated within their own ethnic enclave will not produce the aforementioned benefits, as in this context the exchanges of ideas between ethnically diverse people are limited. At the same time however, the cultural identities of movers are likely to be preserved within the ethnic enclave.[1] Migrants face the trade-off between an environment which maximizes their social utility and an environment which maximizes their economic utility. In other words, the migrants face the dilemma of choosing either a segregated enclave or a multicultural setting. A solution to this dilemma can be reached by means of communication improvements. Progress in transportation and communication helps offsetting the dilution of cultural identities, which occurs when one moves from a segregated to a more integrated environment. Improvements of this type reduce the cost of move, as help migrants maintaining contacts with friends and family of their own ethnic group. This in turn should increase the city heterogeneity, as individuals of different ethnicities eventually locate in a multicultural environment, rather than stay in a segregated context, in order to respond to the positive

[1] Bisin et al. (2006) on the contrary report that ethnic identities are more intense in mixed rather than segregated 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.

externalities generated by a culturally heterogeneous environment. 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. Ottaviano and Prarolo (2009) formalize these hypotheses and predict that two ethnic groups integrate in multicultural cities when communication is easy, whereas they segregate in different cities when diaspora members find it hard to communicate at distance. The objective of this paper is to provide an empirical test to the theoretical predictions of the Ottaviano and Prarolo model (2009), O-P hereafter, on the development of multicultural cities. Given the substitutability between the concept of segregation and integration, this paper draws on the large literature that studies the determinants of segregation to assess the relationship between gradual improvement in distant communication and the rise of multicultural cities.

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

2. Literature

The upward trend in migration has introduced a rising interest in ethnic heterogeneity and its implication for economic performance. One aspect related to ethnic heterogeneity is that of segregation, as the different ethnic groups in a society may cluster in ethnic enclaves, threatening the foundation of the civil societies.

This paper is related to several strand of literature, the first being the empirical studies that analyse the causes of residential segregation. The concept of segregation and integration are complements to one another, as segregation refers to a situation where the two types are concentrated in different cities, whereas integration corresponds to an even distribution of the groups between the cities. The analysis of residential segregation, despite being scarce for European cities, has been prominent in the research agenda of U.S. In this literature, two factors have been traditionally identified as responsible for racial segregation. The first is related to the discrimination of whites against the black or hispanic minorities, while the second are the cultural ties that induce minorities to live in own-race abundant neighbourhood (Patterson, 1997; Thernstrom and Thernstrom, 1997; Ihlanfeldt and. Scafidi, 2002). Cultural identity is under strain in a heterogeneous environment and to preserve it, minorities chose to cluster in ethnic homogenous enclaves. This last view, however, has been challenged by Bisin et al (2006), where the identity of an ethnic minority is found to be more intense, ceteris paribus, in mixed rather than in segregated neighbourhoods.

A test on the two standard determinants of segregation has been offered by Cutler et al. (1999). The authors explore the evolution of racial segregation over time in American cities and test whether segregation is the result of the natives' or alternatively the immigrants' action. They find that the segregation of the black minority arises from discriminating actions taken by incumbent natives. The white majority either enforces separation through rising barriers to keep blacks out of white neighbourhood, or pursues the desire to live with other natives and therefore segregate itself. An additional motivation have been introduced by Cutler et al. (2008) to analyse the rising trend in immigrants' segregation, experienced by U.S. cities. They use the dissimilarity and isolation index, computed at ethnic and city level, to test the discrimination and the cultural theories described above, against a transportation theory. According to the transportation theory, there might be class differences in commuting modes such that natives relocate in automobile-dependent urban periphery. On the contrary immigrants locate in the urban core and rely on public transit as soon as this is a valid alternative to private means. They find that cultural reasons are responsible for increased segregation. In particular, lower experience in United States and linguistic distance between immigrants' native tongue and English increase segregation. They also provide support to the transportation hypothesis, as the metropolitan areas with larger reliance on public transport show greater segregation. This last mechanism introduces the possibility that exogenous factors, such as local government measures, can shape the residential segregation by race.

An alternative explanation for segregation is as well provided by Dawkins (2005), who tests whether increased local government fractionalization, the so called "Tiebout choice", is responsible for larger ethnic segregation, using both metropolitan level and household level data. The author reports first, a positive impact of Tiebout choice on black-white residential segregation in U.S. and second, a larger effect for segregation across jurisdictions than across neighbourhoods. An extended version of the Tibeout model is tested in Rhode and Strumpf (2003). The authors theoretically predict that a decline in mobility costs should result in a greater heterogeneity across communities, as the lower are mobility costs, the greater are the incentives of sorting in homogeneous communities. The empirical analysis however provides little evidence in favour of this mechanism. Along with a secular decline in mobility costs, the analysis reveals a tendency of communities to become more alike. Local land use policies adopted by governments are another possible source of spatial distribution of ethnic groups, as far as they influence the path and composition of population growth. Residential measures that are biased towards low density residential development happen to favour the non-Hispanic White population of California. Hispanic and Asian growths, on the contrary, are boosted by housing stocks biased towards higher-density development (Quigley at al., 2004).

The second related literature is about the economic effects of diversity. While in cross-countries analyses the relationship between diversity and economic performance is found to be quite mixed (Easterly and Levine, 1997; Collier, 2000; Easterly, 2001; Alesina and La Ferrara, 2005) in finer spatial units, such as cities, a positive link is reported. This is because in a dense environment, where differences are more likely to interact, the scope for benefits related to diversity is larger. In Ottaviano and Peri (2005) US-born citizens are the more productive the more culturally diversified is the environment. Moreover, in this paper emerges that local diversity negatively influences the provision of public goods, but this effect is more than offset when a certain degree of integration between communities takes place. Ottaviano and Peri (2006) find that US-born workers living in cities with higher cultural diversity are paid, on average, higher wages and pay higher rents than their counterpart living in more culturally homogenous cities. This joint positive effect of diversity on wage and rent is crucial as it indicates that a positive productivity effect dominates any utility effect that arises from living in cities. A more multicultural urban environment makes US born citizens more productive.

Jacobs (1969) emphasizes the powerful link between 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. The benefit of interactions among different skills is as well evident in Glaeser et al. (1992), where larger innovations in production occur in more diversified cities due to cross-fertilization across different sectors. Workers are highly mobile within a city and carry with them the knowledge acquired in different sectors. A high degree of diversification implies a larger opportunity of contamination between different knowledge and a larger chance that new ideas arise. 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 cultural homogeneity may have prevented China from maintaining this supremacy.

Finally, the paper is related to the literature on the measurement of diversity. The economics literature in this field is still in 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. (2009) 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 index as the Greenberg's (1956) formula, 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, being these socio-economic factors and their interactions 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 distances, 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 linguistic distance.

3. Methodology

The O-P model (2009) analyses the emergence of multicultural cities as the result of cultural diaspora. The model assumes two types of individuals, belonging to two different cultural groups, type 1 and type 2. Each individual is endowed with one unit of labour and the total number of individuals of the two types (L_1 and L_2) is exogenously given. The endogenous number of individuals of type k living in city j is given by L_{jk}, with $L_{jk} + L_{ik} = L_k$.

The model introduces the possibility that the cultural trait, namely the type, is passed from parents to children. The time available to the individuals, normalized to one, is allocated to either pass the cultural trait or work. Utility derives from consumption of a material good (y_{jk}) and a cultural good (C_{jk}):

$$U_{jk} = \beta \log y_{jk} + (1- \beta) \log C_{jk} \qquad (1)$$

Three factors in the model affect the probability that the children end up with the parents' trait. These are the time that the parent devotes to cultural transmission (e_{jk}), the share of the individuals of the same type in the same city (l_{jk}) and the share of the individuals of the same type in the other city (l_{ik}). Clearly, global interactions (l_{ik}) are weaker than local ones (l_{jk}). Given a certain degree of substitutability between the family effort and the peer effects in cultural transmission, it is assumed that the consumption level of the cultural good is:

$$C_{jk} = e_{jk}^{\lambda} (l_{jk}+\phi\, l_{ik})^{1-\lambda} \quad (2)$$

with $\lambda \in (0,1)$ and $\phi \in (0,1)$

In this formulation, λ captures the relative importance of parental education in cultural transmission, $1-\lambda$ is the relative importance of the peer effect and ϕ is a communication parameter that weights the role of the localized peer effect. The expected net labour income is determined by time spent working, net of time spent in cultural transmission (e_{jk}) and the time lost in commuting $\Gamma(L_j)$. Therefore, the budget constraint is:

$$P_j\, y_{jk} = w_{jk}\, \Gamma(L_j)\, (1- e_{jk}) \quad (3)$$

where P_j is the price of the material good, and w_{jk} is the wage per unit of time. Maximizing (1) with respect to e_{jk} subject to (2) and (3) yields:

$$e^*= \frac{\lambda(1-\beta)}{\beta+\lambda(1-\beta)} \quad (4)$$

The model introduces the technology side. Given perfect competition, free trade across cities and labour as the only input in production, a_{jk} represents the productivity of an individual of type k in city j. The aggregate production is given by:

$$Y_j = a_{j1}\, [\Gamma(L_j)\, (1- e_{j1})\, L_{j1}] + a_{j2}\, [\Gamma(L_j)\, (1- e_{j2})\, L_{j2}] \quad (5)$$

The productivity of the two types of individuals is allowed to be different and it depends on the cultural composition of the cities. This is because individuals are assumed to benefit disproportionately from interactions outside their own groups. Only interactions within the same city enhance productivity and this implies that the occurrence of inter-cultural interactions depends solely on home city composition. In city j an individual of type 1 faces a probability l_{j1} and l_{j2} of meeting someone of her own group and of the other group, respectively. It follows that the average productivity of an individual of type 1 is $a_{j1}= a\, (\theta\, l_{j1}+ l_{j2})$, where a is the individual's productivity when exposed to inter-cultural interactions and $\theta \in (0,1)$ is a discount factor for the absence of inter-cultural interactions. The average wage of type k in city j is given by:

$$w_{j1} = a_{j1}= a\, (\theta\, l_{j1}+ l_{j2}) \quad (6)$$

The distribution of individuals is a spatial equilibrium when no individuals of either type may gain a higher utility by changing location. Taking (3), (1), (2) and (4), the indirect utility of an individual of type k in city $j \in (A,B)$ is defined as:

$$V_{jk}\, (L_{A1}, L_{A2}) = \beta\, \log\, (w_{jk}\, \Gamma(L_j)\, (1- e^*)) + (1 - \beta)\, \log\, ((e^*)^{\lambda}\, (l_{jk}+\phi\, l_{ik})^{1-\lambda}) \quad (7)$$

and the a spatial equilibrium arise at $L_{Ak} \in (0, L_k)$ when:

$\Delta V_k(L_{A1}, L_{A2}) \equiv V_{Ak}(L_{A1}, L_{A2}) - V_{Bk}(L_{A1}, L_{A2}) = 0$

Here we are interested in the symmetric cosmopolitan equilibrium, which is a situation where individuals are happy to live in two identical cities, where both types are equally represented. In this situation:

$\Delta V_1(L_{A1}, L_{A2}) = \Delta V_2(L_{A1}, L_{A2}) = 0$ for $L_A = L_1/2 + L_2/2$, $L_{A1} = L_{B1} = L_1/2$, and $L_{A2} = L_{B2} = L_2/2$.

This multicultural outcome is always an equilibrium, but it may be unstable. Assuming an adjustment process, where the only driving force in the migration process is determined by utility differential between location A and B, the stability of the multicultural system is determined by the following formula:

$$\phi > \frac{(1-\beta)(1-\lambda)(1+\theta) - \beta(1-\theta)}{(1-\beta)(1-\lambda)(1+\theta) + \beta(1-\theta)} \equiv \phi_B \qquad (8)$$

Condition (8) implies that the stability of the multicultural equilibrium is fostered by strong materialism (large β), large productivity gains (small θ), important parental cultural transmission (large λ) and easy communication at distance (large ϕ).[2]

To summarize, the model predicts that two groups integrate in multicultural cities when communication is easy, whereas they segregate in different cities when diaspora members find it hard to communicate at distance. Communication improvements help offsetting the dilution of cultural identities, when one moves from a segregated to a more integrated environment. In fact, communication improvements reduce the cost of move, as migrants maintain contacts with friends and family living in the ethnic enclave. This in turn should increase the city heterogeneity, as individuals of different ethnicities choose to locate in a multicultural environment rather than stay in a segregated context. The hypothesis tested is whether gradual improvement in distant communication boosts the generation of multicultural cities, as movers increasingly rely on an enlarged community for identity transmission, rather than on localized peer effects of segregated environment.

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

$$DIV_r = 1 - \sum_s (x_s)^2 \qquad (9)$$

where x_s is the share of foreigners of the specific origin group *s* computed for each province *r*. The index measures the likelihood that two individuals, randomly drawn from the population and living

[2] Details on how to reach condition (8) should be taken from the reference paper (Ottaviano and Prarolo, 2009)

in different provinces, have different ethnicities. This index allows me to measure not only the richness of a city in terms of ethnic groups, namely the number of groups that live in the city, but also 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. A diversity index that incorporates linguistic distance between the ethnic groups (Greenberg, 1956) is therefore computed. 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:

$$\text{Greenberg}_r = \sum_j \sum_s x_j x_s \tau_{js} \tag{10}$$

where the *x* denotes the shares of the different ethnic groups living in provinces 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:

$$\tau_{js} = 1 - \left(\frac{l}{m}\right)^{\delta} \tag{11}$$

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. 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.

A third index used is the Balassa specialization index, also known as Index of Revealed Comparative Advantage (Balassa, 1965). The index is:

$$\text{SPEC}_{rs} = \frac{x_r^s / \sum_r x_r^s}{\sum_s x_r^s / \sum_s \sum_r x_r^s} \tag{12}$$

where x_r^s is the number of foreigners of the specific origin group *s* living in province *r*. In this context, the index compares the share of foreigners from a certain origin area s located in a specific province r with the average share of the same group in the country. The index ranges from 0 to $+\infty$ and a value greater than one identifies a situation of specialization, as the share of the specific foreigners in that province is greater than the same foreigners' share from the totality of the country

provinces.[3] This index has been widely used in international trade to explain comparative advantages.[4]

The concept of segregation and integration are complement to one another. Given two ethnic groups, segregation refers to a situation where the two types are concentrated in different cities, whereas integration corresponds to an even distribution of the groups between the cities. This implies that for an empirical purpose, the analysis of integration can build on the vast literature that studies the determinants of segregation. This strategy is followed here and the typical explanatory variables for segregation are used in the estimation. This set of covariates is then augmented with the variables capturing transportation and telecommunication improvements.

Segregation has been computed in the empirical literature at the level of cities or metropolitan statistical area (MSA). The majority of estimations of segregation has been conducted applying US data, which allow to capture location details at the census tract level. Unfortunately, for European countries such high level of disaggregation is not possible as census data provide location information at only NUTS 3 level. This is the greatest level of disaggregation for the analysis proposed here, which corresponds to provinces. For the purpose of this analysis however, this data limitation should not represent a problem. The objective of the paper is to analyse the emergence of multicultural cities, which implies that we are interested in the degree of integration across cities. In other words we should focus on the way ethnic types sort across cities and not within a city.

4. Data Description

Two main sources of data are used for the analysis. The first is represented by the 1991 and 2001 Italian population Census conducted by the National Institute of Statistics (ISTAT). The Census collects socio-economic information of the Italian population and records the total number of foreigners from main areas of origin. The data are aggregated at NUTS 3 level (province). Nine areas of origin for the foreigners are available using the Census, namely Europe 12, Central Eastern Europe, other Europe, Northern Africa, other Africa, North America, Latin America, Oceania and Asia. Table A1 in the Appendix provides details of the countries entering in the different groups. This information allows me to compute the fractionalization index for each spatial unit of analysis. The dataset is a balanced panel and it contains 95 different provinces for two points in time. The second source of data is the annual collection of information of the foreign population conducted by

[3] More precisely, the upper bound of the index is given by $\frac{\sum_s \sum_r x_r^s}{\sum_s x_r^s}$, which tends to ∞ when the weight of province r in terms of number of foreigners is marginal.

[4] See De Benedictis and Tamberi (2002) for a description of the properties of the index

ISTAT.[5] The data are taken from the register offices of each Italian commune, and 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, new registrations coming from different Italian towns and different countries. This data set allows a larger disaggregation in terms of the area of origins, as 13 groups are identified, namely EU,[6] Central Eastern Europe, other Europe, Northern Africa, Eastern Africa, Western Africa, Central-Southern Africa, North America, Latin America, Eastern Asia, Western Asia, Central-Southern Asia and Oceania (see Table A2 for details). The number of administrative provinces in Italy is not constant, as new provinces have been added in the years. The second data set forms a balanced panel, which contains 103 provinces in four different years, from 2004 to 2007.

The use of different datasets is motivated by data issues. In the theoretical model, communication improvements are as relevant as transportation improvements. Migrants not only may rely on faster transportation modes to maintain social ties with the own ethnic group in different part of the host country, but as importantly, may benefit from strong improvements in telecommunication, such as the possibility to use internet connections to talk to family and friends. While the geographical information on transportation improvements is available for the period 1991-2001, variables related to telecommunication improvements are scant for this period and not disaggregated at geographical level. As it will be described below, telecommunication variables are available only for the most recent data set.

Italy has been historically a country of large emigration. From 1861 it produced nearly 30 million of emigrants. Only in the '70s, 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 1a, 1b

The percentage of foreigners substantially increased between 1991 and 2001, being 0.5 in 1991 and raising to 2.1 in 2001. The geographical distribution of foreigners varied greatly and, among other factors, it is eventually influenced by the economic opportunities that the Italian regions offered. In 1991 North-Central Italy hosted 0.7 percent of immigrants, whereas in the South lived only 0.3 percent of foreigners. The gap widened in 2001, when 3 percent of foreigners were in the North-Central and only 0.9 percent lived in the South (Table 1a). The increasing trend in the presence of foreigners, as well as the disproportional contribution of the North-Central compared to the South, emerge also in the second panel of years (Table 1b).

TABLE 2a, 2b

[5] Rilevazione sulla "Popolazione residente comunale straniera per sesso ed anno di nascita"
[6] The list of EU countries refers to the situation in 2002, when countries of the eastern block were still not included.

The increase in the number of foreigners has produced an increase in diversity in Italy, as indicated by the rising trend in the fractionalization index (Tables 2a, 2b). Moreover, the index of fractionalization is highly correlated with space, being higher in the Northern and Central Italian provinces, and lower in Southern provinces.

TABLE 3

Even if many North-Central provinces displayed high value of diversity in all years, a certain mobility occurred between 2004 and 2007. Between 1991 and 2001 only provinces located in the Northern and Central part of Italy displayed the largest annual percentage change of the diversity index. On the contrary, between 2004 and 2007 the largest scores are displayed by the provinces in the South (Table 3). Despite such large mobility, however, the geographical gap between the North-Centre and the South persists also in the most recent years. Figure 1 shows the geographical distribution of the Italian provinces, grouped in different quartiles. While in 1991 some provinces in the South were represented in the top quartile group, from 2001 onward all provinces in the South were disproportionately represented in the bottom two quartiles of the distribution of the index and the situation did not vary after 2001.

FIGURE 1

The analysis can be further conducted in terms of area of origin of the foreigners. The ethnic mix of the foreign population varied widely during the period considered. In 1991 the most represented macro-areas of origin were Europe 12, followed by Northern Africa. These areas contributed with 21 and 19 percent of total stock, respectively (Table 4a). The situation changed in 2001, when countries of Central Eastern Europe became the largest source of migrants. Not surprisingly, the remarkable presence of foreigners from Central Eastern Europe continued also in the most recent period, when this group picked to more than half of the total stock of migrants. Northern Africa remains an important origin region, with a share varying from 17 percent in 2007 to 21 in 2004. Migrants from Europe 12, classified as EU after 2002, experienced a stable drop during the years.

TABLE 4a, 4b

A different perspective can be taken in what follows. Rather than using an index that captures the overall diversity at province level, it will be considered an index that varies both at province and at ethnic level. The use of a specialization index, such as the Balassa Index, allows me to identify if specific provinces host a disproportionately greater share of foreigners from a certain macro areas of origin. A value greater than one identifies a more than proportional incidence of these foreigners in the specific province, compared to the average Italian incidence. The index can be analysed in terms of mean values and standard deviation. In Table 5a, the groups other Africa

and Asia display low levels of specialization both in 1991 and 2001, as indicated by the less than unitary average value. The tendency of these groups of foreigners to keep low levels of concentration is confirmed in the second panel of years. All Asian groups and the groups Central Southern Africa, Western Africa and Eastern Africa have less than unitary mean values also from 2004 to 2007 (Table 5b). On the contrary, the groups Europe 12 (EU), other Europe, Northern Africa, North America and Oceania tend to have greater than unitary values, despite in some cases the value is very close to one. Finally the group Central Eastern Europe has below unitary Balassa index in 1991 and above unitary from 2001 onward. Tables 6a and 6b report the standard deviation of the index computed over all the provinces. An increase in the dispersion of the index identifies a situation of increasing specialization, namely the tendency of the foreigners to disproportionately localize in selected provinces and avoid some other locations. The groups Central-Eastern Europe and Northern Africa display a reduction in the standard deviation from 1991 to 2001 and this trend continues also from 2004 to 2007. A decreasing trend is displayed also by Central Southern Asia and Western Asia in the second panel years. Finally, there are no groups that exhibit a stable rising trend in the whole period of analysis. The remarkable decrease in the dispersion of the index for Central-Eastern Europe is eventually driven by the strong reduction in the maximum values that the index assumed in the period, as shown in Tables 7a and 7b. In 1991 the concentration of this group in one single province was extremely high, being the maximum 8.70. On the contrary, in the following years there are not episodes of likely high concentration, as indicated by lower extreme values.

TABLE 5a, 5b

To test the O-P theoretical model (2009), the key explanatory variable should be a measure of the transportation and telecommunication improvements, as a proxy for mobility and connection costs. Connections of individuals living in different cities became easier, both because the time to travel from one city to the other declined, and because new technologies made it simpler to people to communicate at short distance. The wide penetration of computers and internet connections is responsible for the drastic improvements in the communication.

TABLE 6a, 6b

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 motorway and railroad available for 1991 and 2001.[7] Alternatively, it can be exploited the improvement that

[7] 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 by the provinces in the specific goods. For the purpose of this study however, these estimates suffer

occurred between 1991 and 2001 in the access to some transportation infrastructure, such as the airports. In Italy, in this period, many small airports that used to be utilized for the transport of goods, switched to airport for the transport of passengers. This implied that the distance to the closest airport decreased significantly in this time interval, allowing a relevant saving in terms of time for travel. Therefore, a measure of the distance to the closest airport at NUTS 3 level and a measure of the total number of airports, located at less than 100 Km from the single province, are the third and fourth variables used, respectively. Another dimension that is exploited is the reduction in the cost of travelling by airplane. The number of low-cost air companies increased substantially in these years. The greater competition and the availability of discounted tickets increased the accessibility to flights and reduced the mobility cost. To capture these improvements, the number of passengers landing and departing in international flights is used as a proxy. This variable is available at NUTS 3 level. Table A3 provides a detailed description of the variables and their source.

TABLE 7

Surprisingly, between 1991 and 2001 the investments in infrastructures in the motorway and railroad are minimal as far as the physical measure is considered. As reported in Table 8, the overall kilometres of motorway in Italy increased annually by only 0.48 percent and of railroad by 0.03 percent. It is reported that Italy lags behind the other European countries in terms of infrastructure endowments and this gap has widened remarkably in the last 20 years, being now three times larger than it was 20 years ago (Gobbo, 2007). For example, large investments in motorway occurred till the ’80, when Italy owned the most modern and developed road network in Europe (ANCE, 2009). However, these investments nearly stopped hereafter, with drastic implications in terms of congestion. This implies that these physical measures represent a bad proxy in Italy for transportation improvements. On the other hand, if one considers the other variables of Table 8, a certain transportation improvement occurred. For example, the accessibility to the airports improved annually by 3 and 5 percent, if one considers the distance to closest airport and the number of near-by airports, respectively. The same is true for the accessibility to flights, as indicated by the 12 percent increase in the number of passengers.

TABLE 8

Regarding the telecommunication variables, one measure is used here and it is the number of internet domains registered at NUTS 2 level for the period 2004-2007. This variable seems to

a major drawback, as the expenses for important infrastructures such as the speed rail line, 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.

adequately capture the large improvements in telecommunication means, which occurred recently, as proved by the annual rate increase of 23 percent (Table 9).

TABLE 9

5. Empirical Estimation

To understand the determinants of the level of diversity across cities, the index of fractionalization is estimated for 1991 and 2001, and for 2004 to 2007. The units of observation are the Italian provinces. Different factors have been traditionally identified in the economics literature as responsible for the location sorting of ethnic groups. These are self-segregation of minorities, racial prejudice and discrimination as well as local level/governmental factors. First, ethnic specific factors are important sources of location choice as they influence the personal propensity for segregation. For example, migrants, in particular newly arrived, may prefer to cluster with their own group, to find support in the destination countries and to recreate the social background of the place of origin. Immigrants chose to live with people that have similar tastes and speak the same language. These factors are likely to be more important, the greater the cultural distance of foreign-born from natives. Second, segregation results from discriminating actions taken by incumbent natives. Discrimination and prejudice are influenced by socio-economic characteristics such as age, gender, household composition, education, marital status and family income. Alternatively, the urban form, the Tiebout choice, housing development laws have been identifies as other important determinants of segregation. After controlling for these factors, the key explanatory variable in this analysis is represented by telecommunication and transportation improvements, which proxy for the costs of mobility. The set of standard covariates is therefore augmented with the variables capturing transportation and telecommunication improvements. The following equation is estimated:

$$DIV_{rt} = \alpha + \beta\, pop_{rt} + \boldsymbol{\gamma}\, \mathbf{DEMO}_{rt} + \delta\, un_{rt} + \phi\, act_{rt} + \eta\, technology_{rt} + \boldsymbol{\lambda}\, \mathbf{ORIGIN}_{rt} + \varepsilon_{rt} \quad (13)$$

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

The dependent variable is ethno-linguistic fractionalization index, for the province *r* in year *t*. The total population (pop) captures the size of the city. The vector of socio economic variables (DEMO) includes the proportion of population in different age categories and education categories. These variables should proxy for discrimination factors. The variables unemployment rate (un) and activity rate (act) proxy for economic opportunities, that represent pull factors in the destination choice. The ethnic group variables (ORIGIN) capture the ethnic level propensity for living in a multicultural environment and include the total number of foreigners from the different areas of

origin. The variables technology improvements related to transportation and telecommunication (technology) are added. Unfortunately, these variables cannot be used jointly in the estimations, as the number of registered domains is available only for the most recent panel and *vice versa*. Table A4 provides a description of the variables and the source. The disturbance contains the province fixed effect, the time effect and the conventional stochastic disturbance term (Baltagi, 2005). Following Moulton (1986), all models are estimated correcting standard errors for clustering by regions.

The results of the two-way error component model in 1991 and 2001 are reported in Table 10. In column one, a parsimonious specification is estimated, where it is only controlled for the time and province fixed effects. A tendency toward rising diversity is experienced by the Italian provinces, and this is in agreement with the descriptive analysis. The coefficient of the 2001 year dummy is positive and statistically significant and it suggests that the average diversity levels were three percent higher in 2001 that they were in 1991.

In column two, the standard covariates described above are added. The time trend is now steeper compared to the parsimonious specification, and it is robust to the inclusion of the additional covariates. The size of the city has a non-statistically significant effect on diversity. Economic opportunities exert a well determined effect on the location choice of foreigners. Cities with a higher rate of activity and a lower rate of unemployment are, on average and ceteris paribus, more multicultural. Labour market features represent important pull factors, as far as foreigners effectively respond to the advantages offered by the economically vibrant cities. Diversity is higher in provinces with a younger population. This again can indicate that foreigners avoid depressed cities, which are also abandoned by the native population in working age. The proportion of people that obtained primary, secondary and tertiary education has a positive effect on diversity, even if the coefficient of secondary education is not statistically significant. Finally, the number of foreigners exerts an influence on the level of diversity, which depends on the groups of origin. An increase in foreigners of the Europe 12 area decreases the level of city diversity. On the contrary, an increase in the number of foreigners from Central-Eastern Europe and from northern Africa is associated with a greater level of multiculturalism.

Columns three to eight add the key transportation variables. The first is the length of the motorway expressed in kilometres. This variable exerts no significant effect on the level of diversity. The second is the length of the railroad and again the coefficient is not statistically significant. These results, however, are not surprising, on the ground that only minimal investments in the motorway and railroad systems have been put in place between 1991 and 2001 in Italy. As discussed before, in the ten years interval, the length of these infrastructures increased by only five

and 0.3 percent, respectively. The third variable, in column five, is the distance to the closest airport, expressed in kilometres. The coefficient has the expected sign and it is statistically significant at conventional level. Ten additional kilometres to the closest airport decreases diversity by 0.1 percent, on average and ceteris paribus. The coefficient of the fourth transportation variable, which is another proxy for the accessibility to airports, is positive and statistically significant. The magnitude of the coefficient suggests that an additional airport located at close distance increases the fractionalization index by 0.4 percent. In agreement with the theoretical model, these results suggest that a reduction in the transportation costs, expressed in terms of time for travelling, increases the degree of multiculturality of the cities. Movers eventually respond to positive externalities generated by a culturally heterogeneous environment and increasingly rely on an enlarged community for identity transmission, rather than on localized peer effects of a segregated environment. Gradual improvements in distant communication allow this process to happen, as it becomes easier for foreigners to maintain their ethnic identity while living outside the ethnic enclave. The fifth variable is the volume of passengers. As reported in column seven, the coefficient of the variable is positive but not statistically significant. In column eight the three transportation variables enter jointly. The coefficient of the variable distance is no more statistically significant, whereas the coefficient of number of airport is invariant, both in terms of magnitude and significance. The coefficient of the number of passengers turned statistically significant, indicating that an increase in the volume of passengers by five percent, which corresponds to an increase of about 100’000 passengers, augments diversity by 0.01 percent.

Table 11 reports the estimations for the second data set, relative to years 2004 to 2007. The estimations are overall in line with those reported for the previous panel, albeit some coefficients turned not statistically significant. As shown in column one, the year dummies are positive and statistically significant, and confirm that the rising trend in diversity continues till the recent years. The average diversity index is three percent higher in 2007 than it was in 2004. The largest increase in diversity occurred between 2006 and 2007, while in the other intervals the increase was less steep.

Column two introduces the standard regressors. As before, the inclusion of the covariates turned the diversity time trend steeper. Contrary to the previous regression, the volume of the population exerts a statistically significant effect on the index of fractionalization, and the negative coefficient indicates that larger cities are less diverse. In agreement with the previous findings, provinces with a younger population tend to be richer in terms of variety of ethnicities. The coefficient of the unemployment rate has a wrong sign while the activity rate has an insignificant coefficient. None of the coefficients of the education variables are statistically significant.

Regarding the main areas of origin, in line with the previous regression, the greater is the number of migrants from Central and Eastern Europe, the more multicultural are the cities. Also migrants from EU increase the level of diversity and the same do migrants from Western Africa and North America. The positive correlation between these ethnic groups and the level of diversity indicates that these migrants tend to avoid clustering in ethnic enclaves and locate homogeneously across the Italian provinces. This process occurs because multicultural provinces eventually offer the largest economic opportunities, which are exploited by migrants from these origin areas. In columns three the telecommunication variable is introduced. The coefficients of the time dummies and the demographic variables are robust to the inclusion of this additional variable, whereas the coefficient of the unemployment rate turns insignificant. The interesting finding is that the telecommunication variable exerts a positive and well determined effect on multiculturality. The number of internet domain is positively correlated with the level of diversity, providing a robust support to the theoretical prediction of the O-P model. Increasing the number of domain by ten units over 10'000 inhabitants, the fractionalization index augments by 0.1 percent.

The latter dataset has the advantage to be disaggregated by single country of origin. Up to now, the different countries have been aggregated by major areas of origin, to make the estimations as comparable as possible with the baseline specification in 1991-2001. However, this aggregation can mask important insights, as single country of origin can drive the result of the entire area. To further exploit the information of the data set, the most important supply countries are distinguished. It has been arbitrarily chosen to select the thirties most represented origin countries, which are, in descending order in terms of size, 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. They overall accounted for nearly 90 percent of the total migration stock in 2007. The empirical results for the alternative aggregation are reported in Table 12. Two findings are worth mentioning. First, the domain variable is robust to this alternative aggregation. The coefficient is positive and statistically significant. The second finding is related to the ethnic group variables. Among the Central-Eastern Europe block, foreigners from Macedonia, Bulgaria and Romania are responsible for increasing the level of diversity, whereas foreigners from Croatia, Ukraine and Russia exert an opposite effect. Three countries of Western Europe are among the top thirty supply countries for Italy, and among these, the foreigners from UK increase the level of diversity. From Eastern Asia, Chinese migrants increase the index of fractionalization whereas Philippines reduce it. None of the migrants from Central-Southern Asia exert a significant effect on the index. From Northern Africa, migrants from Tunisia augment

diversity, whereas migrants from Morocco and Algeria decrease diversity. From Western Africa, foreigners from Senegal are associated with a lower level of multiculturality. Finally, there is no significant effect from any migrants from Latin America. It should be noted however that the inclusion of the communication variable largely affects the significance of the coefficients of the different groups.

The use of diversity indexes that incorporate some measures of distance, such as the linguistic distance, is becoming popular in the empirical exercises, as these indexes add a further dimension to capture the degree of distinctiveness between different groups. For this reason, the Greenberg index (1959) described in equation (10) is estimated on the disaggregated data described above. It should be noted that the correlation between the standard diversity index that does not incorporate linguistic distance and the Greenberg index, that take into account the number of shared branches in a linguistic tree, is very high in this sample, being 0.87. This is because many languages spoken by the migrants in the sample are very distant to one another, having zero shared branches. This implies that for many pairs of languages the distance measure is equal to one. Given the high correlation, it comes with no surprise the very little change in the new estimations, reported in Table 13. Again the coefficient of the variable domain is positive and it is statistically significant at 10 percent level.

One last exercise can be computed, applying the 1991-2001 dataset. Given that the census collects individual information on respondents, including the foreign population living in Italy, this additional detail can be exploited to analyse an ethnic specific index of specialization. The use of an index that varies at ethnic level can inform about important ethnic specific habits. Some groups may exert a greater tendency to concentration for religious motivations, if for example they need to follow religious practise such as attending the Mosque. The index of specialization for each ethnic group s, measured by the Balassa index, is therefore used as dependent variable in the following specification:

$$SPEC_{rts} = \alpha + \boldsymbol{\beta}\, \mathbf{X}_{rts} + \delta\, technology_{rt} + \varepsilon_{rts} \qquad (14)$$

where $\varepsilon_{rts} = \mu_{rs} + \lambda_t + \nu_{rts}$ r=1, …., R; t=1, …, T; s= 1…, S

The vector X captures time-varying characteristics of the ethnic group s in province r and time t. In particular it includes the count of migrants from the different origin areas, a measure of the group economic status, captured by the proportion of occupied migrants, and different demographic variables, such as the proportion of migrants in the different age categories and the proportion of male migrants. The transportation variables described above are also included. The equation incorporates dummy fixed effects for origin and ethnic groups and a year dummy.

Before proceeding with the empirical results, the socio-economic characteristics of the foreign population are analysed. [8] Remarkable differences appear between the ethnic groups in the same year and partially for the same group between 1991 and 2001 (Tables 14, 15). For example, in 1991 the proportion of young migrants ranges from a minimum of 7.3 to a maximum of 19.5 percent, registered by the groups other Africa and Latin America, respectively. These figures suggest that while a certain proportion of movers from Latin America came in host countries as a family, emigration for Africans tended to be an individual strategy. A strong difference in terms of proportion of elderly is identified between migrants from developing countries and from developed ones, with the former displaying the lowest proportion in this age group. The elderly population among the groups Northern Africa, other African, Latin America and Asia are less than three percent, whereas from North America and Oceania reached 15.2 and 12.6 percent, respectively. This suggests that immigration from developing countries just started in 1991 and the first cohorts of movers were still in young age. Another possibility is that foreigners from developing countries have a different emigration strategy than foreigners from developed countries, with the latter being more permanent. On the contrary, foreigners from developing countries may display greater rates of return migration, as they move to destinations when they are young and return to origins in older age.

Emigration from African countries involves not only single and young individuals, but also selects primarily male ones. On the contrary, the large majority of migrants from Central-Eastern Europe and from Latin America are female. Regarding the economic status, some groups display very low levels of occupation. The low rates for migrants from Central-Eastern Europe, Latin America and North America can be somehow justified by the relatively high rates of movers in the non-working age. The low rate of the group from Oceania is more surprising. On the contrary, foreigners from Northern Africa, other Africa and Asia, report much greater occupation rates.

In 2001, minimal changes occurred compared to 1991 but some issues are worth mentioning. First, while in 1991 half of overall migrants were male, in 2001 the majority of movers were female. All groups, except for movers from Central-Eastern Europe, display a reduction in the proportion of male migrants, compared to 1991. Differently from the other groups and in agreement with 1991, migrants from Africa are disproportionately male. On average, the occupation rates among migrants increased from 1991 to 2001. All groups converge to similar rates, except for the groups other Africa and Asia, which display much larger occupation rates than the other groups. Regarding the distribution of foreigners in the age categories, an interesting feature is the large increase in the proportion of migrants younger than 14 years of age, in the group Northern Africa.

[8] For a problem of consistency between 1991 and 2001, the group Europe 12 is included in the other Europe group.

This specific group eventually benefited from the 1998 Italian law, which ensured the right of family reunification to migrants with a regular employment. Finally, the existence of disparities between developing and developed countries regarding the presence of elderly emerges as well in 2001. Given that this feature is valid also in 2001, a support in favour of the life-cycle emigration strategy emerges.

Turning to the estimation of equation [14], Table 16 displays the results. The Balassa index ranges from 0 to $+\infty$, with a value greater than one identifying a situation of specialization. This implies that a positive coefficient indicates a positive effect of the variable on specialization. The first column of the table includes only the year and the province fixed effects. The positive and statistically significant coefficient of the year dummy indicates an overall increase in ethnic specialization among the Italian provinces from 1991 to 2001. In columns (2), the immigrant group dummies are introduced. This exercise explains if specific ethnic groups increased (decreased) specialization between 1991 and 2001, relative to the base group, which is Oceania. According to the table, the groups other Africa and Asia tend to show a decreasing specialization, compared to immigrants from Oceania. It should be noted, that the descriptive analysis on the mean conducted in Table 5a revealed that immigrants from Oceania displayed the second largest index in 1991 and the largest in 2001.

A more complete analysis can be performed, introducing the socio-economic characteristics of the specific ethnic groups. Column (3) reveals that additional migrants from the same ethnic group augment the ethnic specialization. Among the socio-economic characteristics, only the gender variable has a significant effect. Increasing the proportion of male in the gender mix augments the ethnic specialization, on average and ceteris paribus. Regarding the key transportation variables, two out of three have a significant coefficient. An increase in the number of close by airports and in the number of cheaper flights, proxied by the number of passengers, decrease the ethnic specialization. This result is again in line with the prediction of the theoretical model. Finally, controlling for these additional covariates, a greater number of group dummies has statistically coefficients. Only the groups Latin America and other Europe display a likely specialization as the group from Oceania, whereas all other groups tend to be less specialized.

The estimations of equation (13) are conducted using a dependent variable 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.[9] 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 these alternative procedures for both the 1991, 2001 panel and the 2004-2007 panel are reported in Tables 17 and 18. 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 does not affect the coefficients of the transportation variables but it does influence the significance of the communication variable, which turns statistically insignificant.

6. Conclusion

The objective of this paper is to provide an empirical test to the theoretical predictions of the O-P model on the development of multicultural 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 a segregated to a more integrated environment. Improvement in distant communication boosts the generation of multicultural cities, as movers increasingly rely on an enlarged community for identity transmission, rather than on localized peer effects of segregated environment.

The degree of heterogeneity of cities 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. Multiculturalism is maximized by means of a rich mix and an even distribution of the different ethnic groups in the city. Moreover, the Balassa index is used to exploit the difference in specialization both at ethnic level and among provinces and it indicates the existence of a disproportional prevalence of some ethnic groups in specific provinces.

The empirical analysis is conducted using two different datasets. Data are drawn from the Italian Census for 1991 and 2001 and from a demographic assessment of the foreign population living in the different Italian provinces for 2004 till 2007. The data are aggregated at NUTS 3 level (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 ‘70s 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 found

[9] 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 estimates 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.

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 multicultural cities are disproportionately located in Northern and Central Italy. According to the Balassa Index, the groups other Africa and Asia display lower than unitary levels of specialization both in 1991 and 2001, indicating a less than proportional incidence of these foreigners in the specific province, compared to the average Italian incidence.

The empirical estimations provide support to the prediction of the model and reveal that both transportation and telecommunication improvements represent a valid explanation for the birth of cosmopolitan cities and are likely to reduce ethnic specialization in provinces. The two-way error component models for the fractionalization index report that a better access to the airports, lower costs of travelling by airplane and a greater diffusion of internet technologies increases city diversity. An additional airport located at close distance to the provinces increases the fractionalization index by 0.4 percent, on average and ceteris paribus. In agreement with the theoretical model, a reduction in the time for travelling increases the degree of multiculturality of the cities. Second, an increase in the volume of passengers by five percent, which corresponds to an increase of about 100’000 passengers, augments diversity by 0.01 percent. Third, increasing the number of internet domain registered by ten units over 10’000 inhabitants, the fractionalization index augments by 0.1 percent. The positive and significant coefficient of the internet domain is robust to a larger disaggregation of the ethnic groups for migrants, and to the incorporation of a measure of linguistic distance between the different groups.

The estimations that exploit both the ethnic and the province variation in specialization, using the Balassa index, reveal that the groups other Africa and Asia are the least specialized. Different socio-economic characteristics of the ethnic groups are introduced, but only the gender variable exerts a significant effect on specialization. Increasing the proportion of male in the gender mix augments the ethnic specialization, on average and ceteris paribus. Finally, better a access to airports and lower costs for travelling are correlated with specialization in the expected direction, again providing support to the predictions of the theoretical model.

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 transportation variables are robust to this alternative specification.

7. References

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

Alesina A. and E. La Ferrara (2005) "Ethnic Diversity and Economic Performance" Journal of Economic Literature, vol. 43, pp. 762-800

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

Balassa B. (1965) "Trade Liberalization and Revealed Comparative Advantage" Manchester School of Economics and Social Studies, vol. 33, pp. 99-123

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

Bellini E.; G.I.P.Ottaviano, D.Pinelli and G. Prarolo (2009) "Cultural Diversity and Economic Performance: Evidence from European Regions" Working Papers 2009.63, Fondazione Eni Enrico Mattei.

Bisin A.; E. Patacchini; T. Verdier and Y. Zenou (2006) "Bend It Like Beckham: Identity, Socialization and Assimilation" CEPR Discussion Papers 5662

Bossert W.; C. D'Ambrosio and E. La Ferrara (2008) "A Generalized Index of Fractionalization" Cahiers de recherche 2008-01" Universite de Montreal, Departement de sciences economiques

Collier P. (2001) "Implications of ethnic diversity" Economic Policy, vol. 16(32), pp. 127-166

Cutler D. M.; E. L. Glaeser and J. L. Vidgor (1999) "The Rise and Decline of the American Ghetto" Journal of political economy, 107(3) 455-506

Cutler D. M.; E. L. Glaeser and J. L. Vidgor (2008) "Is the Melting Pot Still Hot? Explaining the Resurgence of Immigrant Segregation" Review of Economics and statistics, 90(3), 478-497

Dawkins C. J. (2005) "Tiebout choice and residential segregation by race in US metropolitan areas, 1980-2000" Regional Science and Urban Economics, vol. 35(6), pp. 734-755

De Benedictis L. and M. Tamberi (2002) "A note on the Balassa Index of Revealed Comparative Advantage" *Working Papers 158*, Universita' Politecnica delle Marche

Desmet K.; I. Ortuno-Ortin and S. Weber (2009) "Linguistic Diversity and Redistribution" *Journal of the European Economic Association*, vol. 7(6), pp. 1291-1318

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

Easterly W. and R. Levine (1997) "Africa's Growth Tragedy: Policies and Ethnic Divisions" The Quarterly Journal of Economics, vol. 112(4), pages 1203-50

Glaeser E. L.; H. D. Kallal; J. A. Scheinkman and A. Shleifer (1992) "Growth in Cities" Journal of Political Economy, vol. 100(6), pp. 1126-52

Gobbo F. (2007) "Infrastrutture e sviluppo: alcune riflessioni sulla situazione italiana" *Economia Pubblica*, vol. 5-6, pp. 5-28

Goldsmith R. W. (1951) "A Perpetual Inventory Method of National Wealth" NBER Chapter in Studies in *Income and Wealth*, vol. 14, pp. 5-74, National Bureau of Economic Research

Greenberg J. H. (1956) "The Measurement of linguistic diversity" *Language*, vol.32, pp.109-115

Ihlanfeldt K. R. and B. Scafidi (2002) "Black Self-Segregation as a Cause of Housing Segregation: Evidence from the Multi-City Study of Urban Inequality" Journal of Urban Economics, vol. 51(2), pages 366-390

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

Moulton B. R. (1986) "Random group effects and the precision of regression estimates" *Journal of Econometrics*, vol. XXXII(1), pp.385-397

OECD (2003) Communications Outlook 2003, OECD, Paris

Ottaviano G. I. P. and G. Peri (2005) "Cities and cultures" Journal of Urban Economics, vol. 58(2), pages 304-337

Ottaviano G. I. P. and G. Peri (2006) "The economic value of cultural diversity: evidence from US cities" *Journal of Economic Geography*, vol. 6(1), pp. 9-44

Ottaviano G. I. P. and G. Prarolo (2009) "Cultural Identity And Knowledge Creation In Cosmopolitan Cities" Journal of Regional Science, vol. 49(4), pp 647-662

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

Patterson O. (1997) *The ordeal of integration: Progress and Resentment in America's Racial Crisis*, Basic Civitas, New York

Picci L. (2002) "Le infrastrutture in Italia. Le differenze territoriali e l'efficienza della spesa", in M. Baldassarri, G. Galli and G. Piga (eds) *L'Italia nella competizione globale - Regole per il mercato*, Edizioni il Sole 24 Ore, Milan

Quigley, J. M.; S. Raphael and L. A. Rosenthal (2004) "Local Land-use Controls and Demographic Outcomes in a Booming Economy" Urban Studies, vol. 41, pp. 389 – 421

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

Rhode P. W. and K. S. Strumpf (2003) "Assessing the Importance of Tiebout Sorting: Local Heterogeneity from 1850 to 1990" American Economic Review, vol. 93(5), pp 1648-1677

Shelling T. (1972) "A Process of Residential Segregation: Neighbourhood Tipping" in A. H. Pascal (eds) Racial Discrimination in Economic Life, Lexington Books

Thernstrom S. and A. Thernstrom (1997) *America in Black and White: One Nation, Indivisible*, Touchstone, New York

Appendix

Table A1: Country classification by main area of origin. 1991, 2001

Origin Group	Countries
Europe 12	Belgium, Denmark, France, Germany, Greece, Ireland, Luxembourg, Netherlands, Portugal, Spain, United Kingdom
Central Eastern Europe	Albania, Belarus, Bosnia-Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Macedonia, Moldova, Poland, , Romania, Russia, Serbia, Montenegro, Slovakia, Slovenia, Turkey and Ukraine
other Europe	All other European countries
Northern Africa	Algeria, Egypt, Libya, Morocco, Sudan, Tunisia,
other Africa	All other African countries
North America	Canada, United States
Latin America	All other American countries
Oceania	All countries
Asia	All countries

Table A2: Country classification by main area of origin. 2004-2007

Origin Group	Countries
EU	Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom
Central Eastern Europe	Albania, Belarus, Bosnia-Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Macedonia, Moldova, Poland, , Romania, Russia, Serbia, Montenegro, Slovakia, Slovenia, Turkey and Ukraine
other Europe	All other European countries
Northern Africa	Algeria, Egypt, Libya, Morocco, Sudan and Tunisia
Eastern Africa	Burundi, Comoros, Djibouti, Eritrea, Ethiopia, Kenya, Madagascar, Malawi, Mauritius, Mozambique, Rwanda, Seychelles, Somalia, Tanzania, Uganda, Zambia, Zimbabwe
Western Africa	Benin, Burkina Faso, Cape Verde, Cote d'Ivoire, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Mauritania, Niger, Nigeria, Senegal, Sierra Leone, Togo
Central-Southern Africa	Angola, Botswana, Cameroon, Central African Rep., Chad, Congo, Congo Dem. Rep., Equatorial Guinea , Gabon, Lesotho, Namibia, South Africa, Sao Tome and Principe, Swaziland
North America	Canada, United States
Latin America	All other American countries
Eastern Asia	Brunei Darussalam, Cambodia, China, Korea Dem. Rep., Korea Rep., Indonesia, Japan, Lao, Malaysia, Mongolia, Myanmar, Philippines, Singapore, Taiwan, Thailand, Timor-Leste, Vietnam
Western Asia	Armenia, Azerbaijan, Bahrain, Georgia, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Syria, United Arab Emirates, Yemen
Central-Southern Asia	Afghanistan, Bangladesh, Bhutan, India, Kazakhstan, Kyrgyz Republic, Maldives, Nepal, Pakistan, Sri Lanka, Tajikistan, Turkmenistan, Uzbekistan
Oceania	All countries

Table A3: Description of transportation and telecommunication variables

Variable	Description
Motorway	Total length of railway lines (Km). Source: EUROSTAT**,** Regional statistics (REGIO). NUTS2
Railroad	Total length of motorways (Km). Source: EUROSTAT**,** Regional statistics (REGIO). NUTS2
Airport Distance	Euclidean distance to closest airport (Km), computed using latitude and longitude of main town of each NUTS3 and the airport. The list of italian airports is taken from "Conto Nazionale dei Trasporti" (2003), Ministero dei Trasporti and "Statistiche dei Trasporti-2001", (2003) ISTAT. NUTS3
Airports Number	Number of airports located at less than 100 Km. Source: "Conto Nazionale dei Trasporti" (2003). NUTS3
Passengers	Number of passengers landing and departing in international flights in closest main airport. Source: "Conto Nazionale dei Trasporti" (2003), Ministero dei Trasporti and "Statistiche dei Trasporti-2001", (2003) ISTAT. NUTS3
Domain	Number of domains registered over 10000 inhabitants. Source: Institute for Informatics and Telematics, Italian National Research Council. (NUTS 2)

Table A4: Description of variables and source

Variable	1991; 2001	2004-2007
Migrants by area of origin	Source: Census, ISTAT	Source: register office
ln(population)	Source: Census, ISTAT	Source: register office
Less 14	Proportion of people below 14 years of age. Source: Census, ISTAT	Proportion of people below 14 years of age. Source: register office
15-64	Proportion of people between 15 and 64 years of age. Source: Census, ISTAT	Proportion of people between 15 and 64 years of age. Source: register office
Over 65	Proportion of people over 65 years of age. Source: Census, ISTAT	Proportion of people over 65 years of age. Source: Source: register office
No education	Proportion of people in schooling age who obtained no education. Source: Census, ISTAT	-
Primary Education	Proportion of people in schooling age who obtained primary education. Source: Census, ISTAT	Economically active population with pre-primary, primary and lower secondary education. Source: EUROSTAT
Secondary Education	Proportion of people in schooling age who obtained secondary education. Source: Census, ISTAT	Economically active population with upper secondary and post-secondary non-tertiary education. Source: EUROSTAT
Tertiary Education	Proportion of people in schooling age who obtained tertiary education. Source: Census, ISTAT	Economically active population with tertiary education. Source: EUROSTAT
Unemployment Rate	Source: Census, ISTAT	Source: Labour Force Survey, ISTAT
Activity Rate	Source: Census, ISTAT	Source: Labour Force Survey, ISTAT

Table 1a: The distribution of foreigners (%)

	1991	2001
TOTAL	0.5	2.1
North-Centre	0.7	3.0
South	0.3	0.9

Table 1b: The distribution of foreigners (%)

	2004	2005	2006	2007
TOTAL	3.9	4.3	4.7	5.5
North-Centre	5.1	5.7	6.2	7.2
South	1.5	1.6	1.8	2.3

2a: The average index of fractionalization

	1991	2001
TOTAL	0.011	0.043
North-Centre	0.013	0.057
South	0.007	0.020

Table 2b: The average index of fractionalization

	2004	2005	2006	2007
TOTAL	0.075	0.083	0.090	0.105
North-Centre	0.099	0.110	0.119	0.137
South	0.029	0.032	0.035	0.044

Table 3: Average annual percentage change in the fractionalization index (%)

Province	1991-2001	Location	Province	2004-2007	Location
Terni	76	North-Central	Sassari	47	South
Rovigo	63	North-Central	Nuoro	37	South
Piacenza	63	North-Central	Cagliari	34	South
Treviso	60	North-Central	Cosenza	31	South
Pordenone	60	North-Central	Caltanissetta	30	South
La Spezia	54	North-Central	Siracusa	28	South
Asti	52	North-Central	Enna	27	South
Cuneo	50	North-Central	Vibo Valentia	27	South
Cremona	50	North-Central	Campobasso	25	South
Ancona	49	North-Central	Potenza	24	South
Macerata	49	North-Central	Chieti	23	South
Ascoli Piceno	48	North-Central	Agrigento	22	South
Matera	48	South	Latina	21	North-Central
Belluno	47	North-Central	Ferrara	21	North-Central
Pesaro e Urbino	47	North-Central	Rieti	20	North-Central
Pavia	47	North-Central	Frosinone	19	North-Central
Mantua	46	North-Central	Viterbo	19	North-Central
Alessandria	45	North-Central	Pavia	19	North-Central
Venice	44	North-Central	Pescara	19	South
Verona	43	North-Central	Matera	18	South

Table 4a Macro area of origin: 1991, 2001, (%)

	1991	2001
Europe 12	21.0	10.8
Central Eastern Europe	5.8	30.9
Other Europe	14.0	2.0
Northern Africa	19.4	20.2
Other Africa	10.3	9.0
North America	4.9	1.6
Latin America	8.8	9.2
Asia	15.2	16.2
Oceania	0.6	0.3

Table 4b Macro area of origin: 2004, 2005, 2006, 2007, (%)

	2004	2005	2006	2007
EU	6.68	6.26	5.99	5.23
Central Eastern Europe	44.18	45.00	45.65	51.77
Other Europe	0.73	0.67	0.63	0.52
Northern Africa	20.79	20.33	19.93	17.41
Central Southern Africa	0.44	0.45	0.44	0.40
Western Africa	5.13	4.97	4.91	4.46
Eastern Africa	1.06	1.11	1.12	1.01
North America	0.81	0.73	0.69	0.56
Latin America	6.89	6.82	6.70	5.92
Central Southern Asia	5.47	5.61	5.69	5.33
Western Asia	0.72	0.68	0.64	0.56
Eastern Asia	6.92	7.22	7.48	6.75
Oceania	0.16	0.15	0.13	0.10

Table 5a: Balassa Specialization Index, mean: 1991-2001

	1991	2001
Europe 12	1.04	1.03
Central Eastern Europe	0.91	1.16
Other Europe	1.26	1.04
Northern Africa	1.15	1.10
Other Africa	0.80	0.80
North America	1.12	1.35
Latin America	1.05	0.87
Asia	0.61	0.69
Oceania	1.20	1.56

Table 5b: Balassa Specialization Index, mean: 2004, 2005, 2006, 2007

	2004	2005	2006	2007
EU	1.16	1.17	1.18	1.14
Central Eastern Europe	1.09	1.09	1.09	1.10
Other Europe	1.33	1.34	1.39	1.33
Northern Africa	1.12	1.12	1.12	1.08
Central Southern Africa	0.86	0.87	0.85	0.84
Western Africa	0.81	0.81	0.82	0.82
Eastern Africa	0.84	0.89	0.89	0.86
North America	1.18	1.16	1.18	1.10
Latin America	0.78	0.76	0.75	0.74
Central Southern Asia	0.76	0.75	0.74	0.74
Western Asia	0.85	0.86	0.83	0.83
Eastern Asia	0.79	0.82	0.83	0.82
Oceania	1.55	1.56	1.52	1.34

Table 6a: Balassa Specialization Index, standard deviation: 1991-2001

	1991	2001
Europe 12	0.47	0.61
Central Eastern Europe	1.02	0.47
Other Europe	0.79	0.84
Northern Africa	0.71	0.58
Other Africa	0.68	0.61
North America	0.89	1.28
Latin America	0.50	0.51
Asia	0.41	0.45
Oceania	1.57	1.94

Table 6b: Balassa Specialization Index, standard deviation: 2004, 2005, 2006, 2007

	2004	2005	2006	2007
EU	0.77	0.79	0.81	0.78
Central Eastern Europe	0.35	0.34	0.33	0.27
Other Europe	1.44	1.46	1.54	1.45
Northern Africa	0.60	0.60	0.59	0.54
Central Southern Africa	0.77	0.75	0.73	0.74
Western Africa	0.70	0.69	0.69	0.70
Eastern Africa	1.72	1.85	1.74	1.59
North America	0.93	0.89	0.91	0.84
Latin America	0.59	0.58	0.58	0.61
Central Southern Asia	0.74	0.72	0.70	0.71
Western Asia	0.65	0.62	0.61	0.60
Eastern Asia	0.61	0.63	0.62	0.64
Oceania	1.88	1.96	1.97	1.55

Table 7a: Balassa Specialization Index, min and max: 1991-2001

	1991		2001	
	Min	**Max**	**Min**	**Max**
Europe 12	0.31	3.01	0.22	3.79
Central Eastern Europe	0.18	8.70	0.41	2.32
Other Europe	0.19	4.74	0.08	5.46
Northern Africa	0.06	3.92	0.07	3.47
Other Africa	0.03	2.96	0.06	2.63
North America	0.15	4.28	0.12	5.99
Latin America	0.24	2.67	0.22	3.96
Asia	0.05	2.18	0.10	2.19
Oceania	0.00	11.03	0.13	13.35

Table 7b: Balassa Specialization Index, min and max: 2004, 2005, 2006, 2007

	2004		2005		2006		2007	
	Min	**Max**	**Min**	**Max**	**Min**	**Max**	**Min**	**Max**
EU	0.26	4.45	0.26	4.44	0.26	4.43	0.29	4.63
Central Eastern Europe	0.30	1.83	0.30	1.78	0.30	1.76	0.44	1.59
Other Europe	0.12	7.71	0.00	7.53	0.06	8.67	0.06	7.95
Northern Africa	0.10	3.75	0.10	3.83	0.11	3.76	0.12	3.25
Central Southern Africa	0.09	4.11	0.06	3.90	0.06	3.57	0.06	3.45
Western Africa	0.07	3.23	0.07	3.20	0.10	3.25	0.08	3.40
Eastern Africa	0.03	16.25	0.04	17.42	0.05	15.75	0.05	13.15
North America	0.14	4.23	0.14	4.30	0.14	4.35	0.14	4.30
Latin America	0.21	4.96	0.22	4.87	0.21	4.87	0.19	5.17
Central Southern Asia	0.05	4.17	0.04	3.96	0.06	3.80	0.06	3.69
Western Asia	0.00	2.98	0.00	3.13	0.00	3.29	0.00	3.01
Eastern Asia	0.18	4.54	0.18	4.77	0.21	4.82	0.18	5.12
Oceania	0.00	10.08	0.00	10.30	0.00	10.67	0.00	10.34

Table 8: Improvements in transportation

	1991	**2001**	**Annual percentage change (%)**
Motorway (Km)	394	413	0.48
Railroad (Km)	1012	1015	0.03
Airport Distance (Km)	65	43	3.38
Airports Number	1.53	2.29	4.97
Passengers	1768493	3914621	12.14

Table 9: Improvements in telecommunication

	2004	**2005**	**2006**	**2007**	**Annual percentage change (%)**
Domain	119.635			200.60	23

Table 10: Two-way Error Component Model. Census data

	Dependent Variable: Index of fractionalization *100							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Year 2001	3.229***	4.359***	4.627***	4.234***	4.794***	4.361***	4.532***	4.572***
	[0.437]	[1.017]	[1.094]	[1.081]	[0.998]	[0.952]	[0.956]	[0.962]
Population		0.000002	0.000001	0.000002	-0.0000002	0.000002	-0.000001	-0.000001
		[0.00001]	[0.00001]	[0.00001]	[0.00001]	[0.000006]	[0.00001]	[0.000005]
Less 14		71.34***	73.46***	72.58***	72.33***	60.64***	73.61***	63.23***
		[14.15]	[13.10]	[13.15]	[15.30]	[12.77]	[15.13]	[14.08]
15_64		40.13**	41.55**	40.00**	38.24**	36.90**	44.38**	41.28***
		[16.33]	[15.18]	[16.35]	[16.91]	[13.24]	[18.30]	[14.40]
Primary Education		18.09*	17.27	18.77*	11.29	11.1	18.37*	11.02
		[10.18]	[10.09]	[9.480]	[9.371]	[7.677]	[10.19]	[7.988]
Secondary Education		11.55	11.13	12.95	8.19	3.176	10.26	1.7
		[10.50]	[10.56]	[10.33]	[9.572]	[8.354]	[10.67]	[8.902]
Tertiary Education		35.82*	29.8	37.93**	10.67	22.61	33.39*	18.54
		[18.31]	[18.15]	[18.12]	[18.81]	[16.60]	[16.80]	[19.67]
Unemployment		-0.162***	-0.162***	-0.154***	-0.128**	-0.137***	-0.161***	-0.133***
Rate		[0.042]	[0.041]	[0.0446]	[0.0452]	[0.041]	[0.041]	[0.042]
Activity Rate		17.20***	17.26***	17.66***	19.05***	18.28***	16.70***	17.86***
		[4.976]	[4.906]	[4.983]	[5.163]	[4.681]	[4.772]	[4.404]
Motorway			-0.002					
			[0.004]					
Rail				0.002				
				[0.003]				
Airport					-0.005**			-0.0003
Distance					[0.002]			[0.002]
Airports Number						0.390***		0.385***
						[0.069]		[0.098]
Passengers Number							0.0000001	0.00000006**
							0.00000004	0.00000002

Cont.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Europe 12		-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0004***	-0.0003***	-0.0004***
		[0.0001]	[0.0001]	[0.0001]	[0.0001]	[0.0001]	[0.0001]	[0.0001]
Central-Eastern		0.0001***	0.0001***	0.0001***	0.0001***	0.0001***	0.0001***	0.0001***
Europe		[0.00003]	[0.00003]	[0.00003]	[0.00003]	[0.00003]	[0.00003]	[0.00003]
Other Europe		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
		[0.0002]	[0.0002]	[0.0002]	[0.0002]	[0.0002]	[0.0002]	[0.0002]
Northern Africa		0.0001*	0.0001*	0.0001*	0.0001	0.00003	0.0001	0.00001
		[0.00005]	[0.00005]	[0.00005]	[0.00005]	[0.0001]	[0.00005]	[0.0001]
Other Africa		0.0002	0.0002	0.0002	0.0002	0.0002	0.0003	0.0003
		[0.0002]	[0.0002]	[0.0002]	[0.0002]	[0.0002]	[0.0002]	[0.0002]
North America		0.001	0.001	0.001	0.001	0.001	0.001	0.001
		[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Latin America		-0.0001	-0.0001	-0.0001	-0.0001*	-0.00005	-0.0001	-0.0001
		[0.0001]	[0.0001]	[0.0001]	[0.0001]	[0.0001]	[0.0001]	[0.0001]
Asia		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
		[0.0001]	[0.0001]	[0.0001]	[0.0001]	[0.0001]	[0.0001]	[0.0001]
Oceania		0.0045	0.004	0.004	0.002	0.006	0.005	0.006
		[0.006]	[0.006]	[0.006]	[0.007]	[0.007]	[0.007]	[0.007]
Constant	1.066***	-60.59***	-59.53***	-63.77***	-53.16***	-51.49***	-61.96***	-52.55***
	[0.218]	[16.03]	[16.40]	[13.85]	[17.45]	[14.18]	[18.02]	[16.41]
Observations	190	190	190	190	190	190	190	190
Number of code	95	95	95	95	95	95	95	95
R-squared	0.729	0.968	0.969	0.969	0.971	0.975	0.97	0.976

Notes: robust standard errors in parenthesis. Standard Errors adjusted for 20 clusters. * significant at 10%; ** significant at 5%; *** significant at 1%. The excluded variables are the 1991 year dummy, the proportion of people over 65 years of age and the proportion of people with no education

Table 11: Two-way Error Component Model. Register Office

Dependent Variable: Index of fractionalization *100

	(1)	(2)	(3)
Year 2005	0.781***	0.965***	
	[0.108]	[0.0779]	
Year 2006	1.508***	1.890***	
	[0.199]	[0.154]	
Year 2007	3.001***	3.540***	2.526***
	[0.307]	[0.242]	[0.524]
Population		-0.00000542***	-0.00000652***
		[0.000000856]	[0.00000148]
Less 14		2.580***	2.445***
		[0.225]	[0.260]
15_64		0.893***	0.970***
		[0.244]	[0.233]
Secondary Education		-0.0451	-0.0523
		[0.0285]	[0.0607]
Tertiary Education		0.000537	-0.0447
		[0.0666]	[0.0826]
Unemployment Rate		4.173*	3.196
		[1.999]	[3.754]
Activity Rate		-0.0202	-2.376
		[2.050]	[3.793]
Domain			0.0144**
			[0.00589]
Central-Southern Africa		-0.00182	-0.00333
		[0.00264]	[0.00301]
Western Africa		0.000320*	0.0000767
		[0.000164]	[0.000217]
Eastern Africa		-0.00054	0.0000119
		[0.000408]	[0.000441]
Northern Africa		0.0000723	0.000224
		[0.000126]	[0.000145]
Central-Eastern Europe		0.0000387**	0.0000477*
		[0.0000157]	[0.0000235]
EU		0.000601*	0.000277
		[0.000301]	[0.000372]
Other Europe		0.00531	0.00548
		[0.00393]	[0.00576]
Latin America		-0.000173	-0.000246*
		[0.000103]	[0.000133]
North America		0.00495*	0.00669
		[0.00273]	[0.00408]
Central-Southern Asia		-0.0000536	-0.00011
		[0.0000775]	[0.0000694]
Western Asia		-0.0011	-0.00199
		[0.00107]	[0.00148]
Eastern Asia		0.00016	0.000166
		[0.000172]	[0.000162]
Oceania		-0.00964	-0.02
		[0.00819]	[0.0129]

Cont.

Cont.

	(1)	(2)	(3)
Constant	7.492***	-85.47***	-87.33***
	[0.153]	[17.64]	[17.39]
Observations	412	412	206
Number of codice	103	103	103
R-squared	0.818	0.957	0.971

Notes: robust standard errors in parenthesis. Standard Errors adjusted for 20 clusters. * significant at 10%; ** significant at 5%; *** significant at 1%. The least recent year dummy is always the one removed from regression. The other excluded variables are the proportion of people over 65 years of age and the proportion of people with primary education.

Table 12: Two-way Error Component Model. Register Office. Alternative aggregation

Dependent Variable: Index of fractionalization *100

	(1)	(2)	(3)
Year 2005	0.791***	0.980***	
	[0.109]	[0.0890]	
Year 2006	1.527***	1.893***	
	[0.202]	[0.185]	
Year 2007	3.049***	3.542***	2.147***
	[0.314]	[0.306]	[0.404]
Population		-0.00000649***	-0.00000975***
		[0.00000111]	[0.00000171]
Less 14		2.443***	1.910***
		[0.304]	[0.280]
15_64		0.681*	0.381
		[0.327]	[0.270]
Secondary Education		-0.0958**	-0.182***
		[0.0399]	[0.0625]
Tertiary Education		-0.047	-0.147*
		[0.0524]	[0.0754]
Unemployment Rate		4.393**	6.115
		[1.828]	[4.219]
Activity Rate		-1.152	-5.068*
		[2.181]	[2.825]
Domain			0.0157***
			[0.00394]
Albania		0.000124	0.000119
		[0.000118]	[0.000117]
Serbia and Montenegro		-0.000199	-0.000579
		[0.000247]	[0.000345]
Bosnia and Herzegovina		0.000735	0.00195
		[0.000807]	[0.00128]
Macedonia, FYR		0.000363	0.000937**
		[0.000297]	[0.000363]
Croatia		-0.00126	-0.00373*
		[0.000863]	[0.00179]
Bulgaria		0.000438	0.00173***
		[0.000481]	[0.000267]
Romania		0.0000312*	0.0000640*
		[0.0000154]	[0.0000328]
Poland		-0.000212	0.000158
		[0.000311]	[0.000317]
Ukraine		-0.000192	-0.000431**
		[0.000161]	[0.000205]
Moldova		-0.0000095	0.000032
		[0.000244]	[0.000374]
Russia		-0.000477	-0.00837***
		[0.00202]	[0.00221]
France		-0.00168	0.00189
		[0.00158]	[0.00141]
Germany		0.00195***	0.000568
		[0.000607]	[0.000773]
UK		0.00214**	0.00276***
		[0.000904]	[0.000927]

Cont.

Cont.

	(1)	(2)	(3)
Philippines		0.00028	-0.000905*
		[0.000236]	[0.000520]
China		0.000264	0.000293*
		[0.000191]	[0.000163]
Bangladesh		-0.000449	-0.000396
		[0.000270]	[0.000318]
Sri Lanka		-0.000565**	-0.000288
		[0.000243]	[0.000434]
India		0.000332**	0.000225
		[0.000142]	[0.000195]
Pakistan		-0.000285	-0.000318
		[0.000282]	[0.000376]
Algeria		-0.000888	-0.00270*
		[0.00166]	[0.00143]
Egypt, Arab Rep.		0.000572**	0.000508
		[0.000241]	[0.000383]
Morocco		-0.000145	-0.000378**
		[0.000139]	[0.000174]
Tunisia		0.000161	0.000949*
		[0.000351]	[0.000477]
Ghana		-0.0000991	0.00000891
		[0.000859]	[0.000693]
Nigeria		0.00149*	0.00149
		[0.000738]	[0.00105]
Senegal		-0.000646	-0.00167*
		[0.000613]	[0.000855]
Brazil		0.000959**	0.000911
		[0.000348]	[0.000810]
Ecuador		-0.0000826	0.000126
		[0.000130]	[0.000132]
Peru		-0.000800**	-0.000742
		[0.000358]	[0.000639]
Other		0.0000859	0.000494*
		[0.000161]	[0.000237]
Constant	7.527***	-63.83**	-28.48
	[0.155]	[24.66]	[23.54]
Observations	412	412	206
Number of codes	103	103	103
R-squared	0.816	0.963	0.985

Notes: robust standard errors in parenthesis. Standard Errors adjusted for 20 clusters. * significant at 10%; ** significant at 5%; *** significant at 1%. The least recent year dummy is always the one removed from regression. The other excluded variables are the proportion of people over 65 years of age and the proportion of people with primary education.

Table 13: Two-way Error Component Model. Register Office. Greenberg index

Dependent Variable: Greenberg index *100

	(1)	(2)	(3)
Year 2005	0.252***	0.233***	
	[0.0347]	[0.0249]	
Year 2006	0.492***	0.449***	
	[0.0665]	[0.0536]	
Year 2007	0.738***	0.730***	0.645***
	[0.0950]	[0.0802]	[0.142]
Population		-0.00000231***	-0.00000229***
		[0.000000304]	[0.000000519]
Less 14		0.588***	0.523***
		[0.0979]	[0.0990]
15_64		0.0533	0.0145
		[0.103]	[0.108]
Secondary Education		0.000386	-0.0285
		[0.0134]	[0.0255]
Tertiary Education		-0.00501	-0.0527**
		[0.0133]	[0.0228]
Unemployment Rate		1.712***	2.342**
		[0.490]	[1.036]
Activity Rate		-0.602	-1.799
		[0.607]	[1.158]
Domain			0.00237*
			[0.00121]
Albania		-0.0000119	-0.0000533
		[0.0000315]	[0.0000459]
Serbia and Montenegro		-0.000222**	-0.000305*
		[0.0000934]	[0.000169]
Bosnia and Herzegovina		0.000347	0.000952*
		[0.000354]	[0.000472]
Macedonia, FYR		-0.000153	-0.0000671
		[0.000108]	[0.000127]
Croatia		-0.00105*	-0.00292***
		[0.000591]	[0.000599]
Bulgaria		0.0000578	0.000231**
		[0.000105]	[0.0000991]
Romania		-0.0000126***	-0.0000269*
		[0.00000381]	[0.0000152]
Poland		-0.0000399	-0.000074
		[0.0000724]	[0.0000879]
Ukraine		-0.000106***	-0.000156***
		[0.0000367]	[0.0000506]
Moldova		-0.000204***	-0.000233**
		[0.0000692]	[0.000103]
Russia		-0.00103**	-0.00167**
		[0.000447]	[0.000764]
France		0.000301	0.0000933
		[0.000340]	[0.000502]
Germany		0.000393***	0.000477**
		[0.000120]	[0.000169]
UK		0.0000496	-0.000207
		[0.000217]	[0.000338]

Cont.

Cont.

	(1)	(2)	(3)
Philippines		0.000202***	0.000280***
		[0.0000578]	[0.0000815]
China		0.000500***	0.000505***
		[0.000110]	[0.0000957]
Bangladesh		-0.000190*	-0.000224*
		[0.000109]	[0.000117]
Sri Lanka		-0.000211**	-0.000205
		[0.0000934]	[0.000170]
India		0.0000531	0.0000716
		[0.0000496]	[0.0000550]
Pakistan		0.0000629	-0.00000113
		[0.000104]	[0.0000907]
Algeria		0.000831*	0.000789
		[0.000455]	[0.000552]
Egypt, Arab Rep.		-0.00000686	-0.00012
		[0.000134]	[0.000208]
Morocco		0.0000772**	0.000150**
		[0.0000364]	[0.0000631]
Tunisia		0.000454***	0.000600***
		[0.000131]	[0.000181]
Ghana		-0.000396	-0.000629
		[0.000251]	[0.000366]
Nigeria		0.000698**	0.000925**
		[0.000292]	[0.000353]
Senegal		0.0000269	-0.000129
		[0.000183]	[0.000213]
Brazil		0.000106	0.000184
		[0.000189]	[0.000422]
Ecuador		-0.0000228	-0.0000882**
		[0.0000462]	[0.0000414]
Peru		-0.000350***	-0.000117
		[0.000117]	[0.000245]
Constant	2.656***	-8.117	-2.161
	[0.0488]	[8.055]	[9.341]
Observations	412	412	206
Number of codice	103	103	103
R-squared	0.686	0.953	0.97

Notes: robust standard errors in parenthesis. Standard Errors adjusted for 20 clusters. * significant at 10%; ** significant at 5%; *** significant at 1%. The least recent year dummy is always the one removed from regression. The other excluded variables are the proportion of people over 65 years of age and the proportion of people with primary education.

Table 14: Summary statistics of the demographic characteristics of ethnic groups, 1991. (%)

1991	Central-Eastern Europe	Other Europe	Northern Africa	Other Africa	Latin America	North America	Asia	Oceania	Groups Mean
Less 14	17.7	14.0	9.7	7.3	19.5	16.9	10.7	8.5	13.0
15_64	77.0	78.5	88.9	91.1	77.6	67.8	87.5	78.9	80.9
Over 65	5.3	7.5	1.4	1.6	3.0	15.2	1.8	12.6	6.1
Male	30.2	45.9	78.8	63.8	33.4	45.3	50.5	45.4	49.2
Occupied	27.7	37.0	52.5	49.5	29.4	27.7	45.0	28.9	37.2

Notes: The proportions are calculated for each group of foreigners. To receive 100% one should add vertically the proportions within each group of variables.

Table 15: Summary statistics of the demographic characteristics of ethnic groups, 2001. (%)

2001	Central-Eastern Europe	Other Europe	Northern Africa	Other Africa	Latin America	North America	Asia	Oceania	Groups Mean
Less 14	20.3	4.5	23.1	12.9	9.8	5.7	18.5	2.6	12.2
15_64	77.0	83.2	74.9	85.5	86.7	75.8	80.0	87.3	81.3
Over p65	2.7	12.4	1.9	1.6	3.6	18.5	1.4	10.1	6.5
Male	46.2	33.6	62.8	55.4	28.7	41.4	49.3	39.6	44.6
Occupied	40.5	40.6	44.3	53.2	41.0	40.1	49.1	50.9	45.0

Notes: The proportions are calculated for each group of foreigners. To receive 100% one should add vertically the proportions within each group of variables.

Table 16: Two-way error Component Model on Balassa Index. Census data

Dependent Variable: Balassa Index

	(2)	(2)	(3)
Year 2001	0.091**	0.091**	0.084
	[0.042]	[0.042]	[0.073]
Foreigner Number			0.0001**
			[0.00003]
Less 14			0.301
			[0.806]
15_64			-0.091
			[0.696]
Male			0.801***
			[0.221]
Occupied			0.171
			[0.258]
Passengers Number			-0.00000003*
			0.00000002
Airports Number			-0.002**
			[0.001]
Airport Distance			0.0002
			[0.001]
Central-Eastern Europe		-0.379	-0.580*
		[0.261]	[0.308]
Other Europe		-0.216	-0.333
		[0.255]	[0.263]
Northern Africa		-0.261	-0.688**
		[0.234]	[0.243]
Other Africa		-0.589*	-0.837**
		[0.310]	[0.326]
North America		-0.136	-0.196**
		[0.0892]	[0.0830]
Latin America		-0.424	-0.434
		[0.254]	[0.293]
Asia		-0.731**	-0.955***
		[0.270]	[0.296]
Constant	0.849***	0.601***	-0.191
	[0.021]	[0.145]	[0.684]
Observations	1520	1520	1517
R-squared	0.102	0.16	0.207
province FE	Yes	yes	yes
group FE	No	yes	yes

Notes: robust standard errors in parenthesis. Standard Errors adjusted for 20 clusters. * significant at 10%; ** significant at 5%; *** significant at 1%. The excluded variables are the 1991 year dummy, the ethnic group dummy for Oceania, the proportion of people over 65 years of age and the proportion of female.

Table 17: Alternative estimation. Census data

	Logit Transformation
Year 2001	0.819*
	[0.471]
Population	-0.000000341
	[0.00000177]
Less 14	-0.511
	[5.718]
15_64	-0.305
	[4.898]
Primary Education	0.828
	[2.724]
Secondary Education	8.373
	[5.577]
Tertiary Education	-0.694
	[8.680]
Unemployment Rate	-0.0461***
	[0.0148]
Activity Rate	2.118
	[1.293]
Airport Distance	-0.0004
	[0.000611]
Airports Number	0.0582**
	[0.0251]
Passengers Number	0.0000000168*
	[0.00000001]
Europe 12	-0.000141**
	[0.0000535]
Central-Eastern Europe	0.0000430***
	[0.0000144]
Other Europe	0.000265**
	[0.0000963]
Northern Africa	0.0000122
	[0.0000194]
Other Africa	-0.0000343
	[0.0000552]
North America	0.000799***
	[0.000275]
Latin America	0.00000782
	[0.0000231]
Asia	0.0000133
	[0.0000279]
Oceania	0.00125
	[0.00196]
Constant	-7.161
	[4.529]
Observations	190
Number of code	95
R-squared	0.977

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. The Two-way Error Component Model is applied to the logit transformation of the fractionalization index.

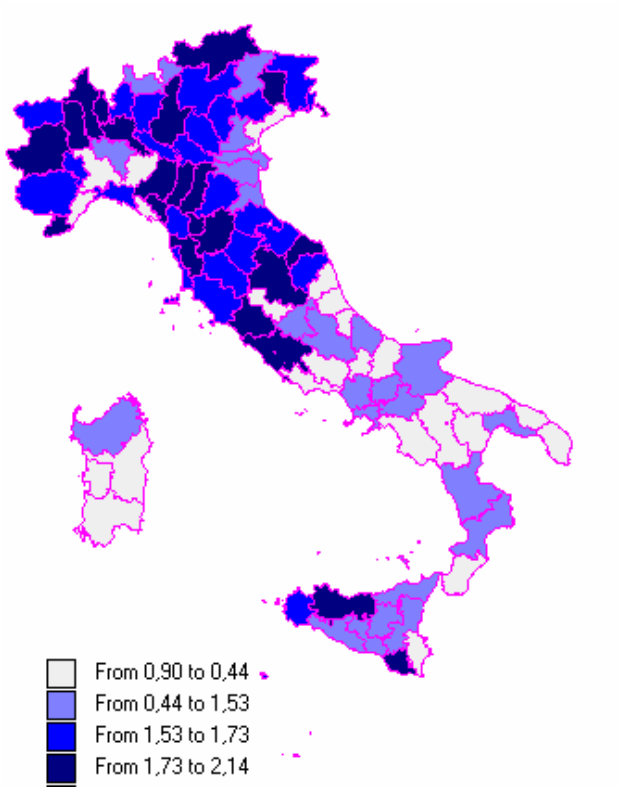
Table 18: Alternative estimation. Register Office

	Logit Transformation
Year 2007	0.373***
	[0.0567]
Population	-0.00000256***
	[0.000000657]
Less 14	-0.0356
	[0.0345]
15_64	0.00816
	[0.0308]
Secondary Education	-0.0117
	[0.00846]
Tertiary Education	-0.0114
	[0.0130]
Unemployment Rate	0.486
	[0.692]
Activity Rate	-0.485
	[0.593]
Domain	0.000673
	[0.000569]
Central-Southern Africa	0.000135
	[0.000338]
Western Africa	-0.000011
	[0.0000357]
Eastern Africa	0.000140*
	[0.0000840]
Northern Africa	0.0000298
	[0.0000243]
Central-Eastern Europe	0.00000275
	[0.00000329]
EU	-0.0000111
	[0.0000773]
Other Europe	0.000728
	[0.000846]
Latin America	-0.0000209
	[0.0000146]
North America	0.000647
	[0.000954]
Central-Southern Asia	-0.0000238*
	[0.0000141]
Western Asia	0.000205
	[0.000357]
Eastern Asia	-0.0000153
	[0.0000125]
Oceania	-0.00287
	[0.00214]
Constant	-0.785
	[2.499]
Observations	206
Number of code	103
R-squared	0.955

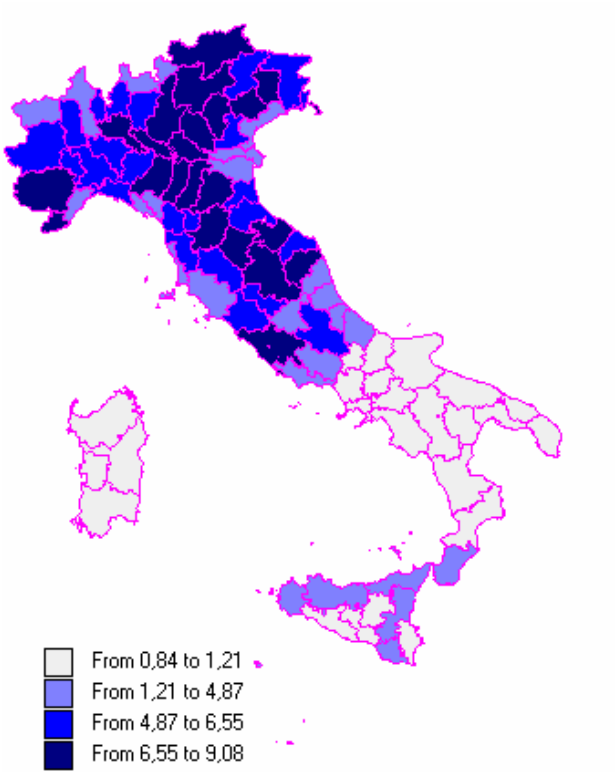
Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. The Two-way Error Component Model is applied to the logit transformation of the fractionalization index.

Figure 1: The index of fractionalization

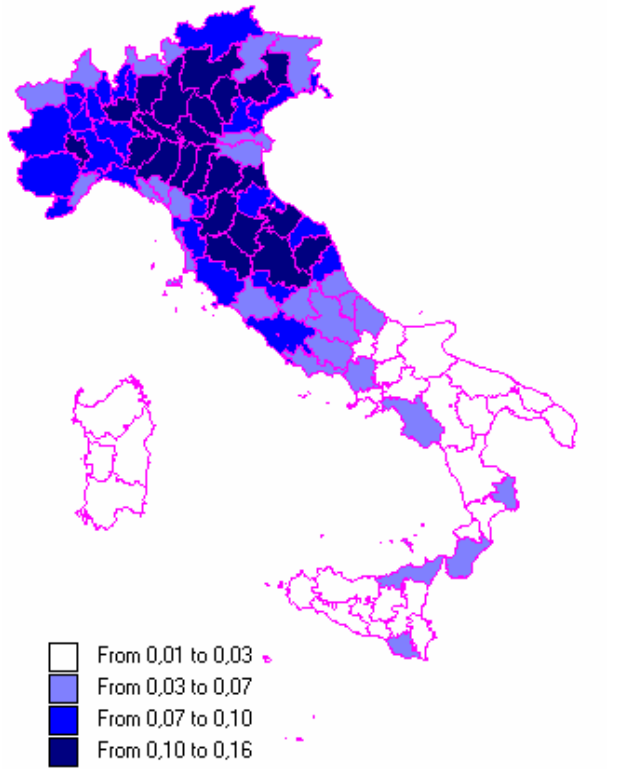
1991



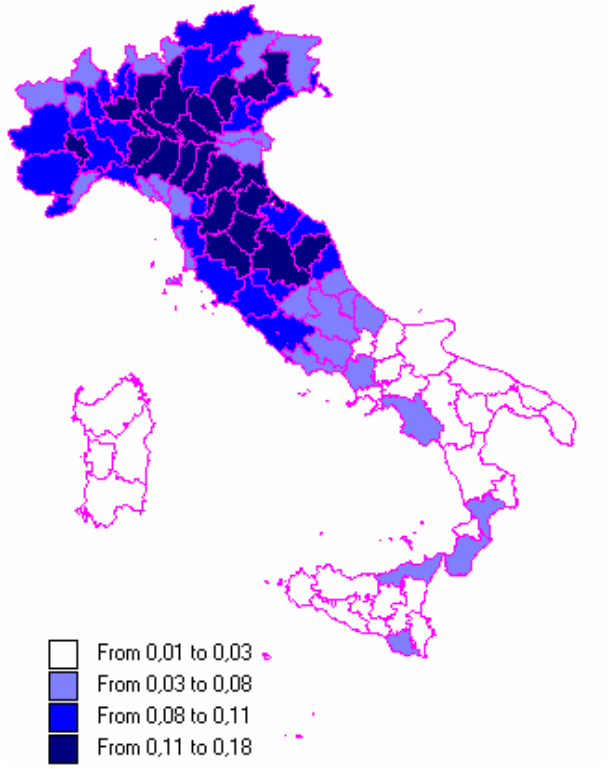
2001



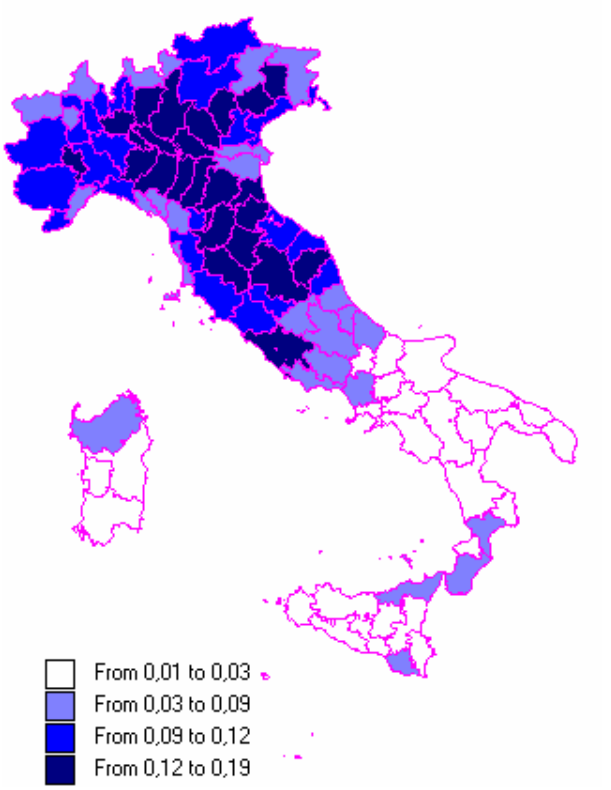
2004



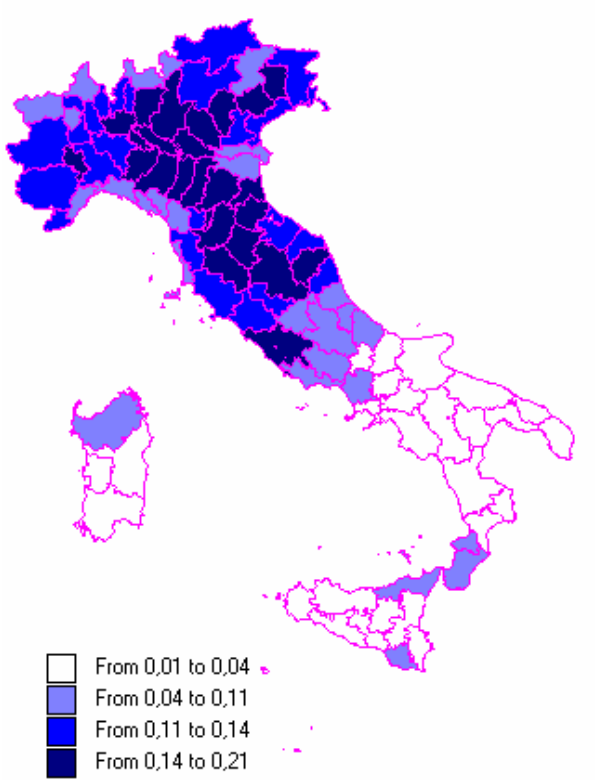
2005



2006


From 0,01 to 0,03
From 0,03 to 0,09
From 0,09 to 0,12
From 0,12 to 0,19

2007


From 0,01 to 0,04
From 0,04 to 0,11
From 0,11 to 0,14
From 0,14 to 0,21