LIKE OIL AND WATER OR CHOCOLATE AND PEANUT BUTTER? ETHNIC DIVERSITY AND SOCIAL PARTICIPATION OF YOUNG PEOPLE IN ENGLAND

ELENA FUMAGALLI E LAURA FUMAGALLI

Like oil and water or chocolate and peanut butter? Ethnic diversity and social participation of young people in England

Elena Fumagalli*
Laura Fumagalli†

very preliminary draft, please do not quote

Abstract

The paper studies the impact of ethnic diversity on social participation of young people. We first propose a theoretical model in which the agents choose between structured and unstructured social activities by taking into account the ethnic composition of the groups they join. We test our predictions using English census data together with the 'Longitudinal Survey of Young People in England' (LSYPE) and we find that ethnic segregation increases the probability of hanging around near home, while ethnic fractionalization decreases it. Furthermore, more structured activities are not affected by ethnic fractionalization. Finally, we use an IV strategy based on both historical and geographical data to correct for endogenous sorting into neighborhoods. The results we get are even stronger than those obtained where the ethnic composition of the neighborhood is taken as exogenous.

Keywords: social participation, fractionalization, segregation
JEL classification: C25, D71, J15

[1]We are grateful to Agar Brugiavini, Sergio Currarini, Emilia Del Bono, John Ermisch, Mario Padula and seminar participants at University 'Ca' Foscari' (Venice), ISER (University of Essex), IZA summer school 2009 and QED meeting 2009 (Amsterdam) for helpful comments and suggestions. This paper is based on work carried on during the visit to the European Centre for Analysis in the Social Sciences (ECASS) at the Institute for Social and Economic Research, University of Essex supported by the Access to Research Infrastructures action under the EU Improving Human Potential Programme

*University 'Ca' Foscari' of Venice. email: efumagalli@unive.it
†Institute for Social and Economic Research, University of Essex. email: lfumag@essex.ac.uk

1 Introduction

Due to the increase in the geographical mobility across the world, a growing number of economists and policy makers is getting interested in the link between ethnic diversity and social participation. Not only is social participation important *per se*, but it is also crucial because it is likely to be correlated with a broad set of labor market outcomes including productivity and wages. A recent article (Borghans, Weel, and Weinberg, 2008) shows that indicators at age 16 of a non-cognitive skill like sociability[1] are good predictors of people's future performance in the labor market, thus claiming the need of studying the pattern of sociability for adolescents. In this respect, Cunha and Heckman (2008) show that, while parental investment for children's cognitive skills should take place between age 6 and age 9, the sensible period for investing in non-cognitive skills occurs later in children's development i.e. between age 8 and age 13. In line with this idea, we study the relationship between ethnic composition of the neighborhood and social participation of young people and we test our predictions on a sample of respondents aged 13-14.

In particular, in our analysis of social participation, we distinguish between 'structured' and 'unstructured' activities where the former require a more developed capacity of planning ahead, while the latter are more spontaneous and they do not require any planning. The reason why we think that such a distinction must be of interest for social scientists and especially for economists is that, according to the well known theory of the big-five personality traits (see Goldberg, 1971), the preference for planned rather than spontaneous behavior is one of the main features characterizing the trait of conscientiousness[2] which has been shown to be closely related with leadership, longevity, college grades (Borghans, Duckworth, and Heckman, 2008) and job performance (Borghans, Duckworth, and Heckman, 2008; Salgado, 1997; Avis, Kudisch, and Fortunato, 2002; Fallon, Avis, Kudisch, Gornet, and A., 2000).

The topic has also direct policy implications in the debate on community cohesion and it can give useful suggestions for the implementation of new strategies to pursue a better integration of ethnic minorities. The importance of social participation in the process of integration of ethnic minorities has its roots in Granovetter's concept of 'weak ties' and 'strong ties' (Granovetter, 1973) and the derived Putnam's idea of 'bounding' and 'bridging' social capital (Putnam, 2000), where the former type of capital indicates 'ties to people who are like you in some important way', while the latter defines 'ties to people who are unlike you in some important way'. We believe that, while spontaneous aggregation is strongly driven by 'bounding social capital', forms of socialization where people share a common aim lead to the creation of 'bridging social capital' and they imply a higher level of interaction among different ethnic groups.[3]

We follow Alesina and La Ferrara (2000), in arguing that young people's social participation might be influenced by the ethnic composition of the neighborhood in which they live,

[1] these indicators include 'going to the cinema or disco', 'going to youth clubs', 'do community work', 'go to political clubs', 'going out with friends without a particular reason'

[2] the remaining four factors are Openness, Extraversion, Agreeableness, and Neuroticism

[3] On this respect, Putnam (2007) observes that in the USA 'Community centers, athletic fields, and schools were among the most efficacious instruments for incorporating new immigrants a century ago, and we need to reinvest in such places and activities once again, enabling us all to become comfortable with diversity'

however, we focus on the case of England since very few pieces of work are available in the economic literature for this country in spite of a growing interest among the policy makers. In fact, it is generally thought that 'the high levels of residential segregation found in many English towns would make it difficult to achieve community cohesion' (H.O., 2001) and that people living in areas characterized by high ethnic homogeneity are less likely to interact with neighbors from a different background. This is confirmed by a survey carried out by the Commission of Integration and Cohesion (CIC, 2007) showing that there is little interaction between people with different ethnic background, especially when spontaneous socialization is taken into account.

The model we propose is based on the idea that more spontaneous and less spontaneous forms of social participation differ in two respects. On the one hand they are different in the level of importance the agents attach to the interaction with other people belonging to the same ethnic group, on the other hand they take place at different geographical levels. In fact, while spontaneous social interaction is likely to occur in the neighborhood in which the adolescents live, the majority of the structured social activities are located in a broader area e.g. the district. Therefore, a changes in the ethnic composition of the area could lead to changes in the rate of participation in social activities as well as a a shift from one type of activity to the other.

We test our predictions by using different data sources including the Longitudinal Study of Young People in England (LSYPE): a large sample of young people at grade nine at school which contains detailed information on the pattern of socialization of the respondents. In particular, to the authors' best knowledge, LSYPE is the only survey which both permits to distinguish between spontaneous and non spontaneous forms of social participation and which contains some indication of the geographical level where the activities take place. The empirical analysis confirms the predictions derived from our theoretical model and it suggests that ethnic diversity, in all its different aspects, has a strong impact on spontaneous interaction, while it has a much weaker effect on more structured activities. In the last part of the paper we carry out an instrumental variable analysis to correct for attenuation bias and to address the problem of potential endogenous Tiebout type (Tiebout, 1956) sorting into areas. The results we got are even stronger than our previous findings.

The remaining of the paper is organized as follows. Section 2 presents the related literature both in sociology and in economics and it explains the location of our contribution in both disciplines. In section 3 we set up a model in which agents have to choose among alternative forms of social participation (including no participation at all). In section 4 we present the empirical application and we provide the instrumental variable analysis. Section 5 concludes.

2 Background

2.1 Ethnic minorities and social participation

The investigation of the causal link between ethnicity and social participation is relatively new in economics, while in sociology the topic has been studied since the first half of the twentieth century. However, most of this literature analyzes the case of the USA and it is based on evidence showing that black Americans seem to participate more than their white counterpart also when social status is controlled for. The contributions trying to explain such a stylized fact refer to two competing theories. The first one, labeled as *'compensation hypothesis'* and proposed by (Myrdal, Steiner, and Rose, 1944) argues that black people use social participation as a mean of getting involved in social life when whites tend to exclude them from it. On the other hand, Olsen's (Olsen, 1966) *'ethnic community hypothesis'* states that high participation among minorities is a consequence of their internal cultural solidity and of their ethnic community consciousness. The core difference between the 'compensation hypothesis' and the 'ethnic community hypothesis' is that they stress different aspects of social participation. In fact, the former interpretation relates participation to the set of social constraints ethnic minorities have to face, while the latter give more importance to the distinctive characteristics of ethnic minorities. Such a dichotomy is often referred in the debate between 'constraints' and 'preferences' as forces impacting on social interaction for ethnic minorities.[4]

The results provided for the USA in most of the literature seem to be robust when blacks and whites are compared. However, allowing for different ethnic groups within these broad categories and for different forms of participation raises some additional problems. Antunes and Gaitz (1975) argue that, while black people show a greater willingness to participate, the level of social participation among Mexican Americans seem to be lower than the one among whites. Moreover, they suggest that the different forms of participation should be divided into *'public'* and *'private'* in order to distinguish activities involving large-scale interaction from those taking place in small aggregations such as family or peer groups. An alternative distinction between different forms of social participation can be found in Putnam (Putnam, 2000) where social activities are divided into 'formal' and 'informal', The former include church attendance, volunteering, and everything related to community based project and political life, the latter, far more frequent, are 'less purposeful', 'less organized' 'more spontaneous and more flexible' forms of interaction like getting together for a drink, a dinner or a barbecue, gossiping with neighbors, watching TV with friends and so on.

Our paper is in line with Putnam (2000) and Antunes and Gaitz (1975) in allowing social participation to take different forms. In particular, we argue that the ethnic composition of the neighborhood where the young people live has differential effects on spontaneous participation and on a more conventional form of interaction in which people share a specific aim, as stated by Putnam (2007).

While the link between ethnicity and participation has been studied for long time for the USA, very few papers address the topic for England. However, these pieces of work have in

[4] see, for England, Platt (2007)

general the advantage of trying to study different forms of participation separately. Platt (2007) analyzes the relationship between illness, caring and ethnicity on social participation and she finds that different ethnic groups are characterized by diverse patterns of socialization, but the different types of social interaction (e.g organized activities and simple going out) turn out to be complements. Finally, using BHPS data, Li (2006) studies whether ethnicity (among other variables) determines social capital. Once again, the results show that the conclusions that can be drawn vary both across models differing for the type of participation analyzed and across different ethnic groups.

2.2 Measures and effects of ethnic diversity

A very interesting literature in economics focuses on the impact of ethnic diversity on several outcomes.[5] The definition of ethnic diversity is an umbrella term encompassing at least two different (although related) aspects and it has been measured through a set of different indicators.

The first aspect regards the degree of heterogeneity characterizing an area.[6] Ottaviano and Peri (2006) define 'culturally diverse cities' those having a larger share of foreign born people and they find that USA born citizens living in such areas pay higher rents than those living in more homogeneous ones. In a companion paper (Ottaviano and Peri, 2005) the same two authors measure cultural diversity by using an index constructed on the basis of the main language spoken at home and they conclude that wages for whites in diverse cities are higher than elsewhere. A tool which has been widely used in the literature in order to describe ethnic diversity is a decreasing transformation of the Herfindhal concentration index which can be interpreted as the probability that two individuals, randomly drawn from the entire population, belong to the same ethnic group. The starting point for the economic literature using such an index can be considered the article by Easterly and Levine (1997), in which the authors found that ethnic heterogeneity affects negatively many economic outcomes at the macro level. The same results are achieved by Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg (2003) by using both a fractionalization index computed by distinguishing groups on the basis of the language they speak and and a version of the same index distinguishing within language groups on the basis of certain physical characteristics.

A second aspect describing ethnic composition is ethnic segregation, measuring the degree of separation between different groups.[7] The definition of ethnic segregation has been used in many different formulations. For example Borjas (1998) defines 'ethnic segregated ghettos' those areas where 'the population that belongs to the respondent ethnic group is at least as large as would have been expected if the ethnic group was randomly allocated to the neighborhood'. Borjas' paper finds that ethnic spill-overs are significant for young people. Although there is no agreement on an index which best measures segregation, the literature

[5] A thorough review is Alesina and La Ferrara (2005).

[6] the words 'heterogeneity', 'fragmentation' and 'fractionalization' has been used by the literature to define the same phenomenon i.e. the degree of (ethnic) homogeneity.

[7] A third concept could be the idea of 'ethnic polarization' which is discussed in Montalvo and Reynal-Querol (2002). Although interesting, such a concept does not enter our model and it will not presented in details.

has mainly used the Duncan and Duncan dissimilarity index (Duncan and Duncan, 1955)[8]. Using the Duncan and Duncan index,[9] Cutler and Gleaser (1997) show that spatial segregation of different ethnic groups has an effect on on schooling, employment and the probability to become single mother and it harms Blacks more than Whites.[10] The Duncan and Duncan index is the indicator we will use to describe ethnic segregation in the reminder of our paper.

To the authors' knowledge, very few papers test the effects of fractionalization and segregation together. Gleaser, Schelnkman, and Shleifer (1995) study the effect of ethnic diversity on population growth for a cross-section of American cities and they find no effect of fractionalization but a positive effect of segregation for cities having large non whites communities. A paper explicitly claiming the need of using both measures is La Ferrara and Mele (2007). The authors analyze the link between ethnic diversity and public school expenditure and focus on the effects of changes in segregation for a given level of fractionalization. In particular, they argue that segregation influences the provision of school spending through two channels. The first channel is related to fractionalization, given that different levels of segregation for an area correspond to different level of fractionalization for the sub-areas. The second channel works through externalities among the sub-areas and it is based on the idea that the groups can cluster within areas in order to form homogeneous jurisdictions. The authors find that segregation has a positive impact on average public school expenditure both at the MSA level and at the district level. [11] However, segregation leads to a more unequal distribution of spending among districts belonging to the same metropolitan area.

2.3 Ethnic diversity and social participation

Very few papers study explicitly the relationship between ethnic diversity and social participation by stressing the importance of the ethnic composition of the neighborhood. A very interesting piece of work in this area, both for the topic addressed and for the methodology adopted, is Alesina and La Ferrara (2000). In their paper the authors explore the socio-economic determinants of social participation in a set of social activities (ranging from professional associations to church groups or literary clubs) by focusing on the role played by ethnic fractionalization.[12] By observing that preferences are likely to be determined by ethnicity and economic status and after testing their hypothesis by using USA data from the General Social Survey for the years 1974-1994, the authors conclude that there might be a link between ethnic fractionalization and social participation.

A recent paper addressing the topic for the UK is Letki (2008) which uses the Citizen Survey and Census data to study how racial fractionalization affects the degree of social capital. The author argues that the results supporting the hypothesis that ethnic fractionalization

[8]Other widely used indices are the square root index of segregation introduced by Hutchens (Hutchens, 2001, 2004) and the index of isolation (Bell, 1954). For a discussion on the properties of some indices of segregation see also Jenkins, Micklewright, and Schnepf (2006)

[9]together with the exposure index

[10]A similar result is achieved by Echenique and Roland (2007) by using a new index called the 'spectral segregation index' (SSI).

[11]MSA is the acronym indicating metropolitan areas in the USA, districts are a lower level aggregation and they are nested into MSA

[12]In this paper and in the one presented in the following paragraph ethnic fractionalization is measured by using the transformation of the Herfindhal concentration index we have introduced in the previous section

has an important impact on social capital might be driven by an omitted variable problem. In fact, after controlling for an index of Multiple Deprivation at the ward level, the impact of ethnic diversity seemed to be barely insignificant.[13] We think that this lack of results might be due to the fact that none of these papers does explicitly model the differences in social participation by aggregating each activity into homogeneous groups.

2.4 Our contribution

This paper makes several contributions.

The first one is the attempt of studying the effects of ethnic composition on people's behavior by explicitly distinguishing between different forms of social participation. In doing so, we try to find a contact point between the sociological literature aiming at classifying different social activities and the papers which, especially in economics, analyze the impact of ethnic diversity on various socio-economic outcomes. For our classification, we borrow Putnam's (Putnam, 2000) distinction between 'formal' and 'informal' activities, but we move the demarcation line in order to take into account the different geographical level where the two groups of activities take place. In particular, we stress the idea that activities can be divided into 'structured' and 'less structured', where the former are, in Putnam's terminology, 'more purposeful' in the sense that people do not meet only for the sake of staying together, but they get together because they share a common aim. As a consequence, we consider 'structured' a very broad range of activities ranging from Putnam's (Putnam, 2000) forms of 'civic and political participation' to less formal groups of actions like playing sport or going to the cinema. We argue that all these different activities share the characteristics of not being completely casual, since they must be planned in advance, perhaps discussing timing and methods with friends or peers. The opposite of such structured forms of socialization is the most casual interaction, i.e. the simple 'hanging around' with friends, which, in the case of the teenagers we are interested in, is a very important and common way of creating social ties. In particular, we argue that 'unstructured' social participation is likely to take place in the neighborhood where the kids live, while more structured activities have a broader geographical scope.

Explicitly modeling young people's choice within a geographical hierarchy is the second contribution of our paper. We focus on two levels of the English geography: the ward and the districts which is the level of aggregation into which wards are nested.[14] The reason why we use wards as our lower level is that, according to the Office for national Statistics the ward is the 'key building block of the English administrative geography', wards are used for the election of local government councillors and they can enjoy a certain degree of autonomy. The higher level of aggregation we use is based on the the local authority districts. The local authority districts constitute the most important geographical unit both because they enjoy greater administrative autonomy than the wards and because, since the 'Local Government

[13]However, as argued by Alesina and La Ferrara (2005), ethnic fractionalization can also have an effect on social participation through deprivation. In fact, racial or ethnic fractionalization can also be the cause of poor investment in physical infrastructures and so be positively correlated with deprivation, As a consequence, controlling for those variables can also lead to underestimation of the impact of ethnic fractionalization.

[14]For a more detailed description of the English geographical hierarchy, see the data appendix

Reorganisation' which took place in the 90s, the districts are in charge of providing many important services, before supplied by the counties. In particular, the districts are responsible for the cultural and recreational functions, which is why we assume that structured social activities take place at that upper level of aggregation.

The third contribution is intimately related to the first two and it derives from the idea that alternative concepts describing the ethnic composition of a given area (namely 'ethnic fractionalization' and 'ethnic segregation') measure two different aspects of ethnic diversity. In our explanation we follow La Ferrara and Mele (2007) in stressing the importance of considering a geographic hierarchy organized on different levels of aggregation and we model explicitly the relationship between segregation and fractionalization and its effects on alternative forms of social participation. In particular, we argue that, in order to explain participation in more structured activities, a probabilistic concept as measured by the index of fractionalization might not be enough to explain the forces driving social interaction and the possibility of searching within the districts must be taken into account.

Moreover, our paper focuses on social participation for teenagers which has hardly been analyzed in the literature, in spite of being of crucial importance for the future of the society (see Cunha and Heckman, 2008). Furthermore, young people are more likely to choose activities just maximizing their own utility with no sons, no old parents, no spouses involved in the decision. Finally, neighborhood effects seem to play an important role in shaping young people's preferences as stated by Case and Katz (1991), who claim that interaction with peers strongly shape young people's behaviors.

The last contribution is an attempt to solve the problem of potential endogenous sorting into districts ant it suggests an historically driven IV strategy to identify the causal effect of ethnic diversity on social participation.

3 The model

The model developed in this section explains how fractionalization and segregation[15] influence social interaction. In our model, which is a modification of Alesina and La Ferrara (2000) one, social interaction is influenced by two different characteristics of the agents: their ethnicity and their geographical location. We will consider two different geographical units: the ward, in which most of the unstructured social interaction takes places, and the district. We chose the district as our higher level of aggregation since we think that the young people rarely create social ties out of it. Furthermore, our choice of allowing structured forms of participation to take place at the district level is particularly meaningful in the case of England where the cultural and recreational functions are now under the responsibility of the Local Authorities.

In addition to ethnicity and geographical location, players are characterized by different levels of 'tolerance' toward different ethnic groups.

We assume that young people derive more utility matching with people similar to them, given that 'homophily' is a broadly observed phenomenon in the adolescent networks (see for example Currarini, Jackson, and Pin, 2007; McPherson, Smith-Lovin, and Cook, 2000). The choice is modeled as a simultaneous game whose payoffs are a function of the share of people of the same ethnicity who choose the same activity. Therefore, all the payoffs will be strictly greater than zero and smaller or equal to one. The model explains why and in which cases equilibria characterized by a certain level of integration can arise.

3.1 Two strategies and one parameter of tolerance

Let us consider a country composed by wards aggregated into districts. Let us assume for simplicity that there is just one district divided in two wards. In any ward two ethnic groups live: British *(b)* and non British *(n)*. Therefore, the population can be divided in 4 groups: British who live in ward 1 *(b1)*, British who live in ward two *(b2)*, non British who live in ward 1 *(n1)* and non British who live in ward two *(n2)*. Each group has to decide simultaneously whether to be involved in a *social activity (SA)*, which takes place within the district or to choose a less structured form of social interaction i.e. to *'hang around' (HA)* in the ward. People derive more utility from social interactions with people of the same ethnicity (*b or n)* and each type has its own level of 'tolerance' towards ethnic diversity, γ_{gi}, which is a parameter between 0 and 1, where γ_{gi} =1 identifies people who do not care about ethnic diversity. We assume that all the individuals of the same type choose the same strategy. In fact, the unit under analysis (the ward) is little and we believe that peer pressure and social influence is very strong, when we analyze small groups of young people.

Every child belonging to each group (*b1, b2, n1, n2)* is characterized by the a utility function with the usual characteristics of monotonicity and concavity according to the formulas:

$$\frac{\partial U_g(A)}{\partial - g} \leq 0 \tag{1}$$

[15] We adopt the definition of segregation measured by the Duncan and Duncan index

$$\frac{\partial U_g(A)}{\partial g} \geq 0 \tag{2}$$

$$\frac{\partial^2 U_g(A)}{\partial^2 g} \leq 0 \tag{3}$$

The utility is increasing in γ_{gi}:

$$\frac{\partial U_g(A)}{\partial \gamma_{gi}} \geq 0 \tag{4}$$

where $A = SA, HA$, $i = 1, 2$ and g= b, n.

From now on, we will consider a specific utility function, which has the stated properties:

$$U_{g1}(SA) = \frac{g1(SA) + g2(SA) + \gamma_{g1}((-g1(SA)) + (-g2(SA)))}{g1(SA) + g2(SA) + (-g1(SA)) + (-g2(SA))} \tag{5}$$

$$U_{g2}(SA) = U_{g1}(SA) \tag{6}$$

$$U_{g1}(HA) = \frac{g1(HA) + \gamma_{g1}(-g1(HA))}{g1(HA) + (-g1(HA))} \tag{7}$$

$$U_{g2}(HA) = \frac{g2(HA) + \gamma_{g2}(-g2(HA))}{g2(HA) + (-g2(HA))} \tag{8}$$

where gi(A) is the number of people belonging to group g who live in ward i and who choose the action A.

In the simultaneous game we constructed, there are 16 (4 players and 2 strategies) candidates to be a equilibrium in pure strategies, but just 4 of them can actually exist.

These are:

$$(b1, n1, b2, n2) = (HA, SA, HA, SA) \tag{9}$$

in which all the British 'hang around' and all the other are involved in social activities, and the opposite case

$$(b1, n1, b2, n2) = (SA, HA, SA, HA) \tag{10}$$

$$(b1, n1, b2, n2) = (HA, SA, SA, HA) \tag{11}$$

and

$$(b1, n1, b2, n2) = (SA, HA, HA, SA) \tag{12}$$

Equilibria (9) and (10) are equilibria with perfect segregation, while (11) and (12) are equilibria with partial integration

Equilibria with perfect segregation. Candidates (9) and (10) are equilibria with perfect segregation for each distribution of the ethnic groups within the district. In this kind of equilibria each ethnicity (in both wards) chooses the same activity, so there is no social

interaction between British and not British and everyone achieves the maximum possible utility. Equilibria with perfect segregation take place even when different people live together, this means that the absence of geographical segregation does not imply different ethnic groups hang around together or are both involved in social activities.

Equilibria with partial integration. Candidates (11) and (12) are equilibria with partial integration for particular sizes of the sub populations *b1, b2, n1, n2* (see Appendix 1).

In equilibria (11) and (12), even if the utility level two of the players[16] achieve is smaller than one, they have no incentive to deviate. This implies that there are some distributions of the population (with some degrees of segregation) which permit equilibria in which people of different ethnic groups find it optimal to interact. Apart from (9)-(12), we can easily rule out other types of equilibria, since there is not a composition of the population that supports them. The detailed description is presented in Appendix 1.

Proposition 1. *(9)-(12) are all equilibria if* $b1 = b2$ *and* $n1 = n2$.

Proof. Recall that candidates (9)-(10) are equilibria for each distribution on the population. Equilibrium (11) exists if *SA* is the best reply both for *n1* and for *b2*, since *b1* and *n2* already achieve the maximum possible pay-off. This means that two conditions have to be met.

$$U_{n1}(SA) \geq U_{n1}(HA) \tag{13}$$

$$U_{b2}(SA) \geq U_{b2}(HA) \tag{14}$$

condition (13) implies that

$$\frac{n1(SA) + \gamma_{n1} b2(SA)}{b2(SA) + n1(SA)} \geq \frac{n1(HA) + \gamma_{n1} b1(HA)}{b1(HA) + n1(HA)} \tag{15}$$

which means $b1 \geq b2$

while condition (14) implies that

$$\frac{b2(SA) + \gamma_{b2} n1(SA)}{b2(SA) + n1(SA)} \geq \frac{b2(HA) + \gamma_{b2} n2(HA)}{b2(HA) + n2(HA)} \tag{16}$$

which means $n2 \geq n1$.

Similarly (12) exists if *SA* is the best reply both for *b1* and *n2*, since *b2* and *n1* already achieve the maximum possible pay-off. It means that two conditions have to be met.

$$U_{b1}(SA) \geq U_{b1}(HA) \tag{17}$$

$$U_{n2}(SA) \geq U_{n2}(HA) \tag{18}$$

[16]Players of type b2 and n1 in 11 and of type b1 and n2 in 12

(17) and (14) imply that

$$\frac{b1(SA) + \gamma_{b1} n2(SA)}{n2(SA) + b1(SA)} \geq \frac{b1(HA) + \gamma_{b1} n1(HA)}{n1(HA) + b1(HA)} \tag{19}$$

which means $n1 \geq n2$

$$\frac{n2(SA) + \gamma_{n2} b1(SA)}{n2(SA) + b1(SA)} \geq \frac{n2(HA) + \gamma_{n2} b2(HA)}{n2(HA) + b2(HA)} \tag{20}$$

which means $b2 \geq b1$.

Therefore, in order to have both equilibrium (11) and equilibrium (12), we need $b1 = b2$ and $n1 = n2$. □

Proposition 2. *The game has only 2 equilibria if* $b1 < b2$ *and* $n1 < n2$ *or if* $b1 > b2$ *and* $n1 > n2$

Proof. In both cases neither conditions of equilibrium (11) nor condition of equilibrium (12) are met. □

Proposition 3. *For a given level of fractionalization, an increase of segregation does not decrease the number of people who play HA.*

Proof. It is easy to prove that an increase of the level of segregation that does not change the number of possible equilibria, increases the number of people who hang around.

Let us now consider an increase of the level of segregation changing the number of possible equilibria. It can be proved that the equilibrium with partial integration and high hanging around is more robust than the equilibrium with partial integration and less hanging around to changes in the number of possible equilibria due to an increase in segregation.

Framework 1 Let us first consider the case in which $n1 = n2$ and $b1 = b2$.

case 1 let us suppose some of the non British in ward 2 move to ward 1, i.e. $b1 = b2$ and $n1 > n2$. The ethnic fractionalization of the district is unaffected, but the two ethnic groups are distributed in a different way within it. In fact, this change increases the level of segregation, which was zero in the previous case.

In this new situation ($b1 = b2$ and $n1 > n2$), the conditions for the existence of equilibrium (12) are not met and just equilibrium (11) survives.

In equilibrium (11) the number of people who hang around ($n1 + b2$) is greater than the number of people who are involved in social activities ($n2 + b1$), since $n1 \geq n2$ and $b1 = b2$.

case 2 Assume now some of the non British in ward 1 move to ward 2, i.e $b1 = b2$ and $n1 < n2$.

In this case the conditions for existence of equilibrium (11) are not met and just equilibrium (12) survives.

In equilibrium (12) the number of people who hang around $(n2 + b1)$ is greater than the number of people who are involved in social activities, since $n1 < n2$

The previous cases show that, starting from the 4 equilibria case, an increase in segregation makes the equilibrium where fewer people hang around unfeasible and it does not affect the other equilibrium of partial integration, in which more people hang around.

case 3 Let us assume now that both some of the non British and some of the British in ward 1 move to ward 2, but let us also assume also the two changes are different in size (therefore we can keep a positive level of segregation) i.e $b1 \leq b2$ and $n1 \leq n2$.

In this case neither the conditions of equilibrium (11) nor the conditions of equilibrium (12) are met, and we are left only with the equilibria of perfect segregation.

Framework 2 Let us now consider the case in which we have 3 equilibria: the 2 equilibria with perfect segregation and the equilibrium of partial integration in which more people hang around

case 1 Reaching a point where there are 4 equilibria by increasing the level of segregation is impossible, since in the four equilibria case segregation is always zero.

case 2 The only way to make the equilibrium with partial integration and high hanging around unfeasible is decreasing the level of segregation

Framework 3 Finally, let us now consider the case with just the equilibria with perfect segregation: an increase of segregation changing the number of possible equilibria, makes the equilibrium with high hanging around feasible and it does not affect the equilibrium with low hanging around

□

We can easily prove that the same is true in the following cases.

$n1 = n2$ and $b1 \geq b2$, $n1 \leq n2$ and $b1 \geq b2$ when equilibrium (11) exists, $n1 = n2$ and $b1 \leq b2$, $n1 \leq n2$ and $b1 \geq b2$, $n1 \geq n2$ and $b1 \leq b2$, $b1 \leq b2$ and $n1 \geq n2$ when equilibrium (12) exists.

3.2 Two strategies and two parameters of tolerance

Now we modify the environment described before by introducing two different parameters of 'tolerance', β_i and γ_i. We can assume that tolerance is higher when we consider social activities, since the interest in the activity itself can compensate part of the disutility created by the interaction with people of a different ethnic group. In order to have $\gamma_i \geq \beta_i$ for all γ_i

and $0 \leq \beta_i \leq 1$, we can fix $\gamma_i = \sqrt{\beta_i}$. Parameter γ is type specific and it is known by the other players.

Thus, the utility functions for HA are modified as follows:

$$U_{g1}(HA) = \frac{g1(HA) + \beta_{g1}(-g1(HA))}{g1(HA) + (-g1(HA))} \tag{21}$$

$$U_{g2}(HA) = \frac{g2(HA) + \beta_{g2}(-g2(HA))}{g2(HA) + (-g2(HA))} \tag{22}$$

Also in this case the utility is increasing in β.

$$\frac{\partial U_{gi}(SA)}{\partial \beta_{gi}} \geq 0 \tag{23}$$

The utility function depends on the percentage of people of the same ethnic group who choose the same activity. The introduction of two different parameters of tolerance helps us studying not only the effect of a change in the geographical collocation of different ethnic groups within the districts, but also the effect of changes in the ethnic mixture of the districts itself. In fact, Since $\gamma_i \geq \beta_i$, a change in the composition of the population affects the two activities in different ways. The following proposition explains this mechanism.

Proposition 4. *A change in district's fractionalization has a greater effect on HA than on SA.*

Proof. Let us assume that b is the majority. Consider first the effect of participation in social activities chosen by British people due to a change in the share of non British people, which means a change in ethnic fractionalization. The utility to be considered is

$$U_b(SA) = Sb(SA) + \gamma_b Sn(SA) \tag{24}$$

Where *Sb(SA)* and *Sn(SA)* are the shares of British and non British who play SA. If we express (24) in terms of the share of the Non British who are involved in that activity, it becomes:

$$U_b(SA) = 1 - Sn(SA) + \gamma_b Sn(SA) \tag{25}$$

so the change in the utility due to a change in the composition of the population[17] is

$$\frac{\partial U_b(SA)}{\partial Sn} = -1 + \gamma_b \leq 0 \tag{26}$$

However, the increase of non British has an effect also on the other activity which is

$$\frac{\partial U_b(HA)}{\partial Sn} = -1 + \beta_b \leq 0 \tag{27}$$

[17] let us assume, without loss of generality that an increase of non British takes place

since $\gamma_b \geq \beta_b$, $-1+\gamma_b \geq -1+\beta_b$, so the utility of SA decreases less than the utility of HA.

The opposite is true, once we consider the choice of non British people.

$$\frac{\partial U_n(SA)}{\partial Sn} = 1-\gamma_n \geq 0 \tag{28}$$

$$\frac{\partial U_n(HA)}{\partial Sn} = 1-\beta_n \geq 0 \tag{29}$$

since $\gamma_n \geq \beta_n$, $1-\gamma_n \leq +1-\beta_n$, i.e. the utility of SA increases less than the utility of HA.

However, since b is the majority, at the aggregate level, the first effect prevails on the second, thus the net effect of a change in the ethnic composition on both activities is negative and it is greater in absolute value for spontaneous forms of social interaction.

□

The feasible equilibria are still the four we have already described, but the introduction of two different parameters of tolerance relaxes the conditions for the existence of the equilibria with partial integration. In general we can say that the higher is the level of tolerance γ, the more likely are the equilibria of partial integration. The following examples explain how equilibria depend on the parameter of tolerance.

Let us suppose that all the agents know the exact value of parameter γ, so they can use this information in order to make their choices.

Example 1. *consider the case in which* $b1 = b2$ *and* $n1 > n2$. *We have the two equilibria with perfect segregation and equilibrium (12)* $\forall\ \gamma$, *as in the case studied in the previous paragraph Furthermore, we have also equilibrium (11) for some values of* γ. *Since in this case players b1 and n2 already obtain utility equal to one, we just have to check the conditions for the other players. Thus, in order to have equilibrium (11), we need the following conditions to be met.*

$U_{n1}(SA) \geq U_{n1}(HA)$ *i.e.* $\frac{n1(SA)+\gamma_{n1}b2(SA)}{b2(SA)+n1(SA)} \geq \frac{n1(HA)+\beta_{n1}b1(HA)}{b1(HA)+n1(HA)}$ *which is always true since*

$$\gamma_i \geq \beta_i \tag{30}$$

by construction $U_{b2}(SA) \geq U_{b2}(HA)$ *i.e.* $\frac{b2(SA)+\gamma_{b2}n1(SA)}{b2(HA)+n1(HA)} \geq \frac{b2(HA)+\beta_{b2}n2(HA)}{b2(HA)+n2(HA)}$ *which is true*

$$\forall \frac{b2n1-b2n2}{n1(b2+n2)} \leq \gamma_{b2} \leq 1 \tag{31}$$

The value $\frac{b2n1-b2n2}{n1(b2+n2)}$ *is admissible, because it is greater than 0 since* $n1 \geq n2$ *and it is smaller than 1 if* $n1 \leq \frac{n1(1+n2)(1+b2^2+b2n2)}{b2^2}$.

The function achieves its maximum when b2 goes to zero and n2 goes to infinite.

So if b2 is a majority and the number of non British is higher in ward one than in ward two and it is not that high, (so it lays under the function $\frac{2b2n2+n2^2}{b2}$*), there are still some individuals with high* γ, *who decide to play SA.*

3.3 Three strategies and two parameters

In this paragraph we introduce a third strategy: staying at home (N). The key assumption of this section is that the utility derived from staying at home is strictly lower than one, i.e. for the ones who decide to stay at home it is impossible to reach the maximum level of utility. This captures the idea that socializing increases people's utility. In this new specification of the model the number of possible equilibria increases considerably (81, 3 strategies for 4 players). In addition to equilibria (9)-(12), just 4 other equilibria are possible.

These are the following:

$$(b1, n1, b2, n2) = (HA, N, SA, HA) \tag{32}$$

$$(b1, n1, b2, n2) = (HA, SA, N, HA) \tag{33}$$

$$(b1, n1, b2, n2) = (N, HA, HA, SA) \tag{34}$$

$$(b1, n1, b2, n2) = (SA, HA, HA, N) \tag{35}$$

Each player chooses to stay at home if her reservation utility ($\overline{U}$) is greater than the utility achieved when she is involved in another activity (HA or SA).

The example 1, described in the previous paragraph, is modified as follows:

Example 2. *we have the 2 equilibria with perfect segregation and equilibrium* (11) *with partial integration if* $\overline{U_{n2}} \leq \frac{n2+\gamma_{n2}b1}{n2+b1}$ *and* $\overline{U_{b1}} \leq \frac{b1+\gamma_{b1}n2}{n2+b1}$

The equilibria in which one group plays N exist when the utility of staying at home is greater than or equal to the utility of being involved in social activities,[18] *so respectively when*

$$\overline{U_{n1}} \leq \frac{n1 + \gamma_{n1}b2}{n1 + b2} \tag{36}$$

$$\overline{U_{b2}} \leq \frac{b2 + \gamma_{b2}n1}{n1 + b2} \tag{37}$$

$$\overline{U_{b1}} \leq \frac{b1 + \gamma_{b1}n2}{b1 + n2} \tag{38}$$

and

$$\overline{U_{n2}} \leq \frac{n2 + \gamma_{n2}b2}{b1 + n2} \tag{39}$$

It comes without saying that equilibrium (12) and equilibria (34)-(35) are mutually exclusive (if we disregard the case in which $\overline{U}$ *is exactly equal to the threshold)*

Proposition 5. *Among the young people involved in a social activity, the ones belonging to a minority are more tolerant.*

[18] It is because γ is always greater or equal β and because $b1 = b2$

Proof. We will prove this proposition in the case of two ethnicities, when the reservation utilities of the different ethnic groups are the same, so that a different level of participation is not the consequence of a different value of the outside option.

Let us consider the case in which all the players derive the same utility from staying at home, $\overline{U_{b1}} = \overline{U_{n1}} = \overline{U_{b2}} = \overline{U_{n2}}$.

Let us now consider the case in which both *b1* and *n2* want to play SA, therefore equilibrium (12) arises[19] and conditions (38)-(39) hold.

Let us define now $Sb1 = \frac{b1}{b1+n2}$ and $Sn2 = \frac{n2}{b1+n2}$, conditions (38)-(39) become:

B1 plays SA if

$$\overline{U_{b1}} \leq Sb1 + \gamma_{b1} Sn2 \tag{40}$$

which means

$$\overline{U_{b1}} \leq 1 - Sn2 + \gamma_{b1} Sn2 \tag{41}$$

$$\overline{U_{b1}} \leq 1 + (-1 + \gamma_{b1}) Sn2 \tag{42}$$

or

$$-1 + \gamma_{b1} \geq \frac{\overline{U_{b1}} - 1}{Sn2} \tag{43}$$

$$\gamma_{b1} \geq \frac{\overline{U_{b1}} - 1 - Sn2}{Sn2} \tag{44}$$

and if we consider *n2*

$$\overline{U_{b1}} \leq Sn2 + \gamma_{n2} Sb1 \tag{45}$$

$$\overline{U_{b1}} \leq Sn2 + \gamma_{n2}(1 - Sn2) \tag{46}$$

$$\gamma_{n2} \geq \frac{\overline{U_{b1}} - Sn2}{1 - Sn2} \tag{47}$$

So $\gamma_{b1} \leq \gamma_{n2}$ if $Sn2 \leq \frac{1}{2}$ □

It means that, when *n2* is the minority ($Sn2 \leq \frac{1}{2}$) they will decide to be involved in a social activity just if their level of tolerance is very high. This result is similar to the result found by Alesina and La Ferrara (2000). Note that in this case the ones who stay at home are not the ones belonging to the minority in their ward, but the ones who does not have a valid outside option in the district.[20]

[19] We can easily prove that the same is true if (12) arises

[20] the implication is the following: if I am a minority and I live in a ward in which I am the majority, I can still decide to stay at home if I can not find people of my type in the district

4 Empirical analysis

4.1 Data

The main data we use come from the first wave of the *'Longitudinal Study of Young People in England' (LSYPE)*, a new dataset created by the *'Department for Children, Schools and Families'(DCSF)* which contains detailed information for around 15000 pupils living in England. Data were collected between 30 March and 19 October 2004 and refer to young people in year 9 at school who were born between 1 September 1989 and 31 August 1990. The dataset is composed of different files, each derived from a separate section of the questionnaire. Apart from the 'young person section', where the questions are asked directly to the child, there is a 'household section', a 'main parent section', a 'young person history section' and an 'individual parent section'. Moreover, the dataset is useful for our research since ethnic minorities have been over-sampled, which implies that it is possible to derive separate results for relatively small ethnic groups.

In addition, a great advantage of the survey is that it is possible to link each respondent to the 'Lower Super Output Areas' (LSOA) in which she lives, which permit to construct the whole geographical hierarchy aggregating LSOA into wards, wards into districts, districts into counties.[21] Given such a rich structure, it is also potentially possible to link any measure computed at area level. In order to construct our indices of ethnic composition, we used data from the 2001 Population Census. The reason for our choice is that census data permit a high level of disaggregation (we used data at ward level) and they are representative of the population, whereas survey data can suffer from bias due to sampling design and differential non-response errors.[22]

Census data report the raw number of people living in each ward classified by ethnic group. The finest ethnic categorization available in the 2001 census distinguishes among: British, Irish, other whites, mixed white and black Caribbeans, mixed white and black Africans, mixed white and Asians, other mixed, Indians, Pakistanis, Bangladeshis, other Asians, Black Caribbeans, Black Africans, Other Blacks, Chinese, Others.[23] There is no agreement on how to aggregate these categories into coarser groups. In particular, it is not clear how 'mixed people' should be classified. On the one hand, one might want to classify, for example, individuals who have a black African parent and a white one in the black African group, but on the other hand it is tempting to group all the 'mixed' together in the same broad category. The latter choice emphasizes the degree of assimilation into the British society, but it adds together people having completely different backgrounds. When the researchers are interested in cultural specific variables, creating the category of 'mixed' can hide important differences. For this reason we chose not to create the 'mixed category' and we added data for mixed people to the closest ethnic minority group.[24]

[21] In order to aggregate census data at ward level, we used the Neighbourhood Statistics provided by the *'Office for National Statistics'*

[22] In principle, we could have used bootstrap techniques applied to LSYPE data as suggested in Jenkins, Micklewright, and Schnepf (2006), but the small sample size for most of the districts would not have permitted us to get reliable estimates.

[23] It is self reported ethnicity

[24] as a consequence 'mixed white and black Caribbean' are considered 'Caribbean' and so on

4.2 Variables involved and descriptive statistics

Summary statistics of the variables we used in our empirical analysis can be found in table 1.

Social participation In order to derive information on social participation, we use a couple of questions in the LSYPE youth questionnaire asking whether, in the four weeks before the interview, the respondent has done any of the activities presented in a list.[25] We chose some of the activities presented and we divided them into several groups.

The theoretical framework we have presented earlier in the paper urges us to distinguish between the simple not very purposeful 'hanging around' and a richer set of social activities implying a certain degree of planning. In order to construct the variable describing the first type of interaction, we use the entry in the list referring to the simple hang around near the respondent's home. Such an entry is really important for our purposes, since it explicitly mentions the geographical level of aggregation in which the activities take place, as stated in our theoretical model. In order to fully exploit this characteristics of our data, we decided not to use the last entry in the list, i.e. the one referring to the hanging around in the high street or in the town/city center. In fact, we think that such a form of social interaction might share some of the characteristics we described for more structured forms of social participation, since it implies making a move from a place to another.

Among the other social activities, we distinguish between forms of civic and political participation (i.e. going to political meetings or demonstrations, doing community work or going to youth clubs) and other forms of participation which we label 'sports and amusements' including: playing snooker, darts or pool, taking part in sports, going to see a football match or other sport event, going to a party, dance nightclub or disco, going to cinema, theater or concert. Although this second distinction is not used in our theoretical model, we decided to keep the traditional demarcation line between more community oriented activities and more self oriented ones as suggested by both Putnam (2000) and Antunes and Gaitz (1975) classification. In that, this paper adds also new pieces of evidence to the literature on the determinants of civic participation for ethnic minorities.

One might have noticed that 'hanging around with friends in the high street', 'going to a pub', 'going to an amusement arcade' and 'playing an instrument' do not belong to any group. The reason for the first exclusion has been already explained above and it relates to the need of testing our hypothesis on the geographical scope of different activities. We excluded playing an instrument because even those playing in a band are likely to spend a lot of time practicing alone, which can hardly be considered a social activity. For a similar reason we also excluded 'going to an amusement arcade', since it does not necessarily imply any form of interaction with peers. Finally, we excluded 'going to a pub' both because it lies between structured and unstructured forms of participation and because of the age of our sample. In fact, being the LSYPE respondents at this stage all younger than 18 years, the proportion of those going to a pub is likely to be a very selected sub sample of people who

[25] the exact list, as well as the question wording is reported in the appendix 2. Notice that the respondents can choose more than one activity.

probably engage in illegal activities, given that drinking alcohol is not allowed by the law for the class of people we are considering.

We can notice from table 2 that, although the participation rate is on average very high, it varies considerably across activities: almost half of those interviewed went to the cinema, more than 50 per cent of them took part in a sport activity, while less than 2 per cent took part in a political demonstration. In addition, a big percentage of young people just hang around both near home (around 55 per cent) and in the city center (around 30 per cent). It is interesting to notice that the percentage of respondents who hang around near home is much higher that the percentage of those going to the city center. This is not surprising, given the age of the individuals interviewed and it confirms our hypothesis that most of the spontaneous socialization takes place in the ward rather than in the district. Moreover, hanging around with friends seems to be the most common social activity, showing that failing to study such a form of socialization like most of the literature does, at least for the sub sample we consider, would hide an important aspect of people's social life. Active political participation is not an option many respondents choose (perhaps just because they are still too young), but, once we adopt a broader definition of politics and we consider all the activities implying a high level of civic engagement, such as being involved in community works or being enrolled in a youth club, participation in civic activities seems to be quite important.

Neighborhood level variables Among the explanatory variables, the ones we are mainly interested in are fractionalization and segregation. In order to measure fractionalization, we used the commonly used transformation of the Herfindahl concentration index according to the formula:

$$F_i = 1 - \sum_k s_{kj}^2 \tag{48}$$

Where F_j is the fractionalization index for the district j, while s_{kj} is the share of ethnic group k in the total population of the district. Like the other indices, this index of fractionalization ranges from 0 to 1, with a value close to zero indicating a low level of heterogeneity within the community and values close to 1 indicating extreme fractionalization. Table 3 reports the least and the most fragmented districts in England, showing that London is the city in which most of the not British people reside, while in the north the percentage of British is close to 100% and the fractionalization index is just above 0.002.

In order to measure segregation, we used the Duncan and Duncan index which can be written as:

$$D_j = \frac{1}{2} \sum_{i=1}^{W_j} \mid \frac{nb_{ij}}{NB_j} - \frac{b_{ij}}{B_j} \mid \tag{49}$$

Where nb_{ij} and b_{ij} are the numbers of non British and British people in the ward level i=1,.., W_i , NB_j and B_j are the numbers of non British people and British people in the district level and W is the number of wards in the district. Segregation is a more complex concept than fractionalization ad it necessarily implies a comparison between a lower and an

upper level (in our case the ward and the district). The indices of segregation compare the level of ethnic homogeneity characterizing each ward with the level of homogeneity characterizing the district, as a consequence, a high value of the index indicates a district where where people are clustered at ward level. The Duncan and Duncan index has been widely used sice it has an easy and useful interpretation, given that it can be interpreted as a 'share of people belonging to one of the groups that should move to another ward (without being replaced) in order to make the proportions between groups at ward level equal to the ones at the district level'.

Table 3 shows that as in the case of fractionalization, segregated districts are not equally distributed across England, since the northern-western areas are the most segregated, while in the south east residential separation between ethnic groups is not an issue. The choice of computing both indices at district level is in line with our theoretical model, since we think it reflects the distances that the young people might travel in order to take part in social activities

In graphs 1, 2 and 3 we plotted the relationship between ethnic diversity and social participation, considering the two different indices and the three groups of activity. Without conditioning on other variables, it is impossible to distinguish the sign of the correlation between the segregation index and the different types of social involvement. On the contrary, in line with proposition 4 of our theoretical model, we can observe an inverse relationship between ethnic fractionalization and spontaneous participation.

It is possible, of course, that the indices presented above might capture some of the heterogeneity among districts which is not necessarily due to differences in their ethnic composition. In order to control for a set of characteristics of the neighborhood which could be correlated with both our measures of ethnic composition, we used the IDACI[26] index of Multiple Deprivation, we took from the Pupil Level Annual School Census (PLASC) for 2004 [27] and we merged it into the main data. We chose the IDACI index since we are interested in the multiple deprivation for children. The IDACI index is available at Super Output Area which is even smaller than the ward. However, district level indices of deprivation can be computed by taking a weighted average of the values computed for each LSOA.[28]

Individual level and family levels controls At the individual level we have obviously controlled for ethnicity, which helps us shed light on the patterns of social participation for different ethnic groups which has has been largely studied by sociologists throughout the last three decades. Due to the big sample size achieved, together with the over sampling of the ethnic minorities, our survey data provide a big number of interviews for non British respon-

[26]The acronym IDACI stand for Income Deprivation Affecting Children Index and it represents the proportion of children aged 0-15 living in deprived households. the index is constructed on the basis of the number of children living in household receiving income support, income-based JSA households, pension Credit (Guarantee) households, Working tax Credit households where they are children receiving Child Tax Credit and whose equivalized income (excluding housing benefits ad before housing costs) is below 60% of the median of the population , child Tax Credit households not eligible for the schemes above whose equivalized income (excluding housing benefits and before housing costs) is below 60% of the median of the population, households composed by asylum seekers in England in receipt of subsistence support, accommodation support or both

[27]PLASC is part of the National Pupil Database (NPD)

[28]The calculation of the indices at district levels required the LSOA Level population at risk estimates available under request for each LSOA in the UK. See the data appendix for a detailed description.

dents, which permits to carry out reliable analysis also for subgroups that can not usually be analyzed separately. In order to avoid using an excessive number of ethnic dummies, we have aggregated the different ethnicities into the nine groups we have already presented earlier in the paper. Table 4 shows the ethnic composition of young people in the full sample. We can notice that more than 15 per cent of the LSYPE respondents define themselves as not British with Indian, Caribbean and Pakistani together representing half of the foreign population.

A look at the patterns of participation of different ethnic groups in table 2 provides some interesting insights. In general, Pakistani and Bangladeshi tend to participate less than their peers belonging to different ethnic groups. On the contrary, the participation rate of Caribbean and Black people, especially in civic activities, turns out to be quite high. On the side of spontaneous participation, British and Caribbean young people are the ones who hang around the most. However, even if ethnicity seems to be a driving force in explaining social participation, these data alone are not able to shed light on the mechanisms driving involvement in social activities.

As further controls at the individual level we use a dummy variable indicating whether the respondent was born in 1989 and another one indicating whether her main language at home is not English. With the first variable we want to capture any effect due to age, together with any other force affecting in a different way people who were born in a different calendar year, irrespective of the cohort they belong to when they enrol into school.[29] We are interested in the main language spoken by the respondents because we want to distinguish the role of constraints from the role of preferences. We expect that people whose main language is not English face linguistic barriers, which makes it difficult for them to interact with English speakers. We can think this linguistic constraint is not very binding for a large set of the population, since just the 2-3 per cent of the young people interviewed declared that at home they are used to speak a language different from English. Finally, we controlled for gender.

Moreover, we construct a variable which takes the value one if the school the young person attends provides any club or sport group. This variable is intended to provide some information about the supply side of participation opportunities with respect to sport. It may be that poor people do not take part in sports because they can not rely on adequate facilities, rather than because they are unwilling to be involved. If this supposition is true, the presence or absence of sport facilities in the school should be significant for the participation in 'sports and amusements'. By looking at the percentage of young people whose schools do not provide sport facilities (less than 8 per cent of the population), we can imagine that the main reason why some young people decide not to play sport is not due to the supply side. One might argue that such a variable could be endogenous, since parents who want their children to participate in sports are likely to choose schools with good sport facilities. LSYPE contains information on the reasons why the respondents' family chose the school the child attends and none of the respondents indicated the presence of sport facilities as a main reason for the children to enrol in a particular school, so we can discount the possibility of this variable being endogenous.

[29] all the respondents in the LSYPE sample belong to the same cohort and they are mainly born in 1989 or 1990. There is a small group of people who were born before or after these dates, but they have been excluded from the sample

The last group of variables we use gives interesting information on respondents' parental background which is also useful to control for selection into districts. The first variable indicates whether the young people usually eat with their family and it takes the value of 1 if the respondent has never eaten with his family in the week before the interview. We are not trying to claim the existence of any causal relationship between having dinner at home and participating in social activities, however, the variable can be seen as a proxy for the importance of family ties as claimed by Putnam (2000).[30] Finally, at the family level control for age and education of the main parent[31] and for a variable indicating which quintile of the distribution of the household income the respondent's family belongs to.

4.3 The Econometric model

In order to understand the effect of the ethnic composition of the neighborhood on each form of social participation we have just presented ('sports and amusements', 'civic activities' and 'hanging around near home'), we estimate the model by using a multivariate probit.[32] The choices of taking part in the three groups of activities are treated as interdependent but not mutually exclusive. At a minimum, interdependency is likely to exist between an individual's choices because of time constraints, while the reason for non-mutually-exclusiveness is that, while the different activities are substitutes in each point in time, they are compatible over the time span considered in the survey question.

The multivariate probit is a multiple-equations extension of the univariate probit allowing for non zero correlation among the error terms. The model is estimated through simulated maximum likelihood using the Geweke-Hajivassiliou-Keane (GHK) simulator (see Cappellari and Jenkins, 2003) and with standard errors clustered at the district level.

Let the latent variables y^*_{1ij}, y^*_{2ij} and y^*_{3ij} be, respectively, the propensity of taking part in 'sports and amusements', in 'civic activities' or in 'hanging around near home' for respondent i in district *j*. By using the latent continuous variable specification for probit models, the multivariate probit can be written as follows:

$$y^*_{1ij} = \alpha_{1ij} + fract_j\beta_1 + segr_j\beta_2 + \mathbf{x}'_{1j}\gamma_1 + \mathbf{x}'_{2ij}\delta_1 + \mathbf{x}'_{3ij}\eta_1 + \epsilon_{1ij} \tag{50}$$

$$y^*_{2ij} = \alpha_{2ij} + fract_j\beta_3 + segr_j\beta_4 + \mathbf{x}'_{1j}\gamma_2 + \mathbf{x}'_{2ij}\delta_2 + \mathbf{x}'_{3ij}\eta_2 + \epsilon_{2ij} \tag{51}$$

$$y^*_{3ij} = \alpha_2 + fract_j\beta_5 + segr_j\beta_6 + \mathbf{x}'_{1j}\gamma_3 + \mathbf{x}'_{2ij}\delta_3 + \mathbf{x}'_{3ij}\eta_3 + \epsilon_{3ij} \tag{52}$$

with

[30] Putnam (2000) observes that having a meal together is a 'traditionally important form of social connectedness' and he uses the frequency of family dinners to describe the variation over time of the strength of family ties

[31] We have included a variable indicating that the main parent has no qualification

[32] We have also estimated the model linearly by using seemingly unrelated models as in Zellner (1962), and the results are substantially the same as those obtained through multivariate probit (see appendix)

$$y_{1ij} = 1 \quad if \quad y^*_{1ij} > 0 \tag{53}$$

$$y_{2ij} = 1 \quad if \quad y^*_{2ij} > 0 \tag{54}$$

$$y_{3ij} = 1 \quad if \quad y^*_{3ij} > 0 \tag{55}$$

Where *fract* is the index measuring fractionalization, *segr* is the Duncan and Duncan index of segregation, $\mathbf{x}_{1j}$ are the neighborhood level controls, while $\mathbf{x}_{2ij}$ and $\mathbf{x}_{3ij}$ are the controls at family and at the individual level. The coefficients β_1, β_2, β_3, β_4, β_5, β_6 are the parameters of interest.

It is assumed that $\epsilon_1, \epsilon_2, \epsilon_3$ are error terms distributed as a multivariate normal with mean of zero and variance-covariance matrix V with V=1 on the leading diagonal and correlations $\rho_{ij} = \rho_{ji}$. If the off-diagonal correlations are equal to zero, the model is equivalent to a set of unrelated probit models, so, even if the hypothesis of interdependency was not correct, the estimates would not be affected.

4.4 Results

Let us start with a setting in which the option 'stay at home' is not feasible i.e. the one described in paragraphs 3.1-3.3. Our main interest is analyzing how the ethnic composition of the district, in both its aspects of segregation and fractionalization, affects social participation. Proposition 3 in our theoretical model predicts that *for a given level of fractionalization an increase of segregation does not decrease the number of people who play HA*, which means that the coefficient for segregation in a model including also a measure of fractionalization, can not be negative in the case of less structured forms of social interaction.

Table 6 presents the set of results obtained on the sub sample of those involved in at least one form of participation, which captures the idea that staying at home is not a valid option at this stage. In the simplest specification we included only the indices describing the ethnic composition of the neighborhood, in a slightly more complete model we added the IDACI index at district level to control for unobserved characteristics of the neighborhood and in the full model we added all the controls at the individual and family level we presented in the previous paragraphs.

It is easy to check that the prediction of the model is confirmed by our empirical analysis, given that the coefficient for segregation in the case of the variable describing spontaneous hanging around with friends in the neighborhood is positive and highly significant. Moreover, the result is robust to different specifications of the model and it becomes even stronger when new controls are added. The rational behind it is that any increase in segregation which leaves fractionalization unaffected is necessarily a result of a redistribution of the ethnic groups within the district which increases the homogeneity within wards. It is worth noticing that the predictions we can derive from the model for SA are exactly the opposite as those for HA, which implies that segregation should have a non positive coefficient in the case of more structured activities, which is what we observe in the data.

Proposition 4, based on the introduction of two different parameters of tolerance, predicts that, ceteris paribus, an increase of fractionalization has a greater (negative) effect on HA than on SA. As a consequence, we expect to get a negative and significant coefficient for fractionalization for the case of spontaneous interaction in the ward and a much more shaded result for more organized forms of social participation. Again, this is confirmed by our regression in table 6 in which the coefficient for ethnic fractionalization in the case of less structured activities is negative and highly significant, while it is insignificant in all the other cases.[33] The negative effect of fractionalization in the case of the simple hanging around with friends in the ward can be seen as a complement of the effect we found for segregation. In fact, in fragmented communities people can not rely on spontaneous forms of interaction because the probability of meeting someone of the same ethnic group simply by 'hanging around' is low, however, in segregated communities, due to the clustering of ethnic groups, British children can be more confident that in wandering around their neighborhood they will meet peers having a similar ethnic background.

The signs of the ethnic dummies deserve further comments. In fact, among those who take part in at least one form of social participation, British young people seem to be more likely to hang around with friends in the ward, while young people belonging to ethnic minorities seem to get more involved in more structured forms of participation. This can be due both to ethnic related preferences and to a mechanism close to the one we used in our theoretical model claiming that, by taking part in more structured social activities, people can engage in a within district process of search, allowing them to join sub groups where they are not the minority. On the contrary, in the ward ethnic minorities are usually the minority, which stops them to choose the simple hanging around.

Both our theoretical model and the results we have just discussed suggest that there is a form of substitutability between structured and less structured forms of participation, at least for those who take part in some forms of socialization. Table 8 (panel above) reports the signs of the estimated correlations between the three groups of activities. Let us point out first that in each model we have estimated, the correlations are always significant and this rules out the hypothesis of independence among equations, thus giving some evidence in favor of the use of the multivariate probit. It is easy to notice that the results are in line with our classification of forms of social participation, given that the correlation between sports and amusements and civic activities is positive, thus confirming the idea that they all belong to the broader group we defined as the group of the 'more purposeful' activities. Similarly, the simple hanging around has negative and significant correlation with any other activity and this, again, is a result confirming our initial hypothesis. All the correlations we have just discussed are significant at 1 percent level of significance.

The regressions on the sub sample of those involved in at least a form of interaction (table 6 and the above panel in table 8) are not useful to shed light on the reasons for participation itself, which are more clearly explained in a model constructed on the whole sample of our young people (table 7 and lower panel in table 8).

[33] It is worth noticing that this lack of significance is mostly due to low estimates of the coefficients rather than to high values of the standard errors, showing that the difference in the significance level is not due to the different number of activities used to construct our alternative forms of social participation

What happens when we consider also those having a higher propensity for staying at home? This is equivalent of making the option 'N' feasible. Recall that introducing the strategy 'staying at home' led to the equilibria (32), (33), (34), (35), which are similar to the equilibria (9), (10), (11), (12), but one of the two groups which was playing SA in the model with two strategies chooses to stay at home when such a strategy becomes a feasible choice. The main difference between table 6 and table 7 is probably the behavior of the ethnic dummies in the case of more structured social activities. Our results show that, when we run the regression on the full sample, the dummies for ethnic minorities for social activities becomes less positive and significant (and in some cases they turn also out to be negative) which means that the group of non British is split in two parts one of those plays SA, while the other plays N as predicted by our model. In fact, for those having a high γ or low $\overline{U}$ staying at home is not an option and those who are not involved in HA (namely the non British) choose to take part in a more structured social activity. The fact that the strategy 'staying at home' crowds out participation in more structured social activities, while it does not affect spontaneous participation has an interesting policy implication. Making structured social activities more attractive can be seen in the light of our model as an exogenous shifting in γ and $\overline{U}$ making γ to rise and $\overline{U}$ to decrease. This means that investing in sports or in youth groups could increase social participation for young people without having any negative effect on spontaneous aggregation.

The models in tables 7 and 8 are also useful to discuss and test some of the explanations for social participation suggested in the early sociological literature on the topic. For example, the negative and significant coefficient for segregation on civic activities is in line with the compensation hypothesis suggested by Myrdal, Steiner, and Rose (1944) and it can be explained by observing that, where segregation is low, ethnic minorities are more dispersed in the territory and they react to such a situation by opting for an Hirshmann's type (Hirschmann, 1970) voice mechanism. Another possible explanation is that there is a sort of substitution between the strength of ties felt within a young person's familiar-peer group and the consciousness of belonging to the society as a whole. Therefore, in segregated societies people prefer the interaction within their own ethnic group and such behavior 'crowds out' more broadly based activities. When we look at the pattern of participation in civic activities, we obtain results confirming both other empirical evidence available for the UK and the literature on social participation in the USA. In fact, our regressions show that, while Indians are less likely to join youth groups or to participate in civic activities, black Africans and black Caribbeans are more likely than British young people to take part in such activities, even when we do not consider church attendance.

In general, the complexity of the ethnic effects we found confirms the need of analyzing separately different ethnic groups and different forms of social participation.

Let us look at the behavior of the IDACI Index both in the model with neighborhood level variables and in the full model. In the model in column (II) the variable has the expected coefficients, given that it is negative and significant for the case of sports and amusements, positive and significant in the case of the spontaneous aggregation and not significant (although positive) in the case of civic activities. The explanation for the first

result can be that being involved in activities belonging to the first group requires paying a ticket or a fee, while the other activities are mainly for free. In this light, the coefficient for hanging around can be read as the other side of the coin, given that in deprived areas the only possible opportunity of socialization is likely to be the simple hanging around in the street. The lack of significance for the coefficient for deprivation in the case of civic activities reflects a complex mechanism, given that on the one hand deprivation can be associated with lack of education and political awareness, while, on the other hand, it can lead to an an Hirshmann's type voice mechanism leading to participation in community oriented activities. [34]

It is interesting to notice that when we control for personal and family level characteristics the significance of the coefficients computed for the IDACI index disappears, while other variables turn out to be significant. In particular, young people raised in wealthier families with older and more educated parents are more likely to get involved in sports and amusement while the opposite is true for the spontaneous aggregation. We think that the behavior of the IDACI index in model (II) picks up average characteristics of those living in a specific area rather than characteristics of the area itself, which explains why the effects found in model (II) disappear when more controls are added. This does not happen to our indices of ethnic composition since their coefficients in the cases we are interested in become even more significant in the full model.

The behavior of the variable proxying for the importance of familiar ties gives other interesting pieces of information. The coefficient of the variable is positive and significant for the generic 'hanging around' variable, while it is negative in the remaining cases (although not significant for civic activities). Although the effect of the variable can not be considered a causal effect, the coefficient seems to capture a form of heterogeneity between families, showing that the more cohesive families are also those that encourage social participation in more structured activities.

The variable indicating that English is not the main language for the respondent is negative and statistically significant for every form of social participation, although we control for ethnic background. Moreover, its effect has comparable size for the three forms of socialization. This can be interpreted as the part of the behavior of ethnic minorities explained by constraints rather than by preferences. This seems to suggest that the ethnic minorities have a lower degree of social interaction, not only because of their different system of values, but also because they face linguistic constraints. However, such a conclusion must be drawn with great caution since not having English as one's mother tongue may also be due to an unobserved heterogeneity which makes the group under study to differ from other migrant belonging to the same ethnic background. In fact, it is not easy to decide whether speaking a language other than English at home causes a lack of social interaction or it is a consequence of a lower propensity of interacting with the English environment. Unfortunately, we are not able to completely distinguish between the two effects.

Unsurprisingly the value of coefficient for the variable proxying for the presence of sport facilities at school is higher and more significant for the model studying participation in the first group of activities which includes sports. It suggests that, ceteris paribus, the presence

[34] Hirschmann (1970)

of many organized activities near home might increase the level of participation of young people. This, again, confirms that the observed pattern in social participation is the result of both the demand and the supply side. However, it is interesting to notice that being able to join a sports group does not crowd out participation in other types of activities which means that demand ad preferences do play an important role in people's choices. Finally, slightly older respondents are a little bit more likely to be involved in spontaneous socialization and males tend to participate more than females, especially when sports and amusements are considered.

Our final comments relate to the correlations between pairs of equations estimated on the full sample and shown in table 8. Now we do not have any negative correlation, which is in line with the idea that people having a higher propensity to stay at home do not take part in any activity, regardless of whether they are spontaneous or more structured.[35] However, the strongest correlation seems to be the one between the two activities we grouped in the SA group and this, once again, confirms our link between the theoretical model we propose and the LSYPE data.

4.5 Tackling the endogeneity problem

One of the main problems in the literature on neighborhood effects is that the identification of causal relationships with respect to neighborhood characteristics can be problematic. In fact, if people endogenously select into areas on the basis of some unobserved characteristics (Tiebout, 1956; Dustmann and Preston, 2001), it becomes difficult distinguishing between the effect of the neighborhood as a whole and the sum of the individual effects of its inhabitants.

In order to address this problem, we adopt different strategies. First of all we used census data for 2001 i.e. three years before the interview. On the one hand such a choice permits to overcome the problem of a simultaneous determination of the ethnic composition of the neighborhood and the choices of social participation of young people, on the other hand the distance in time is not huge, thus permitting to capture the characteristics of the cultural environment where the young people have been leaving and where social norms are formed.

Moreover, the indices for the ethnic composition of the neighborhood are constructed at district level, in order to avoid the sources of endogeneity due to Tiebout type sorting within districts (see Card and Rothstein, 2007). Finally, the endogeneity problem is mitigated by the fact that young people are less mobile than adults (see Cutler and Gleaser, 1997) and they are not directly involved in parental location decisions, thus the ethnic composition of the neighborhood in which they live is not directly correlated with their taste for social cohesion and it depends mainly on parental characteristics we can control for.

Instrumental variables analysis All the abovementioned strategies could not be enough to fully address the potential problem of endogenous sorting. As a consequence, we perform an instrumental variables analysis using different sets of instruments. We first take the fractionalization index as exogenous and we focus only on the index of ethnic segregation which we think is more at risk of being endogenous. The main reason for this choice is that,

[35] This is in line with what observed by Platt (2007)

while in the case of the fractionalization index constructing a district level indicator can fully solve the problem of endogenous selection within district, in the case of the segregation index at district level this is not possible, given that the distribution of ethnic groups within the local authority does play a role in the definition of the index. As a consequence, the degree of internal cohesion of each ethnic group within the district could influence both segregation and social participation. However, in a second stage we instrument both the indices of ethnic composition in order to avoid sources of endogeneity for the index of fractionalization due to the selection into districts.

In order to instrument the index of segregation we use historical data. As we have already mentioned, the data we use to construct our indices of ethnic composition were collected in 2001, three years before LSYPE data, such a distance in time could not be enough to avoid endogeneity, therefore we decided to go back in time and study the historical determinants of ethnic diversity in the UK. We claim that the forces which drove ethnic migration in the past are correlated with social participation in 2004 only through the actual ethnic composition.

Our identification strategy for the case of segregation is based on a well known stylized fact in the UK i.e. that the most racially segregated cities are concentrated in the north west of England and they coincide with the so called mill towns:[36] a definition which groups together a number of towns which used to lead the industrial revolution due to their importance in the textile (mainly cotton) industry. The reason for the correlation between the textile industry and ethnic segregation is that the mills town experienced a huge flows of migrants coming from outside the country which never mixed with the English majority. In line with this idea, we selected a group of cities listed as mill towns in 1830[37], then we computed the geodetic distance (i.e. the shortest path along the earth at sea level between each district in England and these new reference points) according to the formula:[38]

$$\begin{aligned} dist_{ab} &= \arccos(\sin(lat(a)) * \sin(lat(b)) + \cos(lat(a)) \\ &\quad * \cos(lat(b)) * \cos(\mid (long(b) - long(a)) \mid)) * 6371 \end{aligned} \tag{56}$$

where a and b are the two points considered, *lat(a)*, *lat(b)*, *long(a)* and *long(b)* are the latitude and the longitude in radians of each point and the number 6371 is the radius of the earth in kilometers.

Finally, we took the mean distance between each point and the first two closest mill towns[39] as follows:

$$segr\_iv_i = \frac{dist_{i\bar{j}} + dist_{i\bar{k}}}{2} \tag{57}$$

[36] violent racially driven riots took place former mill towns like Oldham, Bradford, Leeds and Burney in 2001 which confirms that ethnic segregation keeps on having consequences nowadays

[37] as a robustness check we performed the same exercise by using a more comprehensive list of mill towns encompassing a longer time span and the results do not change substantially. The results of these regressions are available under request.

[38] in our formula we assume the earth is a regular sphere, which is a commonly used approximation

[39] we used two reference points in order to account for geographical clustering

where

$$\overline{j} : dist_{i\overline{j}} = min(dist_{ij}) \forall j \tag{58}$$

and

$$\overline{k} : dist_{i\overline{k}} = min(dist_{ij}) \forall j \neq \overline{j} \tag{59}$$

In a following step we relax the assumption of exogeneity of the fractionalization index and we instrument both our measures of ethnic diversity. In order to do so, we compute the distance of each district with the main British ports of entry in the first half of the 20th century and we take the mean distance from the two nearest reference points as stated in (57), (58) and (59). There are two reasons why our reference points can be considered exogenous. First of all the position of the ports is quite exogenously determined, given that it depends strongly on the morphology of the land, second, we selected the list of the places of entry in the UK according to an historical document i.e. the *Aliens Act* (1905-1919) through which entry was restricted through a number of approved ports, namely: Cardiff, Dover, Folkestone, Grangemouth, Grimsby, Harwich, Hull, Leith, Liverpool, London, Newhaven, Southampton, the Tyne Ports and Plymouth.[40]

One might argue that using information from the past could lead to the well known weak instrument problem, given that the migration in the past might be only weakly correlated with the actual ethnic composition. There are some historical reasons why we think this is not the case. The first reason is that the composition of the ethnic minority population has not changed much. Holmes (2001) observes that even before the First World War the groups we consider were the biggest non European ethnic minority groups composing the mosaic of the British society, with the Chinese community being the smallest in size, in line with our recent data.

The second reason is related to the history of migration in Britain. The British immigration law in place in the 19th century permitted unrestricted entry of migrants into Britain, however, starting from the 1905 *Aliens Act* and thorough the whole century, more and more restrictions have been posed. We think the British legislative activity with respect to migration can be divided into two phases.

Restrictions related to the ports of entry. From 1905 to 1919 it was possible to enter in the country just through a restricted number of ports. Evidence shows that until the mid 1940s ethnic minorities were confined to London, the ports and some university towns (see Holmes, 2001)[41]

Restrictions through vouchers and quota system. After the Second World War, migrants started moving to the great conurbations, this process took place together with a great diversification in the occupational structure and a greater involvement in political life. In line with such a trend, the *Commonwealth Immigrants Act* of 1962 set up an entry system

[40] Plymouth was added later, (see Pellew, 1989)

[41] Little (1948) observes that in ports like Cardiff *'the most bitter competition between white and coloured seamen took place'* and such a fight was followed by the *Special Restriction (Coloured Alien Seamen) Order* in 1925

for Commonwealth workers based on vouchers[42] which was followed by a quota system under the 1968 Act and by further restrictions based on evidence of partiality in 1971.

The reasons why we do not compute the distances from the great clusters formed in the 60s are twofold. First of all it could not be enough to tackle endogeneity, given that this second wave of migration was influenced by the right movements in America and by the so called 'race relations industry', both of which are likely to be correlated with social participation. Secondly, even in this second phase, the residential choices of ethnic minorities were still strongly influenced by the previous restrictions on the ports of entry. In fact, the immigration control of the 1960s made it impossible temporary migration and forced new migrants to join the pioneers in places where they were already settled down.[43]

A methodology similar to ours is used also in Ottaviano and Peri (2006) for the USA, but it differs from the strategy used here because Britain is an island and it is probably easier defining specific ports of entry, given that we do not have to deal with political boundaries like in the case of the Mexican and the Canadian frontier for the United States. Moreover, our approach is original because it selects the reference points on the basis of an historical analysis and on a specific legislative act.

Being our model non linear, we use a control function approach. In the first step we derive the residuals of the reduced-form equations for the endogenous explanatory variables, then, we include these residuals in the second step as additional explanatory variables.[44] Table 9 shows the first stages both in the case in which fractionalization is taken as exogenous and in the case in which both indices are instrumented. As suggested by Angrist and Krueger (2001), the first stage regressions are simple OLS models[45] where all the covariates are averaged at district level like in Alesina and La Ferrara (2000). Not surprisingly, we found that, in both our specifications, the distance from the mill towns has a good predicting power for the case of segregation with close districts being more segregated than districts which are more far away. Moreover, the same variable is a poor predictor for ethnic fractionalization when the latter is taken as endogenous, which suggests that our identification strategy is corroborated by the data. Similarly, the distance from the main ports of entry listed in the *Aliens Act* has a negative and significant coefficient in the case of fractionalization and an insignificant coefficient in the case of segregation.

Tables 10, 11, 12 show the results of the second stages of our IV strategy. Accounting for endogenous selection into districts makes the results for the coefficients for ethnic composition stronger than the ones we got in our previous models. A reason for that could be that some of the more segregated districts are located in quite big cities having a more complex social structure, better chances of taking part in political activities and smaller emphasis on

[42] Previously commonwealth workers had enjoyed at least formally unrestricted entry in the country

[43] See again Holmes (2001)

[44] Following Angrist and Krueger (2001) have also estimated linear second stages (SUR models) where the potentially endogenous variables are substituted by their fitted counterparts computed in the first stage. The results are reported in the appendix and they are substantially the same as the ones we obtained through the control function approach

[45] We had also run beta models and fractional models like in Papke and Wooldridge (1996), but the results were similar in terms of sign and significance of the coefficients to those obtained with the OLS. These results are available under request.

spontaneous aggregation. This could hide some of the effects of segregation on our different forms of social participation. Another reason why we get stronger results after performing the IV analysis could be due to the well know 'attenuation bias' which can arise as a consequence of the way in which the data are aggregated into our indices (for segregation see (see Card and Rothstein, 2007)). Using predictions based on the structural determinants of segregation, but unrelated to the way in which this is measured, can eliminate this further source of endogeneity.

The effect of applying our instrumental variable strategy is much more shaded in the case of ethnic fractionalization and this is in line with our belief that such an indicator, if constructed at district level, is likely to be less endogenous than the Duncan and Duncan index at district level we used to measure ethnic segregation.

5 Conclusions

This paper addresses the question of whether and how the ethnic composition of the area where young people live affects their pattern of social participation. We distinguish between spontaneous and more structured forms of participation and we argue that while the former are likely to take place in the ward where the individuals reside, the latter have a broader scope and they take place in the district. We show that, due to the existence of the possibility of searching within district, it is possible to have equilibria with partial integration where the separation between ethnic groups is not complete.

Both our theoretical model and our empirical application suggest that ethnic fractionalization discourages spontaneous socialization, while ethnic segregation seems to make it easier. The results on more structured activities are more shaded and they vary a bit when different sub groups of activities are considered. However, we still find a negative effect of segregation on participation in civic activities. Our empirical analysis on the determinants of participation shows that, at least for the sub sample of those taking part in at least one activity, the simple 'hanging around' in the neighborhood is a substitute for more structured activities, namely civic activities and 'sport and amusements'. Our predictions do not change even when we control for endogenous selection into districts: after applying an instrumental variable approach using historical data, the results we got are even stronger especially in the case of ethnic segregation.

In conclusion, our empirical analysis shows that the ethnic composition of the neighborhood, however measured, has a robust effect on spontaneous socialization and a much weaker effect on other forms of social interactions. This finding supports one of the main hypothesis of our theoretical framework i.e. that in the case of more structured activities, people give less importance to the ethnicity of those they interact with. This contains a policy implication which, in line with Putnam (2007), argues that in order to foster ethnic cohesiveness a country should provide people with chances for of sharing a common aim through different opportunities of social participation.

References

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003): "Fractionalization," *Journal of Economic Growth*, 8, 155–194.

Alesina, A., and E. La Ferrara (2000): "Participation in heterogeneous communities," *The Quarterly Journal of Economics*, 115 (3), 847, 904.

——— (2005): "Ethnic Diversity and Economic performance," *Journal of economic Literature*, 43 (3), 762–800.

Angrist, J., and A. Krueger (2001): "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, 15(4), 6985.

Antunes, G., and G. Gaitz (1975): "Ethicity and participation: a study of Mexican-Americans, Blacks and Whites," *The American Journal of Sociology*, 80 No.5, 1192–1211.

Avis, J., J. Kudisch, and V. Fortunato (2002): "Examining the Incremental Validity and Adverse Impact of Cognitive Ability and Conscientiousness on Job Performance," *Journal of Business and Psychology*, 17(1), 87–105.

Bell, W. (1954): "A Probability Model for the Measurement of the Ecological Segregation," *Soc. Forces*, 32, 357–64.

Borghans, L., A. Duckworth, and W. B. Heckman, J.J. (2008): "The Economics and Psychology of personality Traits," *Journal of Human Resources*, 53(4), 972–1059.

Borghans, L., B. Weel, and B. Weinberg (2008): "Interpersonal Styles and Labour Market Outcomes," *Journal of Human Resources*, 43(4), 815–858.

Borjas, G. J. (1998): "To Ghetto or not to Ghetto: Ethnicity and Residential Segregation," *Journal of Urban Economics*, 44, 228–253.

Cappellari, L., and S. Jenkins (2003): "Multivariate probit regression using simulated maximum likelihood," United Kingdom Stata Users' Group Meetings 2003 10, Stata Users Group.

Card, D., and J. Rothstein (2007): "Racial Segregation and the Black-White test Score Gap," Discussion paper, NBER working paper No 12078, Forthcoming in Journal of Public Economics.

Case, A., and L. F. Katz (1991): "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths," Discussion paper, NBER Working Paper No. W3705.

CIC (2007): "Public attitudes towards cohesion and integration, Commission on Integration and Cohesion," .

Cunha, F., and J. Heckman (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 53(4), 738–782.

Currarini, S., O. Jackson, and P. Pin (2007): "An economic model of friendship: homophily, minorties and segregation," Discussion Paper 2007-20, University of Venice, department of economics, Forthcoming in Econometrica.

Cutler, D. M., and E. L. Gleaser (1997): "Are Ghettos Good or Bad?," *The Quarterly journal of Economics*, 112 No. 3, 827–872.

Duncan, O., and B. Duncan (1955): "A methodological analysis of segregation indexes," *American Sociological Review*, 20, 210–217.

Dustmann, C., and I. Preston (2001): "Attitudes to Ethnic Minorities, Ethnic context and Location Decision," *The Economic Journal*, 111 No 470, 353–373.

Easterly, W., and R. Levine (1997): "Africa's Growth Tragedy: Policies and Ethnic Division," *The Quarterly Journal of Economics*, 112(4), 1203–1250.

Echenique, F., and G. Roland (2007): "A measure of segregation based on social interactions," *Quarterly Journal of Economics*, 122 No.2, 441–482.

Fallon, J., J. Avis, J. Kudisch, T. Gornet, and F. A. (2000): "Conscientiousness as a Predictor of Productive and Counterproductive Behaviors," *Journal of Business and Psychology*, 15(2), 339–349.

Gleaser, E., J. Schelnkman, and A. Shleifer (1995): "Economic Growth in a Cross-Section of Cities," *Journal of Monetary Economics*, 36(1), 117–43.

Goldberg, L. (1971): *Advances in Psychological Assessment*chap. A Historical Survey of personality Scales and Inventories, pp. 293–336. Science and Behavior Books.

Granovetter, M. (1973): "The strenght of Weak ties," *The American Journal of Sociology*, 78(6), 1360–1380.

Hirschmann, A. O. (1970): *Exit, Voice and Loyality: Responses to decline in firms, Organizations and States*. Cambridge MA. Harvard University Press.

H.O. (2001): "Community Cohesion. A Report of the Independent Review Team, chaired by Ted Cantle.," .

Holmes, C. (2001): *A tolerant country? Immigrants refugees and Minorities in Britain.*

Hutchens, R. (2001): "Numerical Measures of segregation: desirable properties and their implications," *Mathematical Social Sciences*, 42, 13–29.

Hutchens, R. (2004): "One measure of Segregation," *International Economic Review*, 45, 555–578.

Jenkins, S. P., J. Micklewright, and S. Schnepf (2006): "Social Segregation in Secondary Schools: how does England compare with other Countries," Discussion paper, ISER working papers 2006-2.

La Ferrara, E., and A. Mele (2007): "Racial Segregation and Public School Expenditure," Discussion paper, CEPR Discussion Paper 5750.

Letki, N. (2008): ""Does Diversity Erode Social Cohesion? Social Capital and Race in British Neighbourhoods," Political Studies, Political Studies Association, vol. 56, pages 99-126, 03.," *Political Studies*, 56, 99–126.

Li, Y. (2006): "Social capital, ethnicity and the labour market," *Proceedings the International Conference on Engaging Community*.

Little, K. (1948): *Negroes in Britain a study of racial relations in English society*.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2000): "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 27, 415–444.

Montalvo, J. G., and M. Reynal-Querol (2002): "Why Ethnic Fractionalization? Polarization, Ethnic Conflict and Growth," Economics Working Papers 660, Department of Economics and Business, Universitat Pompeu Fabra.

Myrdal, G., R. Steiner, and A. Rose (1944): *An American Dilemma*. Harper, New York.

Olsen, M. (1966): "Social and Political Participation of Negroes," *American Journal of Sociology*, 72 (July), 32–46.

Ottaviano, G., and G. Peri (2005): "Cities and cultures," *Journal of Urban Economics, Elsevier, vol. 58(2)*, 58(2), 304–337.

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

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*, 11(6), 619–32.

Pellew, J. (1989): "The Home Office and the Aliens Act, 1905," *The Historical Journal*, 32 (2), 369–385.

Platt, L. (2007): "On social activity. Assessing the impact of ilness, caring and ethnicity," Discussion paper, CASE papers 108, Centre for Analysis of Social Exclusion, LSE.

Putnam, R. (2000): *Bowling Alone: The Collapse and Revival of American Community*.

Putnam, R. D. (2007): ""Et pluribus unum": Diversity and community in the Twenty-First Century. The 2006 Johan Skytte Prize Lecture," *Scandinavian Political Studies*, 30 (2), 137–174.

Salgado, J. (1997): "The Five Factor Model of Personality and Job Performance in the European Community," *Journal of Applied Psychology*, 82(1), 30–43.

Tiebout, C. (1956): "A Pure Theory of Local Expenditures," *The Journal of Political Economy*, 64 (5), 416–24.

Zellner, A. (1962): "An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias," *Journal of the American Statistical Association*, 57, 348368.

Appendix 1: Description of the equilibria

Proposition A1.1: In the game with two strategies and one parameter of tolerance $\nexists$ a distribution of the population that supports the following equilibria.

- $(b1, n1, b2, n2) = (SA, SA, SA, SA)$

  Proof. everybody has an incentive to deviate and "hang around" in her own ward, obtaining a payoff equal to one. □

- $(b1, n1, b2, n2) = (HA, HA, HA, HA)$

  Proof. everybody has an incentive to deviate and play SA, obtaining a payoff equal to one. □

- $(b1, n1, b2, n2) = (HA, SA, SA, SA)$, $(b1, n1, b2, n2) = (SA, HA, SA, SA)$, $(b1, n1, b2, n2) = (SA, SA, HA, SA)$, $(b1, n1, b2, n2) = (SA, SA, SA, HA)$

  Proof. Since people from one ward are "hanging around", one of the two wards is "empty", so there is an incentive to play HA there, obtaining a payoff equal to one. □

- $(b1, n1, b2, n2) = (SA, SA, HA, HA)$:

  Proof. people living of ethnicity b living in 1 and people of ethnicity n living in 1 have an incentive to play HA in their own ward (ward one), obtaining the maximum pay off. □

- $(b1, n1, b2, n2) = (HA, HA, SA, SA)$:

  Proof. people of both ethnicities living in 2 have an incentive to play HA in their own ward (ward two), obtaining the maximum pay off. □

- $(b1, n1, b2, n2) = (SA, HA, HA, HA)$, $(b1, n1, b2, n2) = (HA, SA, HA, HA)$, $(b1, n1, b2, n2) = (HA, HA, SA, HA)$, $(b1, n1, b2, n2) = (HA, HA, HA, SA)$

  Proof. Let us consider the equilibrium $(b1, n1, b2, n2) = (SA, HA, HA, HA)$: people living in n2 have an incentive to deviate and to play SA with the players of their own type (living in b1), obtaining the maximum payoff. The same applies to players in b2, n1 and b1 in the other cases □

number and nature of equilibria in a game with two strategies and one parameter of tolerance Let us summarize the number and the nature of the equilibria the different cases. In addition to (9) and (10) we have:

- if $b1 > b2$ and $n1 < n2$: equilibrium (11)
- if $b1 > b2$ and $n1 > n2$: neither equilibrium (11) nor equilibrium (12) can take place
- if $b1 > b2$ and $n1 = n2$: equilibrium (11)
- if $b1 < b2$ and $n1 < n2$: neither equilibrium (11) nor equilibrium (12) can take place
- if $b1 < b2$ and $n1 > n2$: equilibrium (12)
- if $b1 < b2$ and $n1 = n2$: equilibrium (12)
- if $b1 = b2$ and $n1 < n2$: equilibrium (11)
- if $b1 = b2$ and $n1 > n2$: equilibrium (12)
- if $b1 = b2$ and $n1 = n2$: both equilibrium (11) and equilibrium (12)

number and nature of equilibria in a game with two strategies an two parameters of tolerance Let us summarize the number and the nature of the equilibria the different cases. In addition to (9) and (10) we have:

- if $b1 > b2$ and $n1 < n2$: (11) $\forall\gamma_i$ and (12) if $\gamma_{b1} > \frac{b1n2-b1n1}{n1(b1+n2)}$ and $\gamma_{n2} > \frac{b1n2-b2n2}{b2(b1+n2)}$
- if $b1 > b2$ and $n1 > n2$: (11) if $\gamma_{b2} > \frac{b2n1-b2n2}{n2(b2+n2)}$ and (12) if $\gamma_{n2} > \frac{b1n2-b2n2}{b2(b1+n2)}$
- if $b1 > b2$ and $n1 = n2$: (11) $\forall\gamma_i$ and (12) if $\gamma_{n2} > \frac{b1n2-b2n2}{b2(b1+n2)}$
- if $b1 < b2$ and $n1 < n2$: (11) if $\gamma_{n1} > \frac{b2n1-b1n1}{b1(b2+n1)}$ and (12) if $\gamma_{b1} > \frac{b1n2-b1n1}{n1(b1+n2)}$
- if $b1 < b2$ and $n1 > n2$: (11) if $\gamma_{n1} > \frac{b2n1-b1n1}{b1(b2+n1)}$ and ($\gamma_{b2} > \frac{b2n1-b2n2}{n2(b2+n2)}$ and 12) $\forall\gamma_i$
- if $b1 < b2$ and $n1 = n2$: (11) if $\gamma_{n1} > \frac{b2n1-b1n1}{b1(b2+n1)}$ and (12) $\forall\gamma_i$
- if $b1 = b2$ and $n1 < n2$: (11) $\forall\gamma_i$ and (12) if $\gamma_{b1} > \frac{b1n2-b1n1}{n1(b1+n2)}$
- if $b1 = b2$ and $n1 > n2$: (11) if $\gamma_{b2} > \frac{b2n1-b2n2}{n2(b2+n2)}$ and (12) $\forall\gamma_i$
- if $b1 = b2$ and $n1 = n2$: (11) and (11) $\forall\gamma_i$

Proposition A1.2: In the game with three strategies and two parameters of tolerance $\nexists$ a distribution of the population that supports the following equilibria.

- $(b1, n1, b2, n2) = (N, SA, SA, SA)$, (SA, N, SA, SA), (SA, SA, N, SA), (SA, SA, SA, N)

  Proof. everybody has an incentive to deviate and "hang around" in her own ward, obtaining a payoff equal to one. □

- $(b1, n1, b2, n2) = (N, HA, HA, HA)$, (HA, N, HA, HA), (HA, HA, N, HA), (HA, HA, HA, N)

  Proof. everybody has an incentive to deviate and play SA, obtaining a payoff equal to one. □

- $(b1, n1, b2, n2) = (SA, HA, SA, N)$, (SA, N, SA, HA), (HA, SA, N, SA), (N, SA, HA, SA)

  Proof. Players who play N have an incentive to play HA, since the payoff obtained by staying at home is strictly less than one. □

- $(b1, n1, b2, n2) = (N, HA, SA, HA)$, (SA, HA, N, HA),(HA, N, HA, SA), (HA, SA, HA, N)

  Proof. Players who plays N have an incentive to play SA, since the payoff obtained by staying at home is strictly less than one. □

- $(b1, n1, b2, n2)$ =(N , N, N, HA), (N, N, HA, N), (N , HA, N, N), (HA, N, N, N), (N, N, N, SA), (N, N, SA, N), (N, SA, N, N), (SA, N, N, N), (N, N, N, N)

  Proof. Players who plays N have an incentive to play either HA SA, since the payoff obtained by staying at home is strictly less than one. □

- $(b1, n1, b2, n2)$=(N, N, SA, HA), (N, N, HA, SA),(N, N, SA, SA), (N, N, HA, HA),

  Proof. Nobody in ward one is involved in social activity, so people of both ethnic groups living in ward 1 have an incentive to play HA □

- $(b1, n1, b2, n2)$=(SA, HA, N, N), (HA, SA, N, N),(SA, SA, N, N), (HA, HA, S, S),

  Proof. Nobody in ward two is involved in social activity, so both people of both ethnic groups living in 2 have an incentive to play HA □

- $(b1, n1, b2, n2)$=(N, HA, N, HA),

  Proof. Both people of ethnic group b living in 1 and people of ethnic group b living in 1 have an incentive to play SA □

- $(b1, n1, b2, n2)$=(N, SA, N, SA),

  Proof. Both people of ethnic group b living in 1 and people of ethnic group b living in 2 have an incentive to play HA □

- $(b1, n1, b2, n2)$=(N, HA, N, SA),

  Proof. People of ethnic group b living in 2 have an incentive to play HA □

- $(b1, n1, b2, n2)$=(N, SA, N, HA),

  Proof. People of ethnic group b living in 1 have an incentive to play HA □

- $(b1, n1, b2, n2)$=(HA, N, HA, N),

  Proof. People of ethnic group n living in both wards 1 and 2 have an incentive to play SA □

- $(b1, n1, b2, n2)$=(SA, N, SA, N),

  Proof. People of ethnic group n living in both wards 1 and 2 have an incentive to play HA □

- $(b1, n1, b2, n2)$=(HA, N, SA, N),

  Proof. People of type n living in ward 2 have an incentive to play HA □

- $(b1, n1, b2, n2)$=(SA, N, HA, N),

  Proof. People of type n living in ward 1 have an incentive to play HA □

- $(b1, n1, b2, n2)$=(N, HA, HA, N), (HA, N, N, HA)

  Proof. The players who play N has an incentive to play SA □

- $(b1, n1, b2, n2)$=(N, SA, SA, N), (SA, N, N, SA)

  Proof. The players who play N has an incentive to play HA □

- $(b1, n1, b2, n2)$=(N, SA, HA, N)

  Proof. People of type b living in ward 1 have an incentive to play HA , while People of type n living in ward 2 have an incentive to play SA □

- $(b1, n1, b2, n2)$=(N, HA, SA, N)

  Proof. People of type b living in ward 1 have an incentive to play SA , while People of type n living in ward 2 have an incentive to play HA □

- $(b1, n1, b2, n2)$=(SA, N, N, HA)

Proof. People of type n living in ward 1 have an incentive to play HA , while People of type b living in ward 2 have an incentive to play SA □

- $(b1, n1, b2, n2)$=(HA, N, N, SA)

  Proof. People of type n living in ward 1 has an incentive to play SA , while people of type b living in ward 2 have an incentive to play HA □

- $(b1, n1, b2, n2)$=(HA, HA, N, SA), (HA, HA, SA, N), (N, SA, HA, HA)

  Proof. The players who stay at home have an incentive to play HA □

- $(b1, n1, b2, n2)$=(N, HH, SA, SA), (SA, SA, HA, N)

  Proof. All the players who play SA have an incentive to play HA □

- $(b1, n1, b2, n2)$=(N, SA, SA, HA)

  Proof. People of type b living in ward 1 have an incentive to play HA □

- $(b1, n1, b2, n2)$=(HA, SA, SA, N)

  Proof. People of type b living in ward 2 has an incentive to play HA □

- $(b1, n1, b2, n2)$=(SA, SA, N, HA), (SA, SA, HA, N)

  Proof. Players of ward 1 have an incentive to play HA □

- $(b1, n1, b2, n2)$=(SA, HA, N, HA)

  Proof. Players who play N have an incentive to play SA □

- $(b1, n1, b2, n2)$=(SA, N, HA, HA)

  Proof. People of type b living in ward 2 have an incentive to play SA □

number and nature of equilibria in a game with three strategies and two parameters of tolerance The following list summarize the nature and the number of the possible equilibria. In addition to (9) and (10), we have:

- **if** $b1 > b2$ **and** $n1 < n2$: (11) if $\overline{U_{n1}} \leq \frac{n1+\gamma b2}{n1+b2}$ and $\overline{U_{b2}} \leq \frac{b2+\gamma n1}{n1+b2}$; (32) if $\overline{U_{n1}} \geq \frac{n1+\gamma b2}{n1+b2}$; (33) if $\overline{U_{b2}} \geq \frac{b2+\gamma n1}{n1+b2}$; (34) if $\overline{U_{b1}} \geq \frac{b1+\gamma n2}{b1+n2}$; (35) if $\overline{U_{n2}} \geq \frac{n2+\gamma b1}{b1+n2}$;
- **if** $b1 > b2$ **and** $n1 > n2$: (32) if $\overline{U_{n1}} \geq \frac{n1+\gamma b2}{n1+b2}$; (33) if $\overline{U_{b2}} \geq \frac{b2+\beta n2}{n2+b2}$; (34) if $\overline{U_{b1}} \geq \frac{b1+\beta n1}{b1+n1}$; (35) if $\overline{U_{n2}} \geq \frac{n2+\beta b2}{b2+n2}$;
- **if** $b1 > b2$ **and** $n1 = n2$: (11) if $\overline{U_{n1}} \leq \frac{n1+\gamma b2}{n1+b2}$ and $\overline{U_{b2}} \leq \frac{b2+\gamma n1}{n1+b2}$;(32) if $\overline{U_{n1}} \geq \frac{n1+\gamma b2}{n1+b2}$; (33) if $\overline{U_{b2}} \geq \frac{b2+\gamma n1}{n1+b2}$; (34) if $\overline{U_{b1}} \geq \frac{b1+\gamma n2}{b1+n2}$; (35) if $\overline{U_{n2}} \geq \frac{n2+\gamma b1}{b1+n2}$;
- **if** $b1 < b2$ **and** $n1 < n2$: (32) if $\overline{U_{n1}} \geq \frac{n1+\beta b1}{n1+b1}$; (33) if $\overline{U_{b2}} \geq \frac{b2+\gamma n1}{n1+b2}$; (34) if $\overline{U_{b1}} \geq \frac{b1+\beta n1}{b1+n1}$; (35) if $\overline{U_{n2}} \geq \frac{n2+\gamma b1}{b1+n2}$;
- **if** $b1 < b2$ **and** $n1 > n2$: (11) if $\overline{U_{n1}} \leq \frac{n1+\gamma b2}{n1+b2}$ and $\overline{U_{b2}} \leq \frac{b2+\gamma n1}{n1+b2}$; (32) if $\overline{U_{n1}} \geq \frac{n1+\gamma b2}{n1+b2}$; (33) if $\overline{U_{b2}} \geq \frac{b2+\gamma n1}{n1+b2}$; (34) if $\overline{U_{b1}} \geq \frac{b1+\beta n1}{b1+n1}$; (35) if $\overline{U_{n2}} \geq \frac{n2+\beta b2}{b2+n2}$;
- **if** $b1 < b2$ **and** $n1 = n2$: (12) if $\overline{U_{n2}} \leq \frac{n2+\gamma b2}{n2+b1}$ and $\overline{U_{b1}} \leq \frac{b1+\gamma n2}{n2+b1}$;(32) if $\overline{U_{n1}} \geq \frac{n1+\beta b1}{n1+b1}$; (33) if $\overline{U_{b2}} \geq \frac{b2+\gamma n1}{n1+b2}$; (34) if $\overline{U_{b1}} \geq \frac{b1+\beta n1}{b1+n1}$; (35) if $\overline{U_{n2}} \geq \frac{n2+\beta b2}{b2+n2}$;
- **if** $b1 = b2$ **and** $n1 < n2$: (11) f $\overline{U_{n1}} \leq \frac{n1+\gamma b2}{n1+b2}$ and $\overline{U_{b2}} \leq \frac{b2+\gamma n1}{n1+b2}$;(32) if $\overline{U_{n1}} \geq \frac{n1+\gamma b2}{n1+b2}$; (33) if $\overline{U_{b2}} \geq \frac{b2+\gamma n1}{n1+b2}$; (34) if $\overline{U_{b1}} \geq \frac{b1+\beta n1}{b1+n1}$; (35) if $\overline{U_{n2}} \geq \frac{n2+\gamma b1}{b1+n2}$;
- if $b1 = b2$ **and** $n1 > n2$: (12) if $\overline{U_{n2}} \leq \frac{n2+\gamma b2}{n2+b1}$ and $\overline{U_{b1}} \leq \frac{b1+\gamma n2}{n2+b1}$;(32) if $\overline{U_{n1}} \geq \frac{n1+\beta b1}{n1+b1}$;32) if $\overline{U_{n1}} \geq \frac{n1+\gamma b2}{n1+b2}$; (33) if $\overline{U_{b2}} \geq \frac{b2+\gamma n1}{n1+b2}$; (34) if $\overline{U_{b1}} \geq \frac{b1+\gamma n2}{b1+n2}$; (35) if $\overline{U_{n2}} \geq \frac{n2+\gamma b1}{b1+n2}$;
- **if** $b1 = b2$ **and** $n1 = n2$: (11) and (12); (32) if $\overline{U_{n1}} \geq \frac{n1+\gamma b2}{n1+b2}$; (33) if $\overline{U_{b2}} \geq \frac{b2+\gamma n1}{n1+b2}$; (34) if $\overline{U_{b1}} \geq \frac{b1+\gamma n2}{b1+n2}$; (35) if $\overline{U_{n2}} \geq \frac{n2+\gamma b1}{b1+n2}$;

Appendix 2: variables definition

We considered two different questions about young people's participation. The question wording is the following:

Question 1:

Here is a list of things people can do when they are not at school. Can you please tell me which, if any, you have been to or done in the last four weeks?

- Played snooker, darts or pool
- Took part in any kind of sport
- Gone to see a football match or other sports event
- Gone to an amusement arcade
- Gone to a party, dance, nightclub or disco
- Gone to a pub or bar
- Gone to a cinema, theatre or concert
- Played a musical instrument
- All of these
- None of these

Question 2:

Here is a list of some more things people do when they are not at school. Can you please tell me which, if any, you have been to or done in the last four weeks? Just read out the numbers.

- Gone to a political meeting/march, rally or demonstration
- Done community work (such as helping elderly, disabled or other dependent people; cleaning up the environment; helping volunteer organizations or charities)
- Gone to a youth club or something like it (including scouts or girl guides)
- Just hubv ng around/messed about near to your home
- Just hung around/messed about in the high street or the town/city centre
- All of these
- None of these

Question 3: 'How many times had taken part in religious classes in last 7 days?'

- more than once a week
- about once a week,

- two or three times a month,
- about once a month
- less than once a month.

Question 4:
'How many times had friend round the house in last 7 days?' The possible outcomes are:

- None
- Once or twice
- 3-5 times
- More than 6 times

Question 5:
How many times gone out with friends in last 7 days?
The possible outcomes are, as before:

- None
- Once or twice
- 3-5 times
- More than 6 times

Appendix 3: Geography and data

The Longitudinal studies of Young People in England contains a variable indicating the output area (OA) where each respondent lives and this permits to link each observation with a huge number of variables describing the neighborhood characteristics measured at different levels of aggregation.

English geographical hierarchy In order to understand the structure of our data, is worth giving some basic information on the English geographical hierarchy. The lowest available level of aggregation is the output area (OA) containing around 150 households which are quite homogeneous in terms of characteristics of the dwelling. The OA are aggregated into LSOAs (Lower Super Output Areas containing around 600 households), then into MSOAs (Middle Super Output Areas containing around 2500 households). The complete structure would include also Upper Super Output Areas (USOAs), which are not available yet.

OA are nested into wards. in England there are almost 8000 wards with an average population of around 5500 individuals, they are used for the election of local government councillors and they are nested into Local Authority districts (LA), which are composed, on average, by 23 electoral wards. The districts can be of four different types: metropolitan districts (36), non-metropolitan districts (239), London Boroughs (32[46] plus the city of London) and unitary authorities (46). Among those, the first three groups have been created as the lower level of two-tier authorities, being higher level authorities in the two-tier structure known as metropolitan and non metropolitan counties in case of metropolitan and non metropolitan districts, while London boroughs are part of Greater London. On the contrary, the unitary authorities are single-tier authorities and they do not belong to any county. Finally, both types of local authorities are grouped into 9 Government Office Regions (GORs[47]), corresponding to level 1 of the 'Nomenclature of Territorial Units for Statistics' (NUTS).

An alternative geography used in England is the postal geography. The UK is divided into many postcode units (about 1,78 million), which are grouped in more than 10000 postcode sectors, then into around 3000 postcode districts and finally into around one hundred postcode areas. The postcodes do not have any geographical meaning and they are simply groups of up to 100 adjacent addresses used for organizing the delivery of the mails. For this reason, we chose to limit our use of the postal geography just to the computation of the distances between cities we used for our instrumental variables. Using the postal geography to derive locational attributes is a common practice and it is due to the fact that the Office of National Statistics assigns centroids to postcodes thus making it possible to compute distances among between each couple of units in the country. The postal geography and the geography used in the census are not exactly compatible, but at our level of aggregation they can be jointly used with limited loss of precision by using the look up tables provided by the ONS.

[46]City of London, Barking and Dagenham, Barnet, Bexley, Brent, Bromley, Camden, Croydon, Ealing, Enfield, Greenwich, Hackney, Hammersmith and Fulham, Haringey, Harrow, Havering, Hillingdon, Hounslow, Islington, Kensington and Chelsea, Kingston upon Thames, Lambeth, Lewisham, Merton, Newham, Redbridge, Richmond upon Thames, Southwark, Sutton, Tower Hamlets, Waltham Forest, Wandsworth, Westminster.

[47]North East, North West, Yorkshire and The Humber, East Midlands, West Midlands, East of England, London, South East, South West

data Sources For our analysis we used data taken from different data sources we merged into the main LSYPE dataset. The indices of ethnic composition are derived the 2001 census for England. Census data permit to compute the exact number of people for each ethnic group living in each ward in 2001. Moreover, we have linked the LSYPE records to the relevant 'index of deprivation affecting children' available through the Pupil Level Annual School Census (PLASC) dataset for 2004. As we have already mentioned, the IDACI index is released at LSOA level so that the figure for the districts must be computed as a weighted average of the single records for al the LSOA in the district. Setting up the system of weights requires the SOA Level population at risk estimates which can be obtained from the 'Department for Communities and Local Government'. This permits to compute at the district level both the number of children actually used to compute the IDACI index and the number of those at risk of being deprived (the denominators in the calculus of the index). Once computed the indices at district level, they have been added to the main data.

Appendix 4: Tables

Table 1: Descriptive statistics for the selected sample

variable	obs	mean	st.dev	min	max
sports amusements	9772	0.852	0.08	0	1
political social	9772	0.247	0.086	0	1
hang about	9772	0.548	0.122	0	1
fractionalization	308	0.152	0.147	0.02	0.676
segregation	308	0.197	0.108	0.058	0.619
idaci	308	0.185	0.091	0.052	0.585
age main parent	9772	41.63	5.982	20	72
no qualification	9772	0.176	0.381	0	1
no dinner	9772	0.089	0.284	0	1
no mother tongue	9772	0.023	0.152	0	1
clubs	9772	0.934	0.246	0	1
male	9772	0.505	0.499	0	1
born in the 1989	9772	0.322	0.467	0	1
income	9772	23378.38	11979.74	0	40000
other white	9772	0.016	0.127	0	1
indian	9772	0.029	0.169	0	1
caribbean	9772	0.028	0.165	0	1
mixed	9772	0.016	0.126	0	1
african	9772	0.018	0.136	0	1
pakistani	9772	0.016	0.126	0	1
bangladeshi	9772	0.006	0.081	0	1
chinese	9772	0.003	0.055	0	1

Table 2: Participation in social activities (%)

Ethnic group	Sports and amusements					Civic activities			Spontaneous participation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Other white and Irish	27.82	66.36	21.11	27.05	58.09	1.59	4.95	16.27	45.51	31.41
Caribbean	18.47	53.25	17.11	36.38	53.00	1.38	3.74	29.32	52.13	28.97
African	15.88	49.69	13.14	24.16	48.23	1.16	3.52	24.12	34.59	19.08
Pakistani	25.36	56.32	10.64	7.85	33.36	2.02	2.58	14.24	30.71	12.97
Bangladeshi	18.89	37.92	11.29	9.81	29.75	1.75	3.20	17.86	31.89	14.25
Indian	24.01	52.19	10.99	17.26	52.13	1.71	3.12	13.11	32.93	20.09
Chinese	17.10	53.09	3.35	17.82	47.64	0	0	16.44	34.88	23.45
British	27.20	50.70	20.63	28.74	49.36	1.23	4.5	20.76	58.47	30.92
other	20.19	60.14	17.17	20.55	53.40	2.19	3.53	14.99	38.03	23.27
mean	26.33	57.85	19.62	27.60	49.25	1.29	4.32	20.43	55.33	29.54

Table 3: Segregation and fractionalization in English districts

District	Region	Index
Least segregated districts		**dissimilarity index**
Castle Point	East Anglia	0.0584
Ashfield	East Midlands	0.064
Sedgemoor	South West	0.0696
Basildon	East Anglia	0.0697
Hertsmere	East Anglia	0.0701
Most segregated districts		
Burnley	rest of north west	0.5207
Bradford	West Yorkshire	0.5242
Pendle	rest of north west	0.5531
Oldham	Greater Manchester	0.5777
Blackburn with Darwen	rest of north west	0.6198
Least fragmented districts		**fractionalization index**
Easington	Rest of North	0.0202
Sedgefield	Rest of North	0.0237
Derwentside	Rest of North	0.024
Wear Valley	Rest of North	0.0261
Alnwick	Rest of North	0.0275
Most fragmented districts		
Lambeth	Inner London	0.6548
Tower Hamlets	Inner London	0.6586
Haringey	Inner London	0.6662
Newham	Inner London	0.6714
Hackney	Inner London	0.6761

Table 4: Composition of the sample and the population

Ethnicity	number	Percentage in the sample	Percentage in the population
Other white and Irish	232	1.51	1.86
Caribbean	1061	6.88	2.87
African	704	4.57	1.97
Pakistani	940	6.10	2.26
Bangladeshi	722	4.68	0.90
Indian	1195	7.75	3.28
Chinese	44	0.29	0.39
British	10103	65.55	84.27
other	411	2.67	2.20

Table 5: Participation in social activities

Activity	Percentage
sports and amusements	
Snooker, Dart, pool	26.33
Sport	57.85
Sport event	19.62
Party, dance	27.60
Cinema	49.25
Civic activities	
Political demonstrations	1.29
community work	4.32
Youth club	20.43
Hang around near home	55.33
excluded activities	
hang around in the city centre	29.54
Pub or bar	15.01
Instrument	23.35
amusement arcade	16.70

Table 6: Results for the multivariate probit on the sub sample of those involved in at least one form of participation

	(1) just ethnic diversity						(2) neigh level variablrd						(3) full model					
	sport amusement		civic activities		hang around		sport amusement		civic activities		hang around		sport amusement		civic activities		hang around	
fract	-0.0712		0.1788	**	-0.9083	***	0.1747		0.1747	*	-1.0563	***	-0.0573		0.0946		-0.4033	***
	(0.122)		(0.085)		(0.088)		(0.135)		(0.103)		(0.103)		(0.143)		(0.125)		(0.123)	
segr	0.0249		-0.3004	***	0.2837	***	0.1322		-0.3011	***	0.2138		0.1322		-0.2709	**	0.5003	***
	(0.154)		(0.089)		(0.102)		(0.134)		(0.106)		(0.151)		(0.185)		(0.133)		(0.159)	
idaci	1.2888	***	-0.6024	***	0.2847	***	-0.7461	***	0.0071		0.4554	**	-0.0586		0.1992		0.1885	
	(0.034)		(0.029)		(0.036)		(0.253)		(0.193)		(0.193)		(0.212)		(0.230)		(0.203)	
age mp							1.3828	***	-0.6028	***	0.2293	***	0.0069	***	0.0015		-0.0093	***
							(0.044)		(0.035)		(0.040)		(0.003)		(0.003)		(0.002)	
no qual													-0.1610	***	-0.0604		0.0869	**
													(0.047)		(0.041)		(0.037)	
no dinner													-0.1851	***	-0.0213		0.1651	***
													(0.054)		(0.057)		(0.049)	
no mother t													-0.0788		-0.2287	***	-0.1922	***
													(0.095)		(0.073)		(0.070)	
no clubs													0.3915	***	0.0269		0.0105	
													(0.074)		(0.058)		(0.057)	
gender													0.3823	***	0.0171		0.1019	***
													(0.031)		(0.027)		(0.024)	
born in 1989													0.0329		-0.0128		0.0404	
													(0.048)		(0.031)		(0.029)	
inc-2nd													0.1177	**	0.0050		0.0632	
													(0.058)		(0.061)		(0.054)	
inc-3rd													0.1223	**	-0.0935	*	-0.0169	
													(0.053)		(0.053)		(0.040)	
inc-4th													0.1717	***	-0.0061		0.0044	
													(0.056)		(0.044)		(0.039)	
inc-5th													0.3355	***	0.0012		-0.1505	***
													(0.059)		(0.049)		(0.033)	
other white													0.5195	***	-0.0237		-0.3043	***
													(0.187)		(0.125)		(0.097)	
asian													0.1756	**	-0.1696	***	-0.5671	***
													(0.081)		(0.060)		(0.057)	
caribbean													0.0883		0.2743	***	-0.1690	**
													(0.074)		(0.064)		(0.070)	
other													0.2849	*	-0.0024		-0.4107	***
													(0.152)		(0.108)		(0.095)	
african													0.2601	**	0.1889	**	-0.4108	***
													(0.105)		(0.083)		(0.063)	
pakistani													0.2590	**	0.0887		-0.7564	***
													(0.114)		(0.098)		(0.079)	
bangladeshi													0.0578		0.1990		-0.4040	***
													(0.121)		(0.129)		(0.092)	
chinese													-0.3188		0.3260		-0.2449	
													(1.086)		(0.279)		(0.412)	
Constant	1.2888	***	-0.6024	***	0.2847	***	1.3828	***	-0.6028	***	0.2293	***	0.3218	**	-0.7057	***	0.5619	***
	(0.034)		(0.029)		(0.036)		(0.044)		(0.035)		(0.040)		(0.140)		(0.159)		(0.112)	
Observations	13719						13718						9292					
Standard	errors	in	par	*	p< .10	**	p< .05	***	p< .01									

Table 7: Results for the multivariate probit on the full sample

	(1) just ethnic diversity						(2) neigh level variables						(3) full model					
	sport amusement		civic activities		hang around		sport amusement		civic activities		hang around		sport amusement		civic activities		hang around	
fract	-0.6478	***	0.0162		-1.0544	***	-0.3850	***	0.0383		-1.1655	***	-0.2634	**	0.0264		-0.4617	***
	(0.176)		(0.094)		(0.064)		(0.141)		(0.116)		(0.118)		(0.124)		(0.132)		(0.129)	
segr	-0.2735		-0.3552	***	0.1346		-0.1820		-0.3418	***	0.0891		0.0168		-0.2701	**	0.4519	***
	(0.198)		(0.097)		(0.118)		(0.186)		(0.095)		(0.166)		(0.174)		(0.118)		(0.137)	
idaci	1.2204	***	-0.6059	***	0.2670	***	-0.7588	***	-0.0681		0.3304	*	0.0713		0.2151		0.2207	
	(0.070)		(0.041)		(0.031)		(0.293)		(0.175)		(0.191)		(0.249)		(0.237)		(0.195)	
age mp							1.3177	***	-0.5983	***	0.2261	***	0.0054	**	0.0013		-0.0087	***
							(0.059)		(0.032)		(0.039)		(0.003)		(0.002)		(0.002)	
no qual													-0.1496	***	-0.0628		0.0735	**
													(0.045)		(0.043)		(0.031)	
no dinner													-0.1560	***	-0.0248		0.1538	***
													(0.059)		(0.045)		(0.046)	
no mmother t													-0.2439	***	-0.2724	***	-0.2495	***
													(0.059)		(0.066)		(0.065)	
no clubs													0.3860	***	0.0458		0.0460	
													(0.060)		(0.054)		(0.047)	
gender													0.4570	***	0.0532	**	0.1503	***
													(0.029)		(0.027)		(0.028)	
born in 1989													0.0440		-0.0075		0.0461	
													(0.031)		(0.035)		(0.030)	
inc-2nd													0.1087	**	0.0135		0.0709	*
													(0.048)		(0.047)		(0.039)	
inc-3rd													0.1266	**	-0.0773		0.0022	
													(0.050)		(0.053)		(0.041)	
inc-4th													0.2091	***	0.0134		0.0359	
													(0.052)		(0.042)		(0.042)	
inc-5th													0.3754	***	0.0302		-0.1016	***
													(0.058)		(0.048)		(0.035)	
other white													0.4963	***	-0.0021		-0.2664	**
													(0.174)		(0.124)		(0.123)	
asian													-0.0063		-0.1833	***	-0.5803	***
													(0.066)		(0.066)		(0.049)	
caribbean													0.0831		0.2761	***	-0.1523	**
													(0.077)		(0.073)		(0.061)	
other													0.1034		-0.0216		-0.4152	***
													(0.116)		(0.105)		(0.089)	
african													0.1260		0.1697	**	-0.4131	***
													(0.080)		(0.084)		(0.072)	
pakistani													-0.1952	***	-0.0107		-0.8281	***
													(0.064)		(0.107)		(0.066)	
bangladeshi													-0.2874	***	0.0695		-0.5215	***
													(0.081)		(0.116)		(0.092)	
chinese													-0.4455	*	0.2050		-0.3641	
													(0.265)		(0.322)		(0.278)	
Constant	1.2204	***	-0.6059	***	0.2670	***	1.3177	***	-0.5983	***	0.2261	***	0.2274		-0.7640	***	0.4196	***
	(0.070)		(0.041)		(0.031)		(0.059)		(0.032)		(0.039)		(0.168)		(0.107)		(0.098)	
Observations	14692						14691						9772					
Standard	errors	in	par	*	p¡.10	**	p¡.05	***	p¡.01									

Table 8: Correlations between the dependent variables

	sub sample					
	just ethnic		neighborhood		full	
ρ_{21}	0.0712	***	0.0767	***	0.0588	***
	(0.018)		(0.018)		(0.023)	
ρ_{31}	-0.0985	***	-0.0924	***	-0.0703	***
	(0.015)		(0.015)		(0.021)	
ρ_{32}	-0.0551	***	-0.0535	***	-0.0526	***
	(0.014)		(0.015)		(0.018)	
Observations	13719		13718		9292	
	full sample					
	just ethnic		neighborhood		full	
ρ_{21}	0.2045	***	0.2130	***	0.1741	***
	(0.013)		(0.020)		(0.018)	
ρ_{31}	0.1759	***	0.1699	***	0.1307	***
	(0.026)		(0.014)		(0.019)	
ρ_{32}	0.0151		0.0135		0.0033	
	(0.018)		(0.016)		(0.018)	
Observations	14692		14691		9772	
*	p<.10	**	p<.05	***	p<.01	

Table 9: First stage regression (linear, means at the district level)

	just segregation endogenous		both indices endogenous			
	segregation		segregation		fractionalization	
dist mill towns	-0.0003	***	-0.0002	***	0.0000	
dist ports			0.0002		-0.0003	**
fract	0.3940	***				
idaci	0.2912	***	0.3420	***	0.0775	
no dinner	0.0337		0.0280		-0.0088	
no mother tongue	-0.1546		0.0503		0.5359	***
no clubs	0.0463		0.0339		-0.0124	
other white	-0.1865		-0.0339		0.3825	***
indian	0.3337	***	0.4830	***	0.4248	***
caribbean	-0.3234	**	0.0516		0.9927	***
mixed	-0.1051		0.0033		0.2227	
african	-0.6125	***	-0.3610	**	0.5440	***
pakistani	0.6465	***	0.8241	***	0.4220	***
bangladeshi	-0.4416	***	-0.1869		0.6471	***
chinese	0.7073		0.7117		-0.0348	
fam controls	yes		yes		yes	
ind controls	yes		yes		yes	
Constant	0.2802	**	0.1502		-0.2408	**
F test on the excluded instruments	14.83		29.00		9.35	
Observations	308		308			

Table 10: Second stages on the sub sample of those involved in at least one form of participation

	just segregation endogenous						both endogenous					
	sport amusement		civic activities		hang around		sport amusement		civic activities		hang around	
fract	-0.0657		0.1520		-0.4495	***	-0.1432		0.1910		-0.4651	***
	(0.157)		(0.126)		(0.133)		(0.184)		(0.175)		(0.150)	
segr	0.2177		-0.7288	***	0.8372	***	0.1091		-0.6474	***	0.8534	***
	(0.290)		(0.220)		(0.235)		(0.299)		(0.250)		(0.248)	
Residuals segr	-0.1675		0.8613	***	-0.6341		-0.0771		0.7766	**	-0.6877	*
	(0.398)		(0.321)		(0.403)		(0.424)		(0.333)		(0.407)	
residuals fragm							0.4119		-0.5048		0.3238	
							(0.417)		(0.382)		(0.335)	
idaci	-0.0920		0.3542		0.0775		-0.0055		0.3031		0.0755	
	(0.215)		(0.232)		(0.240)		(0.230)		(0.231)		(0.247)	
age mp	0.0070	***	0.0015		-0.0092	***	0.0070	***	0.0014		-0.0092	***
	(0.003)		(0.003)		(0.002)		(0.003)		(0.003)		(0.002)	
no qual	-0.1599	***	-0.0635		0.0888	**	-0.1587	***	-0.0645		0.0893	**
	(0.047)		(0.041)		(0.037)		(0.047)		(0.042)		(0.037)	
no dinner	-0.1845	***	-0.0204		0.1641	***	-0.1846	***	-0.0208		0.1645	***
	(0.054)		(0.057)		(0.049)		(0.054)		(0.056)		(0.048)	
no mmother t	-0.0788		-0.2327	***	-0.1900	***	-0.0762		-0.2342	***	-0.1890	***
	(0.095)		(0.074)		(0.071)		(0.095)		(0.074)		(0.070)	
no clubs	0.3900	***	0.0283		0.0096		0.3896	***	0.0288		0.0093	
	(0.074)		(0.058)		(0.057)		(0.074)		(0.058)		(0.057)	
gender	0.3832	***	0.0162		0.1037	***	0.3841	***	0.0160		0.1037	***
	(0.032)		(0.027)		(0.024)		(0.032)		(0.027)		(0.024)	
born in 1989	0.0332		-0.0111		0.0388		0.0333		-0.0109		0.0385	
	(0.048)		(0.032)		(0.030)		(0.048)		(0.032)		(0.030)	
inc-2nd	0.1184	**	0.0068		0.0617		0.1190	**	0.0069		0.0618	
	(0.058)		(0.062)		(0.053)		(0.058)		(0.062)		(0.053)	
inc-3rd	0.1230	**	-0.0920	*	-0.0185		0.1230	**	-0.0919	*	-0.0187	
	(0.053)		(0.053)		(0.039)		(0.053)		(0.053)		(0.039)	
inc-4th	0.1742	***	-0.0063		0.0037		0.1740	***	-0.0061		0.0036	
	(0.056)		(0.045)		(0.038)		(0.056)		(0.045)		(0.038)	
inc-5th	0.3365	***	-0.0002		-0.1489	***	0.3381	***	-0.0013		-0.1483	***
	(0.059)		(0.049)		(0.033)		(0.059)		(0.049)		(0.033)	
other white	0.5250	***	-0.0370		-0.2934	***	0.5316	***	-0.0396		-0.2918	***
	(0.186)		(0.127)		(0.096)		(0.187)		(0.127)		(0.096)	
asian	0.1723	**	-0.1507	**	-0.5786	***	0.1887	**	-0.1592	**	-0.5782	***
	(0.082)		(0.060)		(0.057)		(0.083)		(0.065)		(0.056)	
caribbean	0.0908		0.2682	***	-0.1640	**	0.1022		0.2636	***	-0.1625	**
	(0.075)		(0.063)		(0.071)		(0.072)		(0.068)		(0.071)	
other	0.2844	*	-0.0057		-0.4039	***	0.2927	*	-0.0104		-0.4028	***
	(0.153)		(0.108)		(0.095)		(0.155)		(0.111)		(0.098)	
african	0.2634	**	0.1708	**	-0.3981	***	0.2748	***	0.1659	*	-0.3952	***
	(0.107)		(0.083)		(0.063)		(0.106)		(0.087)		(0.063)	
pakistani	0.2498	**	0.1282		-0.7841	***	0.2624	**	0.1205		-0.7859	***
	(0.116)		(0.102)		(0.076)		(0.113)		(0.105)		(0.074)	
bangladeshi	0.0653		0.1648		-0.3798	***	0.0762		0.1607		-0.3761	***
	(0.126)		(0.123)		(0.093)		(0.127)		(0.123)		(0.092)	
chinese	-0.3215		0.3390		-0.2586		-0.3122		0.3347		-0.2586	
	(1.091)		(0.276)		(0.408)		(1.087)		(0.278)		(0.403)	
Constant	0.3096	**	-0.6461	***	0.5174	***	0.3218	**	-0.6568	***	0.5156	***
	(0.151)		(0.157)		(0.116)		(0.156)		(0.159)		(0.117)	
Observations	9292						9292					
Standard	errors	in	par.	*	p<.10	**	p<.05	***	p<.01			

Table 11: Second stages on the full sample

	just segregation endogenous						both endogenous					
	sport amusement		civic activities		hang around		sport amusement		civic activities		hang around	
fract	-0.2704	**	0.0861		-0.4995	***	-0.3546	**	0.1100		-0.5277	***
	(0.133)		(0.130)		(0.129)		(0.170)		(0.168)		(0.119)	
segr	0.0690		-0.6974	***	0.7725	***	-0.0449		-0.6400	***	0.7673	***
	(0.293)		(0.162)		(0.240)		(0.281)		(0.202)		(0.254)	
Residuals segr	-0.1235		0.8130	***	-0.6022	*	-0.0296		0.7512	**	-0.6332	**
	(0.452)		(0.311)		(0.312)		(0.458)		(0.340)		(0.296)	
residuals fragm							0.4297		-0.4211		0.3661	
							(0.413)		(0.286)		(0.315)	
idaci	0.0489		0.3591		0.1080		0.1432		0.3256		0.1223	
	(0.245)		(0.230)		(0.209)		(0.276)		(0.233)		(0.221)	
age mp	0.0055	**	0.0014		-0.0084	***	0.0056	**	0.0014		-0.0084	***
	(0.003)		(0.002)		(0.002)		(0.003)		(0.002)		(0.002)	
no qual	-0.1484	***	-0.0644		0.0742	**	-0.1465	***	-0.0650		0.0749	**
	(0.045)		(0.043)		(0.031)		(0.045)		(0.044)		(0.032)	
no dinner	-0.1562	***	-0.0302		0.1495	***	-0.1568	***	-0.0305		0.1497	***
	(0.059)		(0.045)		(0.047)		(0.058)		(0.045)		(0.046)	
no mmother t	-0.2424	***	-0.2794	***	-0.2484	***	-0.2385	***	-0.2805	***	-0.2469	***
	(0.059)		(0.064)		(0.066)		(0.059)		(0.065)		(0.065)	
no clubs	0.3823	***	0.0456		0.0435		0.3809	***	0.0461		0.0428	
	(0.060)		(0.055)		(0.047)		(0.060)		(0.055)		(0.047)	
gender	0.4561	***	0.0536	**	0.1503	***	0.4563	***	0.0536	**	0.1502	***
	(0.029)		(0.027)		(0.027)		(0.029)		(0.027)		(0.027)	
born in 1989	0.0437		-0.0087		0.0437		0.0443		-0.0084		0.0435	
	(0.031)		(0.035)		(0.031)		(0.031)		(0.035)		(0.031)	
inc-2nd	0.1099	**	0.0155		0.0706	*	0.1104	**	0.0154		0.0708	*
	(0.049)		(0.047)		(0.039)		(0.049)		(0.047)		(0.039)	
inc-3rd	0.1272	**	-0.0763		0.0009		0.1271	**	-0.0763		0.0007	
	(0.050)		(0.052)		(0.041)		(0.050)		(0.052)		(0.041)	
inc-4th	0.2072	***	0.0136		0.0352		0.2066	***	0.0137		0.0351	
	(0.052)		(0.041)		(0.043)		(0.052)		(0.041)		(0.043)	
inc-5th	0.3749	***	0.0270		-0.1045	***	0.3763	***	0.0262		-0.1038	***
	(0.057)		(0.047)		(0.036)		(0.057)		(0.047)		(0.036)	
other white	0.4935	***	-0.0069		-0.2531	**	0.5019	***	-0.0083		-0.2505	**
	(0.174)		(0.122)		(0.122)		(0.174)		(0.123)		(0.123)	
asian	-0.0068		-0.1721	***	-0.5980	***	0.0115		-0.1774	***	-0.5945	***
	(0.064)		(0.062)		(0.048)		(0.065)		(0.067)		(0.048)	
caribbean	0.0869		0.2679	***	-0.1503	**	0.0988		0.2653	***	-0.1471	**
	(0.077)		(0.071)		(0.062)		(0.076)		(0.076)		(0.062)	
other	0.1034		-0.0218		-0.4188	***	0.1124		-0.0249		-0.4164	***
	(0.116)		(0.105)		(0.089)		(0.116)		(0.106)		(0.089)	
african	0.1252		0.1540	*	-0.3982	***	0.1376	*	0.1511	*	-0.3936	***
	(0.079)		(0.084)		(0.071)		(0.078)		(0.087)		(0.070)	
pakistani	-0.1993	***	0.0281		-0.8588	***	-0.1859	***	0.0226		-0.8582	***
	(0.065)		(0.110)		(0.067)		(0.064)		(0.112)		(0.066)	
bangladeshi	-0.2855	***	0.0351		-0.5049	***	-0.2755	***	0.0328		-0.4999	***
	(0.085)		(0.108)		(0.098)		(0.085)		(0.108)		(0.098)	
chinese	-0.4616	*	0.2418		-0.3507		-0.4535	*	0.2391		-0.3494	
	(0.265)		(0.318)		(0.274)		(0.264)		(0.317)		(0.272)	
Constant	0.2208		-0.7136	***	0.3747	***	0.2334		-0.7214	***	0.3758	***
	(0.173)		(0.109)		(0.097)		(0.174)		(0.113)		(0.098)	
Observations	9772						9772					
Standard	errors	in	par.	*	p<.10	**	p<.05	***	p<.01			

Table 12: Second stages: correlations between the dependent variables

subsample				
	just segregation		both indices	
	endogenous		endogenous	
ρ_{21}	0.0526	**	0.0526	**
	(0.022)		(0.022)	
ρ_{31}	-0.0742	***	-0.0743	***
	(0.018)		(0.018)	
ρ_{32}	-0.0631	***	-0.0629	***
	(0.016)		(0.017)	
Observations	9292		9292	
full sample				
ρ_{21}	0.1865	***	0.1865	***
	(0.020)		(0.020)	
ρ_{31}	0.1430	***	0.1429	***
	(0.018)		(0.019)	
ρ_{32}	-0.0140		-0.0139	
	(0.017)		(0.017)	
Observations	9772		9772	
Standard	errors	in	parentheses	
*p<.10	**	p<.05	***	p<.01

Table 13: Results for the seemingly unrelated models

	sub-sample						full sample					
	sports amusements		civic activities		hang about		sports amusements		civic activities		hang about	
fract	-0.0059		0.0329		-0.1547	***	-0.0531		0.0129		-0.1713	***
	(-0.025)		(-0.052)		(-0.049)		(-0.036)		(-0.045)		(-0.052)	
segr	0.0197		-0.0894	*	0.188	***	0.0034		-0.0848	**	0.1695	***
	(-0.028)		(-0.047)		(-0.053)		(-0.029)		(-0.042)		(-0.054)	
idaci	-0.0124		0.0695		0.0708		0.0117		0.0695		0.0802	
	(-0.037)		(-0.071)		(-0.061)		(-0.054)		(-0.07)		(-0.07)	
age main parent	0.0012	**	0.0005		-0.0035	***	0.0011	*	0.0005		-0.0032	***
	(-0.001)		(-0.001)		(-0.001)		(-0.001)		(-0.001)		(-0.001)	
no qualification	-0.0302	***	-0.0198		0.0322	**	-0.0357	***	-0.0199		0.0267	**
	(-0.008)		(-0.012)		(-0.014)		(-0.012)		(-0.013)		(-0.011)	
no dinner	-0.0354	***	-0.0073		0.062	***	-0.0356	**	-0.0071		0.0583	***
	(-0.013)		(-0.013)		(-0.017)		(-0.014)		(-0.016)		(-0.018)	
no mother tongue	-0.0099		-0.0718	***	-0.0723	**	-0.0688	***	-0.0788	***	-0.0906	***
	(-0.015		(-0.023)		(-0.031)		(-0.015)		(-0.023)		(-0.024)	
no clubs	0.0827	***	0.0083		0.0037		0.0997	***	0.0144		0.0173	
	(-0.015		(-0.021)		(-0.022)		(-0.019)		(-0.015)		(-0.019)	
gender	0.0623	***	0.0051		0.0392	***	0.0965	***	0.0161	*	0.0566	***
	(-0.006		(-0.01)		(-0.01)		(-0.007)		(-0.009)		(-0.011)	
born in 1989	0.0048		-0.0039		0.0152		0.0086		-0.0022		0.0173	**
	(-0.006		(-0.01)		(-0.01)		(-0.007)		(-0.01)		(-0.009)	
inc-2nd	0.0237	**	0.0018		0.0229		0.0279	**	0.0038		0.026	
	(-0.011		(-0.016)		(-0.018)		(-0.012)		(-0.015)		(-0.016)	
inc-3rd	0.0235	**	-0.0297	**	-0.0072		0.0322	***	-0.0243		0.0004	
	(-0.01		(-0.015)		(-0.014)		(-0.012)		(-0.015)		(-0.016)	
inc-4th	0.0316	***	-0.0019		0.0014		0.048	***	0.0041		0.0128	
	(-0.011)		(-0.013)		(-0.016)		(-0.012)		(-0.013)		(-0.018)	
inc-5th	0.055	***	-0.0001		-0.0573	***	0.077	***	0.0076		-0.0404	**
	(-0.01)		(-0.016)		(-0.014)		(-0.011)		(-0.016)		(-0.017)	
other white	0.0688	***	-0.0082		-0.1193	**	0.0882	***	-0.0026		-0.1071	**
	(-0.019)		(-0.036)		(-0.047)		(-0.019)		(-0.037)		(-0.044)	
indian	0.0264	**	-0.0506	**	-0.2195	***	-0.0012		-0.0535	***	-0.2264	***
	(-0.011)		(-0.02)		(-0.022)		(-0.017)		(-0.017)		(-0.018)	
caribbean	0.0149		0.0962	***	-0.064	***	0.0183		0.0944	***	-0.0602	***
	(-0.012)		(-0.026)		(-0.023)		(-0.016)		(-0.022)		(-0.022)	
mixed	0.0436	**	-0.0023		-0.1587	***	0.0213		-0.0059		-0.1626	***
	(-0.019)		(-0.032)		(-0.035)		(-0.024)		(-0.03)		(-0.036)	
african	0.0398	***	0.0639	**	-0.1601	***	0.0246		0.0558	**	-0.1628	***
	(-0.014)		(-0.03)		(-0.027)		(-0.017)		(-0.027)		(-0.027)	
pakistani	0.0392	***	0.0275		-0.2893	***	-0.057	***	-0.0013		-0.3134	***
	(-0.014)		(-0.032)		(-0.027)		(-0.02)		(-0.025)		(-0.025)	
bangladeshi	0.0034		0.0658		-0.1557	***	-0.1001	***	0.0239		-0.2023	***
	(-0.024)		(-0.04)		(-0.037)		(-0.024)		(-0.037)		(-0.034)	
chinese	-0.0552		0.1094		-0.0954		-0.1145		0.0715		-0.1364	
	(-0.083)		(-0.094)		(-0.129)		(-0.084)		(-0.085)		(-0.106)	
Constant	0.7161	***	0.2402	***	0.7155	***	0.6517	***	0.2194	***	0.6599	***
	(-0.032)		(-0.042)		(-0.04)		(-0.042)		(-0.04)		(-0.039)	
Observations	9292						9772					
st	errors	in	par	*	$p < .10$	**	$p < .05$	***	$p < .01$			

[48]The data in tables 2, 1, 4 and 5 are weighted. In table 8 the variable 1 is 'sports and amusements', the variable 2 is 'civic activities' and the variable 3 is 'hanging around'. In table 9 the covariates are means at the district level of all individual controls in tables 7 and 8. In tables 6, 7, 8, 9, 10, 11, 12 standard errors are clustered by districts and, in tables 10, 11, 12, 13, 14, 15 they are also bootstrapped to account for the generated variable problem.

Table 14: Results for the seemingly unrelated models: second stages: sub sample

	just segregation endogenous						both endogenous					
	sports amusements		civic activities		hang about		sports amusements		civic activities		hang about	
fract	-0.0077		0.0524		-0.1735	***						
	(0.026)		(0.048)		(0.053)							
fract (fitted)							-0.0204		0.0688		-0.1793	***
							(0.031)		(0.051)		(0.055)	
segr (fitted)	0.0330		-0.2413	***	0.3261	***	0.0232		-0.2243	***	0.3271	***
	(0.045)		(0.075)		(0.087)		(0.044)		(0.068)		(0.088)	
idaci	-0.0172		0.1235		0.0208		-0.0043		0.1041		0.0206	
	(0.038)		(0.081)		(0.088)		(0.042)		(0.069)		(0.077)	
age main parent	0.0012	**	0.0004		-0.0035	***	0.0012	**	0.0004		-0.0034	***
	(0.000)		(0.001)		(0.001)		(0.001)		(0.001)		(0.001)	
no qualification	-0.0301	***	-0.0200	*	0.0329	**	-0.0301	***	-0.0198		0.0332	**
	(0.009)		(0.011)		(0.014)		(0.008)		(0.013)		(0.015)	
no dinner	-0.0354	***	-0.0076		0.0624	***	-0.0354	***	-0.0076		0.0626	***
	(0.012)		(0.017)		(0.021)		(0.013)		(0.017)		(0.019)	
no mother tongue	-0.0098		-0.0735	***	-0.0706	***	-0.0094		-0.0739	***	-0.0701	**
	(0.016)		(0.022)		(0.027)		(0.016)		(0.021)		(0.028)	
no clubs	0.0826	***	0.0086		0.0030		0.0826	***	0.0088		0.0029	
	(0.018)		(0.020)		(0.025)		(0.015)		(0.019)		(0.022)	
gender	0.0623	***	0.0051		0.0393	***	0.0622	***	0.0053		0.0394	***
	(0.007)		(0.010)		(0.011)		(0.005)		(0.010)		(0.010)	
born in 1989	0.0047		-0.0037		0.0146		0.0048		-0.0037		0.0145	
	(0.007)		(0.009)		(0.012)		(0.007)		(0.011)		(0.011)	
inc-2nd	0.0237	*	0.0020		0.0229		0.0236	**	0.0021		0.0230	
	(0.012)		(0.019)		(0.018)		(0.012)		(0.017)		(0.016)	
inc-3rd	0.0236	**	-0.0295	**	-0.0071		0.0235	**	-0.0294	*	-0.0071	
	(0.012)		(0.014)		(0.014)		(0.011)		(0.015)		(0.015)	
inc-4th	0.0317	***	-0.0019		0.0015		0.0317	***	-0.0019		0.0014	
	(0.011)		(0.016)		(0.016)		(0.011)		(0.013)		(0.015)	
inc-5th	0.0550	***	-0.0003		-0.0573	***	0.0552	***	-0.0007		-0.0571	***
	(0.011)		(0.015)		(0.016)		(0.011)		(0.013)		(0.014)	
other white	0.0691	***	-0.0126		-0.1153	***	0.0706	***	-0.0145		-0.1149	***
	(0.020)		(0.041)		(0.044)		(0.018)		(0.031)		(0.040)	
indian	0.0260	**	-0.0441	**	-0.2239	***	0.0284	***	-0.0475	***	-0.2238	***
	(0.011)		(0.019)		(0.025)		(0.010)		(0.018)		(0.024)	
caribbean	0.0151		0.0945	***	-0.0617	***	0.0175		0.0909	***	-0.0617	***
	(0.013)		(0.027)		(0.021)		(0.013)		(0.023)		(0.022)	
mixed	0.0435	**	-0.0022		-0.1594	***	0.0454	**	-0.0052		-0.1592	***
	(0.021)		(0.027)		(0.040)		(0.018)		(0.033)		(0.038)	
african	0.0403	***	0.0575	*	-0.1548	***	0.0428	***	0.0540	*	-0.1540	***
	(0.014)		(0.031)		(0.029)		(0.014)		(0.030)		(0.028)	
pakistani	0.0381	**	0.0419		-0.3010	***	0.0407	***	0.0369		-0.3024	***
	(0.016)		(0.031)		(0.033)		(0.013)		(0.032)		(0.030)	
bangladeshi	0.0044		0.0557		-0.1459	***	0.0064		0.0531		-0.1448	***
	(0.022)		(0.038)		(0.042)		(0.022)		(0.044)		(0.043)	
chinese	-0.0557		0.1140		-0.1001		-0.0543		0.1119		-0.1001	
	(0.085)		(0.104)		(0.116)		(0.074)		(0.112)		(0.125)	
Constant	0.7145	***	0.2604	***	0.6990	***	0.7155	***	0.2585	***	0.6988	***
	(0.035)		(0.042)		(0.045)		(0.032)		(0.046)		(0.035)	
Observations	9292						9292					
st	errors	in	par	*	$p < .10$	**	$p < .05$	***	$p < .01$			

Table 15: Results for the seemingly unrelated models: second stages: full sample

	just segregation endogenous						both endogenous					
	sports amusements		civic activities		hang about		sports amusements		civic activities		hang about	
fract	-0.0542		0.0300		-0.1889	***						
	(0.038)		(0.046)		(0.048)							
fract (fitted)							-0.0773	*	0.0433		-0.1977	***
							(0.045)		(0.049)		(0.052)	
segr (fitted)	0.0128		-0.2197	***	0.2964	***	-0.0143		-0.2081	***	0.2894	***
	(0.067)		(0.056)		(0.101)		(0.061)		(0.073)		(0.099)	
idaci	0.0086		0.1153	*	0.0367		0.0334		0.0997		0.0411	
	(0.056)		(0.059)		(0.078)		(0.064)		(0.069)		(0.082)	
age main parent	0.0011	**	0.0004		-0.0032	***	0.0012	**	0.0004		-0.0032	***
	(0.001)		(0.001)		(0.001)		(0.001)		(0.001)		(0.001)	
no qualification	-0.0357	***	-0.0200		0.0273	*	-0.0353	***	-0.0197	*	0.0277	**
	(0.012)		(0.013)		(0.014)		(0.010)		(0.011)		(0.013)	
no dinner	-0.0356	**	-0.0073		0.0585	***	-0.0357	**	-0.0072		0.0586	***
	(0.014)		(0.014)		(0.015)		(0.016)		(0.015)		(0.018)	
no mother tongue	-0.0687	***	-0.0799	***	-0.0894	***	-0.0677	***	-0.0801	***	-0.0887	***
	(0.017)		(0.019)		(0.024)		(0.018)		(0.024)		(0.026)	
no clubs	0.0997	***	0.0147		0.0165		0.0993	***	0.0148		0.0163	
	(0.015)		(0.017)		(0.020)		(0.017)		(0.015)		(0.022)	
gender	0.0965	***	0.0161	*	0.0567	***	0.0965	***	0.0163	**	0.0568	***
	(0.007)		(0.009)		(0.011)		(0.007)		(0.008)		(0.010)	
born in 1989	0.0086		-0.0019		0.0168	*	0.0086		-0.0019		0.0167	*
	(0.007)		(0.008)		(0.009)		(0.006)		(0.009)		(0.009)	
inc-2nd	0.0279	**	0.0041		0.0257		0.0278	**	0.0042		0.0259	*
	(0.012)		(0.016)		(0.016)		(0.012)		(0.017)		(0.015)	
inc-3rd	0.0321	***	-0.0239	*	0.0003		0.0321	***	-0.0238		0.0002	
	(0.012)		(0.014)		(0.015)		(0.011)		(0.015)		(0.015)	
inc-4th	0.0480	***	0.0041		0.0129		0.0479	***	0.0041		0.0127	
	(0.012)		(0.014)		(0.017)		(0.012)		(0.014)		(0.015)	
inc-5th	0.0770	***	0.0074		-0.0405	***	0.0773	***	0.0071		-0.0401	***
	(0.012)		(0.014)		(0.015)		(0.011)		(0.014)		(0.014)	
other white	0.0884	***	-0.0065		-0.1034	**	0.0906	***	-0.0082		-0.1029	**
	(0.025)		(0.034)		(0.045)		(0.021)		(0.037)		(0.040)	
indian	-0.0017		-0.0476	***	-0.2305	***	0.0030		-0.0504	***	-0.2296	***
	(0.016)		(0.018)		(0.022)		(0.015)		(0.016)		(0.018)	
caribbean	0.0184		0.0930	***	-0.0581	**	0.0218		0.0896	***	-0.0580	**
	(0.014)		(0.022)		(0.025)		(0.015)		(0.021)		(0.025)	
mixed	0.0213		-0.0060		-0.1626	***	0.0238		-0.0087		-0.1623	***
	(0.026)		(0.031)		(0.033)		(0.025)		(0.028)		(0.040)	
african	0.0250		0.0502	*	-0.1577	***	0.0286		0.0469		-0.1568	***
	(0.018)		(0.030)		(0.025)		(0.022)		(0.030)		(0.026)	
pakistani	-0.0579	**	0.0121		-0.3248	***	-0.0541	***	0.0075		-0.3257	***
	(0.023)		(0.024)		(0.022)		(0.019)		(0.028)		(0.026)	
bangladeshi	-0.0995	***	0.0151		-0.1936	***	-0.0965	***	0.0127		-0.1925	***
	(0.026)		(0.027)		(0.038)		(0.028)		(0.033)		(0.038)	
chinese	-0.1147		0.0744		-0.1390		-0.1126		0.0725		-0.1388	
	(0.098)		(0.086)		(0.107)		(0.088)		(0.087)		(0.102)	
Constant	0.6504	***	0.2373	***	0.6448	***	0.6536	***	0.2361	***	0.6457	***
	(0.037)		(0.040)		(0.044)		(0.037)		(0.041)		(0.041)	
Observations	9772						9772					

st errors in par * $p < .10$ ** $p < .05$ *** $p < .01$

Appendix 4: Graphs

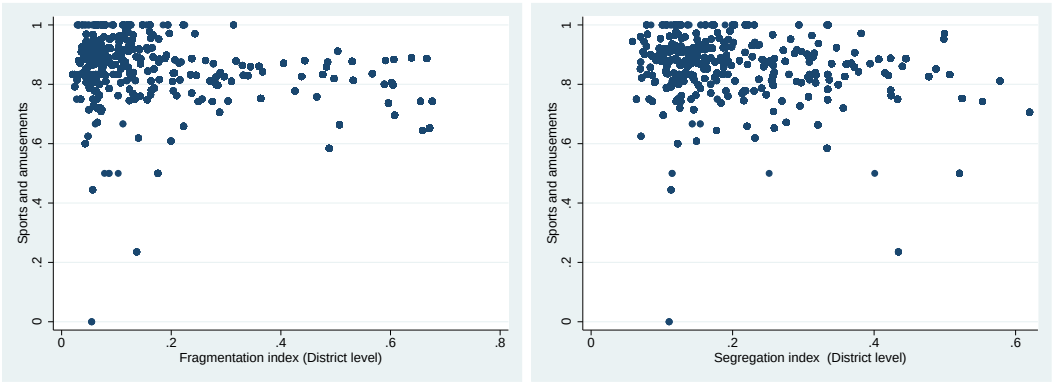

Figure 1: Ethnic diversity and participation in sports-amusements

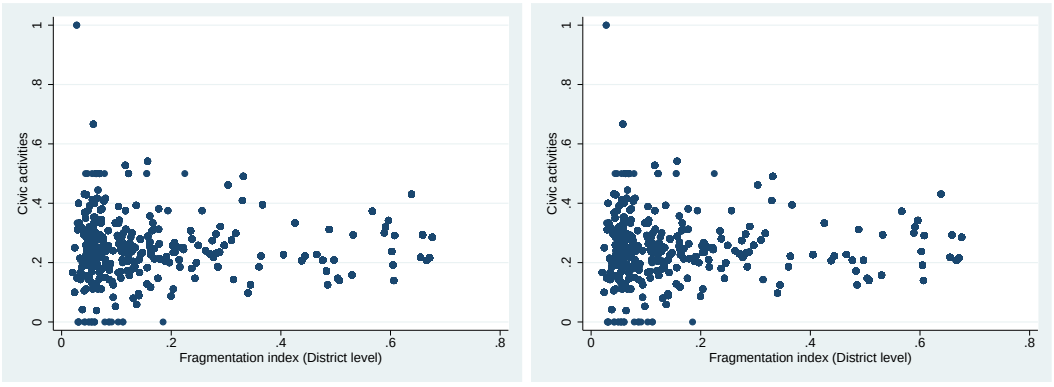

Figure 2: Ethnic diversity and civic participation

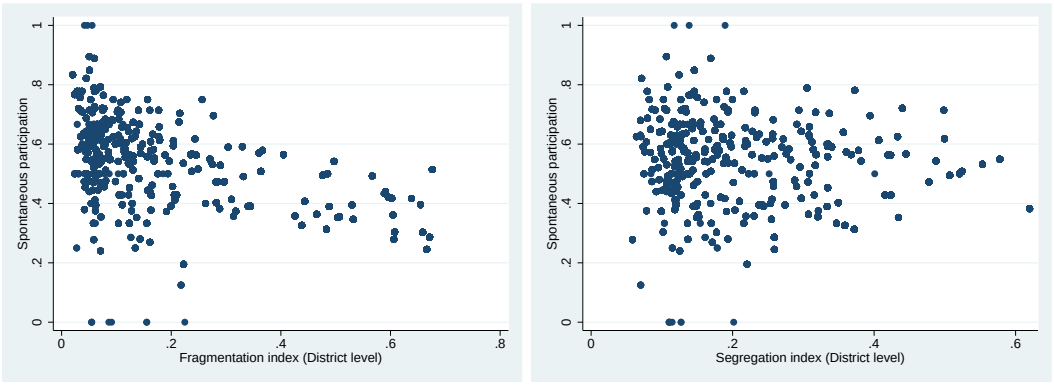

Figure 3: Ethnic diversity and spontaneous participation