



TI 2001-028/3

Tinbergen Institute Discussion Paper

# **Effects of Ethnic Geographical Clustering on Educational Attainment in the Netherlands**

*Thomas de Graaff*

*Cees Gorter*

*Peter Nijkamp*

### **Tinbergen Institute**

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

### **Tinbergen Institute Amsterdam**

Keizersgracht 482  
1017 EG Amsterdam  
The Netherlands  
Tel.: +31.(0)20.5513500  
Fax: +31.(0)20.5513555

### **Tinbergen Institute Rotterdam**

Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31.(0)10.4088900  
Fax: +31.(0)10.4089031

Most recent TI discussion papers can be downloaded at  
<http://www.tinbergen.nl>

# Effects of Ethnic Geographical Clustering on Educational Attainment in the Netherlands

Thomas de Graaff, Cees Gorter and Peter Nijkamp  
Department of Spatial Economics and Tinbergen Institute  
Free University, Amsterdam  
De Boelelaan 1105, 1081 HV Amsterdam  
E-mail: tgraaff@econ.vu.nl  
Tel: +31 (0)20 4446098

February 26, 2000

## **Abstract**

Contemporary migration studies witness an increasing interest in the socio-economic role of networks of migrants. Such networks are sometimes even regarded as the most important attraction and location factors for migration, and may even exceed purely economic factors like unemployment and wage levels in importance. The empirical measurement and analysis of migrants' networks however, is far from easy. Usually, the size of immigrant networks in a city is proxied by means of the share of foreigners, while much less attention is given to the spatial distribution of immigrants. This paper aims to address the empirical assessment of spatial clustering of socio-cultural groups in the city. It does so by modifying a geographical concentration measure developed by Ellison and Glaeser (the gamma coefficient), with a view to the measurement of spatial clustering of migrants in the Netherlands. Because of the scale-independent character of the gamma coefficient, we are able to investigate the degree of ethnic clustering at two different spatial levels, namely urban districts and urban neighborhoods.

The second research aim of the present paper centers around the explanation of the educational attainment of ethnic children with the help of this clustering index in combination with parental attributes and social network characteristics. The results obtained indicate that educational attainment may depend on geographical clustering, but that the geographical scale of analysis is highly influential on the findings.

**JEL Classification:** I2, R2

**Key words:** Ethnic minorities, geographical concentration, educational attainment

# 1 Introduction

The recognition that space matters in socio-economic processes is already half a century old (see Isard, 1956). It is noteworthy that the past decade has witnessed a revival of interest in spatial economic analysis, beginning with the seminal article of Lucas (1988), where economic growth is explained from different geographical perspectives (both opportunities and barriers). Adoption of innovation and access to knowledge are increasingly recognized as drivers of economic progress. In the 90s several authors (e.g. Bénabou, 1996; Durlauf, 1994) have focussed their attention on the transmission of knowledge between economic agents at the level of community neighborhoods. Knowledge distribution acts in this case as the primary force for spatial discrepancies in economic growth and thus welfare.

From this perspective, the influx of immigrants and their tendency to cluster (see e.g. LaLonde and Topel, 1991; Bartel, 1988; and Carrington et al., 1996) may create deviant economic processes in the form of human capital spillovers, and thus different outcomes in welfare in the long run. If immigrants are prone to cluster geographically, then it is likely that along with the formation of ethnic minorities also various network externalities arise. Although Carrington et al. have shown that network externalities cause a decline in the migration costs for new migrants, it may be less beneficial for the transmission of human capital.<sup>1</sup> The slow catching up of second generation immigrants (see for evidence for the Netherlands Tesser et al., 1999, and for the United States Borjas, 1993) compared to the indigenous population may thus be explained from their propensity to cluster.

Borjas (1992; 1995) even separates individual human capital spillovers into a neighborhood effect and an ethnic capital effect, where the last effect represents an ethnic externality generated by human capital spillovers from the ethnic group to their offspring. However, as also mentioned by Borjas, these ethnic capital externalities are created by different average levels of human capital among ethnic groups and not by different structural parameters in the intergenerational human capital transmission. If we assume that ethnic groups have different starting values of human capital and that network effects are present, then the concept of ethnic capital may be regarded as a phenomenon created by the presence of network externalities, where the transmission of human capital operates within the network, formed by the influx of immigrants.

Rather than trying to determine whether structural differences between ethnic groups or (ethnic) networks are responsible for lagging ethnic socio-economic behavior, we aim to investigate in this paper the intergenerational human capital transmission from a spatial perspective. In particular, we will examine whether the ethnic population composition within Dutch neighborhoods may have an impact on the forming of human capital. In addition to this geographical component, we will also incorporate social network effects (with no spatial boundary) and characteristics of the household head, in order to determine the causes for differences in the transmission of human capital.

To detect the impact of population composition on the accumulation and transfer of human capital,

---

<sup>1</sup>These are mainly economic-oriented arguments, while in sociology and geography other examples of network externalities are found, like the devaluation of social norms and values (Wilson, 1987) and the lagging educational performance of schools in segregated areas (Massey and Denton, 1993). On the other hand, ethnic entrepreneurship could be seen as a positive ethnic externality (see e.g. Kloosterman et al., 1999, and Van Delft et al., 2000).

a proper tool is needed for measuring socio-cultural and spatial ethnic clustering. In our research, we will deploy the gamma index of Ellison and Glaeser (1997) to measure whether neighborhoods in Dutch cities have a deviant population composition. The gamma index has two advantages, namely that it is both scale-independent and a result of utility maximization theory.

The remainder of the paper is organized in the following way. The next Section will introduce the ethnic clustering index. The subsequent Section will start with a concise presentation of the data, and with the identification of possible methodological ‘pitfalls’. Thereafter, a model for educational attainment is constructed, which assumes that the transmission of human capital can be studied by looking at the determinants of the schooling level of ethnic children compared to that of the parents. The final part of the Section presents the results, while the last Section will interpret these results.

## 2 An Ethnic Clustering Index

In this Section we will develop an index of ethnic clustering based on the geographic concentration index developed by Ellison and Glaeser (1997). This index has two major advantages compared to other concentration indices. The first one is that the index is the result of profit maximizing behavior by each individual immigrant (or household), which implies that the index originates directly from micro-economic theory (in contrast to e.g. the segregation index<sup>2</sup>). Secondly, Ellison and Glaeser (1997) have proven that the index is scale-independent, which makes comparisons on different geographical scales possible. This in contrast with e.g. Gini’s coefficient, which is sensitive to differences in geographical scale (Bartels, 1977). The next subsection will present the underlying economic model. The subsequent subsection will offer a presentation of the geographical index based on the preceding model.

### 2.1 An Assignment Model of Migrants

Consider the ex post realization of migrants  $k$  ( $k = 1, \dots, N$ ) who belong to a particular ethnic minority  $v_k$ . Assume now that each migrant  $k$  will receive a profit  $\pi_{km}$ <sup>3</sup> from locating in a particular area  $i$  and belonging to ethnic minority  $m$  ( $m = 1, \dots, M$ ). Furthermore, assume also that the migrant’s profit function is given by:

$$\log \pi_{km}^i = \log \xi_m^i + g_m^i(v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_N) + \varepsilon_{km}, \quad (1)$$

where  $\xi_m^i$  is a random variable reflecting the characteristics of ethnic group  $m$  in area  $i$ . Typically, these variables represent characteristics from the source country, like average level of education, religious groups, unemployment levels and the like, or ties between the source country and the destination country;  $g_m^i(\cdot)$  captures benefits established by the network effect by the whole community of immigrants with

---

<sup>2</sup>The segregation index is often used in geography and is defined as the percentage of a population which has to move in order to obtain a homogeneous distribution of that population. See for an application of this index to the Dutch situation e.g. Van Kempen and Van Weesep, 1997.

<sup>3</sup>Profits can be regarded here as the expected net present value of living in area  $i$ .

nationality  $m$ . The  $\{\varepsilon_{km}\}$  are idiosyncratic characteristics of immigrant  $k$ , like stamina, motivation, or risk aversion.<sup>4</sup>

This model is able to account for the fact that the probability of belonging to a certain ethnic minority may also depend on historical factors, like colonial ties and guest worker agreements. This would imply that, without the presence of network effects, the expectation of  $\xi_m^i$  acts as the average probability for a potential migrant to be of ethnicity  $m$ . We can extend this by assuming that the idiosyncratic factors  $\varepsilon_{ki}$  have an extreme value distribution. Then, with  $g_m^i(\cdot) = 0$  and conditional on the realization of  $\xi_1^i, \dots, \xi_{k-1}^i, \xi_{k+1}^i, \dots, \xi_N^i$ , our model can be regarded as a conditional logit model (see McFadden, 1973), where the immigrants' group 'assignment' is modelled by:

$$\Pr(v_k = m | \xi_1^i, \dots, \xi_{k-1}^i, \xi_{k+1}^i, \dots, \xi_N^i) = \frac{\xi_m^i}{\sum_n \xi_n^i}. \quad (2)$$

In order to analyze the ethnic concentration we are interested in defining the moments of the distribution of  $\{\xi_m^i\}$ . Because we want (2) to represent our actual data, we impose the expectation:

$$E_{\xi_1^i, \dots, \xi_{k-1}^i, \xi_{k+1}^i, \dots, \xi_N^i} \left[ \frac{\xi_m^i}{\sum_n \xi_n^i} \right] = x_m^i, \quad (3)$$

where  $x_m^i$  is the share of ethnic minority  $m$  living in a given area  $i$ . In this framework an ethnic group may have a high share and thus a high probability to attract immigrants, because the conditions for the specific ethnic group in that area are favorable. One may think of e.g., a good labor complementarity with workers from country  $m$ , comparable climate conditions, a common language, etcetera.

Next, we will define the variance of (2) for a particular ethnic group  $m$  as follows:

$$\text{Var}_{\xi_1^i, \dots, \xi_{k-1}^i, \xi_{k+1}^i, \dots, \xi_N^i} \left[ \frac{\xi_m^i}{\sum_n \xi_n^i} \right] = \gamma_i^r x_m^i (1 - x_m^i), \quad (4)$$

where  $\gamma_i^r \in [0, 1]$ , with the superscript  $r$  denoting the amount of variation due to the characteristics in area  $i$ . The parameter  $\gamma_i^r$  captures the importance of belonging to an ethnic group  $m$  in a certain area  $i$ . If  $\gamma_i^r = 0$ , then unobserved ethnic characteristics have no effect on attracting immigrants, so that each immigrant has ethnicity  $m$  with probability  $x_m$ . In other words, the probability that in area  $i$  an immigrant is of ethnicity  $m$  is totally random with a probability equal to the nation-wide average. If  $\gamma_i^r = 1$ , then ethnic characteristics are so important that every immigrant in area  $i$  will be of the same ethnicity  $m$ , because the economic, social and environmental conditions for that ethnicity are by far the most favorable in area  $i$ . In this case we witness complete clustering of ethnic groups, so that  $\gamma_i^r$  is a

---

<sup>4</sup>This model can also be seen as an assignment model of immigrants to ethnic minorities, where each individual  $k$  chooses that ethnic group to maximize his expected profits.

measure of ethnic clustering due to ethnic group characteristics. Note that  $x_m^i (1 - x_m^i)$  is the maximum variance of (2) with mean  $x_m^i$ .

The second cause of ethnic clustering in (2) is the presence of networks (see for an insightful economic explanation of the pull force of an ethnic network Carrington et al., 1996). By imposing that networks are beneficial for an immigrant *ceterus paribus*, one assumes the presence of network externalities. Although difficult to express in pecuniary terms, these network externalities consist of the provision of a living place, a provisional job, or even a lowering of the adaptation costs in general. The importance of network externalities can be captured by assuming the following specification for  $g_m^i(\cdot)$ :

$$\log \pi_{km}^i = \log \xi_m^i + \sum_{j \neq k} e_{kj}^i (1 - u_{ji}) (-\infty) + \varepsilon_{km}, \quad (5)$$

where  $e_{ki}^i$  are Bernouilli random variables (0 or 1) equal to 1 with a probability of  $\gamma^{ne}$ , if migrant  $j$  is potentially able to help migrant  $k$ . If migrant  $j$  lives in area  $i$ , then  $u_{ji} = 1$ , otherwise  $u_{ji} = 0$ . The main assumption in (5) is that migrants must live in the same area to support each other. Furthermore, by using the Bernouilli variables, each pair of migrants has a positive probability of forming a network (with a penalty of infinitive negative profits if they do not), indexed by  $\gamma_i^{ne}$ .

Define  $s_{mi}$  now as the actual share of ethnic group  $m$  in area  $i$ ; then a candidate measure of ethnic clustering is given by:

$$G = \sum_m (s_{mi} - x_m)^2. \quad (6)$$

$G$  is quadratic, because we want to measure excess clustering of ethnic groups, regardless whether an ethnic minority inhabits an area in more or less numbers than would be justified by (3).

We can now state that with the use of (3),(4), and (5) the expectation of (6) equals (the proof is given by Ellison and Glaeser, 1997):

$$E(G) = \left(1 - \sum_m x_m^2\right) \left(\gamma_i + (1 - \gamma_i) \sum_m \left(\frac{1}{N_m}\right)^2\right), \quad (7)$$

where  $N_m$  is the total population of the ethnic minority  $m$ , and  $\gamma_i = \gamma_i^r + \gamma_i^{ne} - \gamma_i^r \gamma_i^{ne}$ . This is a remarkable result, because it states that without knowing which clustering forces – ‘natural’, or due to network externalities – are present, it is possible to measure excess clustering. Of course, with raw geographical data the underlying socio-economic processes are impossible to be found, but this proof provides us with a theory-based model in order to construct an index for clustering. The next subsection will develop such a clustering index.

## 2.2 An index of ethnic clustering

First of all we note that even for relatively small ethnic populations the following holds:  $\lim_{N_m \rightarrow \infty} \left(\frac{1}{N_m}\right)^2 \downarrow 0$ . Therefore, by using (7) we may state that

$$\hat{\gamma}_i = \frac{\sum_m (s_{mi} - x_m)^2}{\left(1 - \sum_m x_m^2\right)} \quad (8)$$

acts as an unbiased estimator of  $\gamma_i$ , where  $\hat{\gamma}_i$  is by definition constrained on  $[0, 1]$ .  $\hat{\gamma}_i$  or the gamma coefficient can now be seen as a clustering index of ethnic groups within a spatial area. If ethnic minorities are perfectly randomly dispersed according to their nation-wide average and no network externalities are at work, then  $\hat{\gamma}_i$  is zero. If  $\hat{\gamma}_i$  is equal to one, then our data show that all immigrants in area  $i$  are only member of one and the same ethnic group, so that clustering is at its maximum. Whether this is because ethnic characteristics and area characteristics have a good match or whether social networks are highly influential in the socio-economic performance of the immigrant is impossible to discern at this stage. Note that the denominator in most cases also tends to become very small, so that (8) is mostly determined by (6).

Besides this easy interpretation,  $\hat{\gamma}_i$  has two useful properties. Firstly, the index is straightforwardly developed from (2), which implies that the index can be directly linked to structural economic theory. Secondly, it can be derived that  $\hat{\gamma}_i$  is scale-independent (see Ellison and Glaeser, 1997), which implies that the gamma coefficients can be compared at different spatial scales and also for different definitions of ethnic groups. The next Section will make use of the scale independency of  $\hat{\gamma}_i$ , where the influence of different spatial scales is investigated, in order to investigate whether physical proximity has an effect on socio-economic processes.

### 3 Educational Attainment

For several decades already there has been a long debate whether and how the socio-economic status (SES) of individuals has an effect on their labor market position and educational performance. One of the explanations of the importance of the socio-economic status is that it acts as a transmission channel for human capital. Parental capital (see e.g. Becker, 1974) and the transfer of knowledge between friends and more distant relatives can all be examined in this light and it is generally accepted that these processes have indeed quite some influence on the accumulation of human capital of individuals.

Whether spatial characteristics affect the accumulation of human capital (as advocated by Bénabou, 1996; and Durlauf, 1994) is more subject to debate. If individuals learn from their neighborhood, then human capital accumulation is strongly affected by characteristics of the neighborhood. On the other hand, with the rise of better communication and transportation possibilities one may question the importance of physical proximity. Furthermore, empirical validation of social interaction effects is severely limited because of identification problems (see for a good exposition of this problem Manski, 1993). In this case, it is almost impossible – without some further restricting assumptions – to distinguish between endogenous effects (e.g. neighborhoods effects), correlation effects (e.g. historical sample selection) or contextual effects (e.g. individuals in the neighborhood face the same external constraints). We will return to this issue, and especially to the issue of individual selection effects when we discuss the results.



In the present Section we will concentrate on the effects of spatial distribution of ethnic minorities on educational attainment. For this purpose we will use the gamma coefficient described in the Section above and apply it at two spatial scales, namely neighborhoods and districts of selected cities in the Netherlands. GIS results and estimations will both show that the choice of scale is crucial here.

### 3.1 Data

We will use here the so-called SPVA survey, which was conducted in 1994. The SPVA surveys were conducted among the four largest ethnic groups in the Netherlands, namely the Moroccans, the Surinamese, the Antilleans and the Turks. As a reference group also the Dutch were included. The SPVA was essentially constructed to obtain a good view of the socio-economic position of ethnic minorities in the Netherlands, and is focussed especially on the schooling attainment, the housing situation and the position on the labor market. Because both first and second generation<sup>5</sup> immigrants are included in the survey, a good overview may be provided of the transformation of immigrants to ethnic minorities.

Unfortunately, the survey is not a perfect random subsample of the ethnic communities in the Netherlands. Because the survey is conducted in 13 cities<sup>6</sup> in the Netherlands, more or less an urban view is given. However, with the use of secondary data sets provided by the Dutch Central Bureau of Statistics, the representativeness of the SPVA was checked and – besides of course of urban characteristics – no important differences between the two populations can be found (see Tesser et al., 1995). In other words, with the use of the SPVA around 50 percent of all persons in the Netherlands belonging to the ethnic minorities considered had a chance of being surveyed. Within the cities, ethnic respondents were aselectly sampled with the relative probabilities compared to the size of the respective ethnic community. Indigenous respondents were also aselectly sampled, but have to be regarded as a reference group and can in no way be seen as an aselect sample of the Dutch population.

If we now assume that sampling was totally random, then we can define  $N_m$  as the total population of ethnic minority  $m$  over all cities. If all individuals are now completely randomly distributed over all neighborhoods in the cities under consideration, then the share  $x$  of ethnic minority  $m$  in all neighborhoods is:

$$x_m = \frac{N_m}{\sum_{j=1}^M N_j}, \quad (9)$$

with  $M$  the total number of ethnic groups under consideration (which is in our case five). We are now interested in how we can measure deviations from the ‘perfect’ distribution as defined in (9).

---

<sup>5</sup>Second generation immigrants are here defined as those people who are born in the Netherlands and from whom at least one parent is born in a foreign country. If both parents are born in a foreign, but different country, then the child is assigned to the country of its mother. Surinam and the Dutch Antilles are considered here as foreign countries, although they are (former) colonies.

<sup>6</sup>The following cities were used in the survey: Amsterdam, Rotterdam, The Hague, Utrecht, Eindhoven, Enschede, Almere, Alphen aan de Rijn, Bergen op Zoom, Hoogezand-Sappemeer, Delft, Dordrecht and Tiel. This list comprises also the five largest cities in the Netherlands.

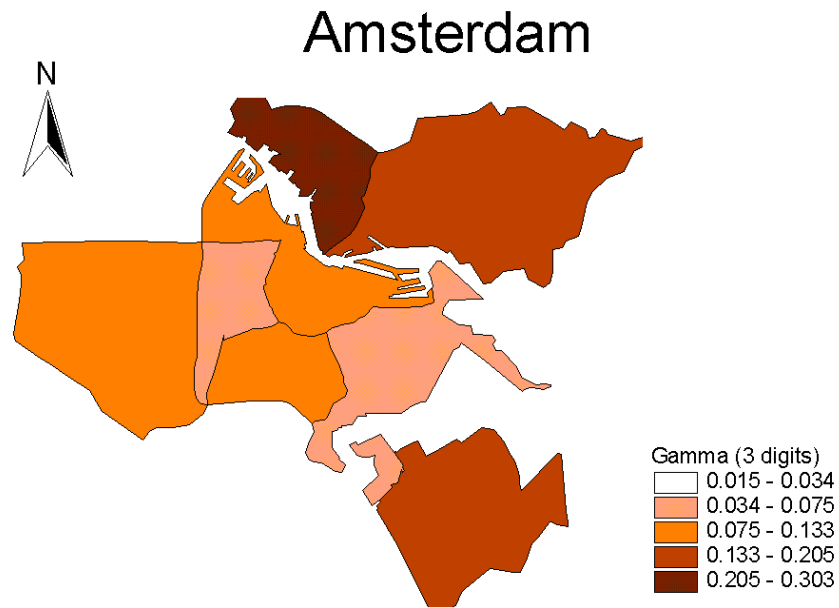


Figure 1a: Distribution of gamma coefficients over the city of Amsterdam on a 3-digit postal code level (districts) (Source SPVA-94).

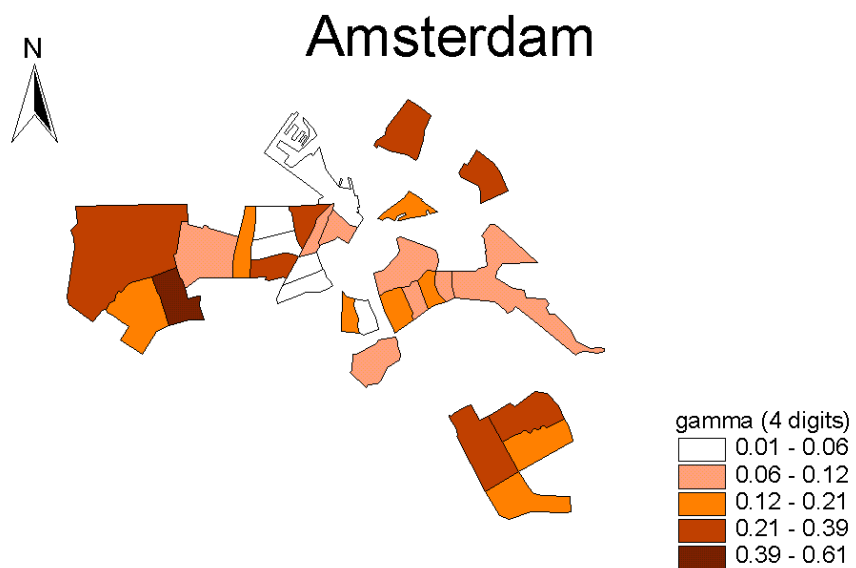


Figure 1b: Distribution of gamma coefficients over the city of Amsterdam on a 4-digit postal code level (neighborhoods) (Source SPVA-94).

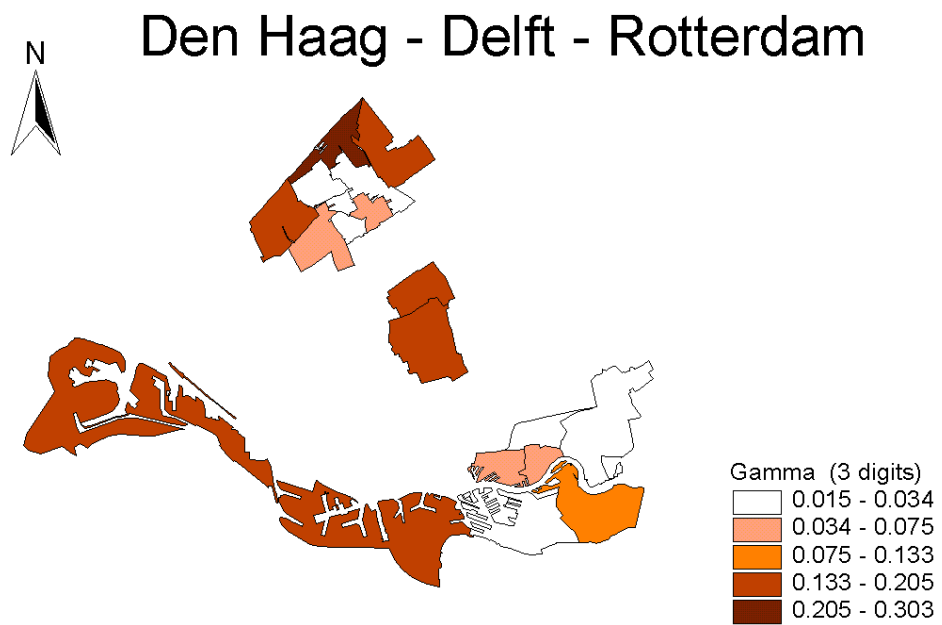


Figure 2a: Distribution of gamma coefficients over the cities of Den Haag, Delft, and Rotterdam on a 3-digit postal code level (districts) (Source SPVA-94).

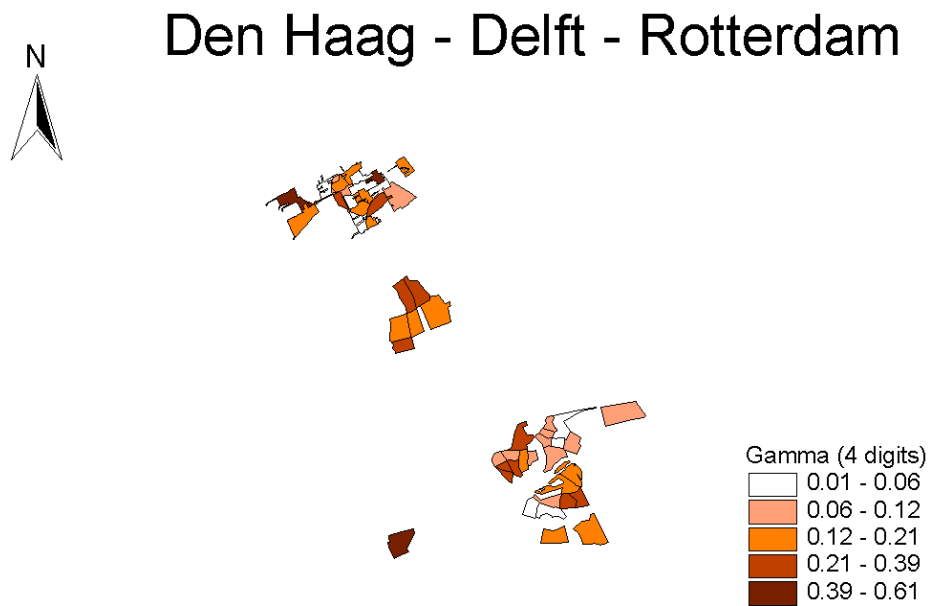


Figure 2a: Distribution of gamma coefficients over the cities of Den Haag, Delft, and Rotterdam on a 3-digit postal code level (districts) (Source SPVA-94).

Note that this ‘perfect’ distribution is only perfect in the framework of the survey. Furthermore, the percentage of the Dutch in the survey is much lower than the percentage of the Dutch in the real population. This may give biased results in the calculation of the gamma coefficient, compared to the real underlying coefficient. Rich urban districts e.g. will have a large share of Dutch inhabitants, which may give an overestimation of their gamma coefficients. However, because we are only interested in relative differences between spatial units, this does not pose a large problem. Because the Dutch population in the survey has more or less comparable socio-economic characteristics, we treat their clustering in the same way as that of the ethnic minorities.

Figure 1 and figure 2 display the distribution of  $\hat{\gamma}_i$  over the districts (these coincide with 3 digit postal codes) and neighborhoods (these coincide with 4 digit postal codes) of Amsterdam and over Den Haag, Delft and Rotterdam, respectively. There are less data on the neighborhood level, because observations in neighborhoods and districts with less than ten inhabitants are removed from the subsample. The distributions show clearly a totally different pattern for each geographical scale. The gamma coefficient calculated on district level in Amsterdam shows excess clustering mainly in the north and the southeast. The southeast (also called the ‘Bijlmer’) is characterized by a high share of Surinamese and Antillean inhabitants (see for even a more micro-area ethnic cluster analysis in Amsterdam, Deurloo and Musterd, 1998).

If clustering however is calculated on a neighborhood level, then another pattern emerges. Here we see again that the southeastern part of Amsterdam shows extreme clustering, just as some northern neighborhoods. However, in the western part of Amsterdam one can see that also here hotspots of ethnic clustering emerge. Here, especially the Moroccan community of Amsterdam is known to be located.

The figures for Den Haag, Delft and Rotterdam offers even a more ambiguous picture. At the district level we observe clustering in the western part of Rotterdam (the southern agglomeration), Delft (the middle agglomeration) and in the rims of Den Haag (the northern agglomeration). However, if we look at the neighborhood level, a different picture emerges. Now, the central and north-western part of Rotterdam, which are early 20th century or early post World War II houses, are clustered. Delft remains largely clustered, but in Den Haag we can now observe that also the central part is clustered. These neighborhoods date all back to the late 19th or early 20th century.

It is not very surprising that different levels of geographical scale can turn into a totally different picture. The spatial scale has to be determined a priori and has to be justified for the purpose of the study. Figure 1 and figure 2 reflect essentially different heterogeneity levels and it seems that even within a city these levels can differ widely, depending on the spatial scale.

For our empirical part of our explanatory analysis we aim to investigate the effect of clustering on schooling attainment. In addition to the location of the household, we have also information on the composition of the household, the characteristics of the household members including schooling attainment and several social interaction variables, which we consider as proxies for ethnic networks. We are interested in the educational attainment of the children in the households above the age of 12.

Figure 3 displays the distribution of schooling attainment of children over the various ethnic groups considered here. Clearly, there are large differences in the level of schooling among the five ethnic groups

considered. Where especially the Turks and the Moroccans have low levels of education, the Dutch control group is significantly overrepresented in the two highest education levels. The Antilleans and especially the Surinamese have schooling attainment levels that are comparable with the Dutch group, except for the highest level of education (higher vocational and academic). The question now arises what are the causes of these differences in educational attainment between the ethnic groups, and whether these differences can be explained by (a selected set of) characteristics of the individuals within the group. Because there is a priori no reason to believe that ethnicity by itself causes deviant schooling performance, the explanatory power of the model may be tested by the significance levels of the ethnic minority dummies.

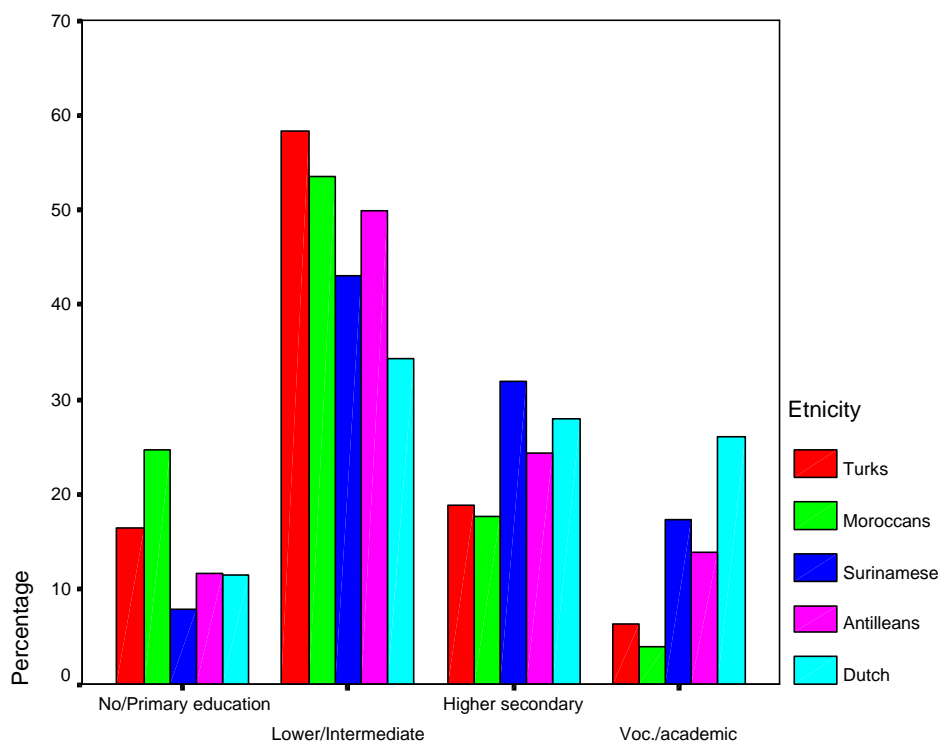


Figure 3: The distribution of schooling attainment (highest level of education attended) by children in the age category 12-25 over the considered ethnic groups in the Netherlands (source: SPVA-94).

### 3.2 A Model of Educational Attainment

Due to data limitations we have to pool the observations of those children who still receive education (around 80%) and those children who have finished their schooling career (around 20%). Therefore, we define here schooling attainment as the highest education level attended by the children above the age of 12 present in the households under consideration. To correct for the large group that is still at school, we have chosen to include a dummy for schooling participation. An alternative could be the use of censoring techniques, which has however the disadvantage of losing much information. Because of the composition

of our data set and thus the chance of losing important information, we have chosen for the dummy approach.

Our schooling attainment,  $y_i$ , is a discrete variable defined as:

$$y_i = \begin{cases} 0 & \text{no or primary education} \\ 1 & \text{lower and intermediate secondary education} \\ 2 & \text{higher secondary education} \\ 3 & \text{higher vocational and academic education} \end{cases} \quad \text{if}$$

We believe that schooling attainment of children is highly dependent on the schooling attainment of the parents ( $y^p$ ). In our research,  $y^p$  will measure the direct intergenerational human capital spillover between parents and children (Becker, 1974). Because schooling attainment is not a perfect measure for human capital, some control variables for the parents are desirable. Therefore, we like to include characteristics ( $Z_h$ ) of the household (measured by characteristics of the household heads), like age and gender of the household head. If both parents are still present, the household head is mostly male (especially in the case of the ethnic minorities). In addition, we include a variable indicating whether the household head speaks Dutch with her child, where the language proficiency of a child is regarded as part of her human capital.

The network variables, here measured by whether someone is member of a society and whether someone has frequent contacts with Dutch individuals, are denoted by ( $Z_n$ ). Therefore, we hypothesize that the variable  $Z_n$  captures human capital spillovers as not taking place within the household. So  $Z_n$  should be affected by relatives outside the household, friends, societies and also the environment in general of the child. Thus,  $Z_n$  also includes the gamma coefficient  $\hat{\gamma}$ , where we will perform a sensitivity analysis by letting  $\hat{\gamma}$  vary between the district level and the neighborhood level. If human capital accumulation is influenced by the geographical environment of the individuals, then a significant impact of  $\hat{\gamma}$  on schooling attainment may be expected.

Let  $X$  be the matrix of characteristics of the child (like gender, age, and being a second generation immigrant), then the following model can be constructed:

$$y^* = X\beta + Z_h\varphi_1 + Z_n\varphi_2 + \varphi_3y^p + \varepsilon, \quad (10)$$

where  $(\beta, \varphi_1, \varphi_2, \varphi_3)$  is the parameter vector to be investigated and where  $\varepsilon$  is a normal variable.  $y^*$  is now a latent variable with the following definition:

$$y_i = \begin{cases} 0 & y_i^* \leq 0 \\ 1 & \text{if } 0 < y_i^* \leq \mu_1 \\ 2 & \mu_1 < y_i^* \leq \mu_2 \\ 3 & \mu_2 < y_i^* \end{cases} \quad (11)$$

This leaves us with the well familiar ordered probit model (see for an exposition of these types of models Maddala, 1983). As a last remark, we mention that the variable ‘second generation immigrant’ in this model is more like a general variable. If human capital accumulation of an individual, conditional on family, direct environment and networks, is still growing, then this is most probably due to more general or national characteristics, like a better schooling system. Therefore, the impact of being a second generation immigrant most likely captures the difference in (pre-)schooling in the Netherlands and (pre-)schooling in the foreign countries under consideration, and their importance for present schooling (results) in the Netherlands. This variable can be expected to be positive, because it will always incorporate country-specific skills acquired in the Netherlands. The next subsection presents the results of the estimation of (11).

### 3.3 Results

The ordered probit model applied to (2) leads to the results presented in Table 1. Because some areas at the neighborhood level had less than 10 inhabitants we have chosen to omit those observations from the subsample, resulting in less observations for the estimation on a neighborhood level. Table 1 shows that the impact and significance of the gamma coefficient clearly changes when working on an different geographical scale. Even the sign reverses (though insignificantly). Most probably this is due to the fact that the information content of data on different spatial aggregation levels differs, a phenomenon closely related to the usual problems with aggregation in economics (see for an early presentation Theil, 1954). In this case we regressed the dummies for the ethnic minorities on the gamma coefficient at the two spatial levels. Only the Dutch regression coefficient changed significantly, from significantly positive for the 3-digit postal codes to insignificantly negative for the 4-digit postal codes. Because we already observed from Figure 3 that the Dutch are the highest educated, the previous finding indicates that on a district level mainly the significant clustering of the Dutch causes the positive impact of  $\hat{\gamma}_i$  on educational attainment. On the other hand, because the Dutch dummy does not attribute significantly to the gamma coefficient at the neighborhood level, the impact of  $\hat{\gamma}_i$  becomes also insignificant. This is highly remarkable, because it states that geographical clustering may cause positive externalities for the Dutch, but if ethnic minorities are clustered no effects can be found. We refer to the Discussion for possible explanations of this result.

To discuss the results of Table 1, we have to calculate the marginal effects of the coefficients for each of the four education levels (*cf.* Greene, 1992). This is because the effects and even the signs of the coefficient are not immediately clear (especially not for the subgroups  $y_i = 1, 2$ ). Instead, we calculated the effects of changes in the covariates on the cell probabilities. Table 2 and 3 present the marginal effects for all cells, where the marginal effects of the dummy variables are considered as the differences between the impacts of the extremes of the variables (0 or 1). All marginal effects are evaluated at the means.

Table 1: Ordered probit results for a schooling attainment model with  $y_i^*$  level of education.

	Gamma coefficient on 3-digit postal codes		Gamma coefficient on 4-digit postal codes	
	<i>Coeff.</i>	<i>St. error</i>	<i>Coeff.</i>	<i>St. error</i>
Constant	-1.55	0.31	-1.30	0.37
<i>Individual characteristics</i>				
Gender (Female = 1)	0.10	0.07	0.11	0.08
Age	0.14	0.01	0.12	0.01
Generation ( $2^{nd}$ = 1)	0.19	0.10	0.17	0.11
Schooling participation (yes = 1)	1.06	0.10	0.95	0.11
Turkish-dummy	-0.19	0.17	-0.02	0.19
Moroccan-dummy	-0.37	0.16	-0.30	0.18
Surinam-dummy	-0.04	0.15	0.04	0.18
Antillean-dummy	-0.16	0.17	-0.05	0.20
<i>Household head characteristics</i>				
Gender household head (female =1)	-0.19	0.10	-0.14	0.11
Age household head	0.00	0.01	0.00	0.01
Schooling attainment household head	0.16	0.05	0.18	0.06
Not speaking Dutch with their children	-0.16	0.07	-0.15	0.07
<i>Network variables</i>				
No member of a society	-0.10	0.04	-0.11	0.05
No contact with the Dutch	-0.19	0.06	-0.17	0.07
Gamma coefficient ( $\hat{\gamma}_i$ )	1.17	0.60	-0.44	0.43
$\mu_1$	1.63	0.06	1.63	0.07
$\mu_2$	2.59	0.07	2.62	0.09
Log likelihood		-1144		-882
$N$		1067		830

Tables 2 and 3 clearly show more or less the same results for all variables, except the gamma coefficient, where the sign is clearly reversed, which has already been explained above. Gender does not seem to have a significant impact on educational attainment. Higher age leads to a higher educational attainment, while the dummy for schooling participation suggests that individuals who have finished their schooling career have less chance to obtain a higher education level. These two results are not surprising, because they point to the fact that there are more possibilities for education when still at school and being at young age.



Table 2: Marginal effects for the coefficients of the schooling attainment estimation with the gamma coefficient calculated at a 3-digit postal code level (variables significant at 5% level in bold and dummy variables evaluated at their extreme values).

	$\partial P_0/\partial \mathbf{x}_i$	$\partial P_1/\partial \mathbf{x}_i$	$\partial P_2/\partial \mathbf{x}_i$	$\partial P_3/\partial \mathbf{x}_i$
<i>Individual characteristics</i>				
Gender (Female = 1)	0.00	-0.02	-0.02	0.04
Age	<b>-0.03</b>	<b>-0.02</b>	<b>0.03</b>	<b>0.02</b>
Generation (2 <sup>nd</sup> = 1)	<b>0.00</b>	<b>-0.04</b>	<b>-0.03</b>	<b>0.08</b>
Schooling participation (yes = 1)	<b>-0.03</b>	<b>-0.29</b>	<b>-0.07</b>	<b>0.39</b>
Turkish-dummy	0.00	0.04	0.03	-0.08
Moroccan-dummy	<b>0.01</b>	<b>0.09</b>	<b>0.06</b>	<b>-0.15</b>
Surinam-dummy	0.00	0.01	0.01	-0.02
Antillean-dummy	0.00	0.04	0.02	-0.06
<i>Household head characteristics</i>				
Gender household head (female =1)	<b>0.00</b>	<b>0.05</b>	<b>0.03</b>	<b>-0.08</b>
Age household head	0.00	0.00	0.00	0.00
Schooling attainment household head	<b>-0.03</b>	<b>-0.02</b>	<b>0.03</b>	<b>0.02</b>
Not speaking Dutch with their children	<b>0.03</b>	<b>0.02</b>	<b>-0.03</b>	<b>-0.02</b>
<i>Network variables</i>				
No member of a society	<b>0.02</b>	<b>0.01</b>	<b>-0.02</b>	<b>-0.01</b>
No contact with the Dutch	<b>0.04</b>	<b>0.03</b>	<b>-0.04</b>	<b>-0.03</b>
Gamma coefficient ( $\hat{\gamma}_i$ )	<b>-0.25</b>	<b>-0.16</b>	<b>0.25</b>	<b>0.16</b>

The dummies for the ethnic groups show that only the Moroccans children on the district level have significantly lower education levels. For the Turks, Surinamese and Antilleans no significant impact can be found. This is in line with theory, which states that schooling attainment is dependent on background characteristics, parental capital and the direct environment as well as on the schooling system in general and not with the ethnicity of a child. In addition, the insignificance levels of the variable show that the variation between the ethnic groups is adequately explained.

Being a second generation immigrant significantly improves the probability of a higher education level, which is in line with most other research. In this case our findings suggest that schooling or pre-schooling in the Netherlands is more efficient than schooling abroad. This is not surprising, because the country-specific human capital, which is needed for further Dutch education, is not taught in foreign countries.

Table 3: Marginal effects for the coefficients of the schooling attainment estimation with the gamma coefficient calculated at a 4-digit postal code level (variables significant at 5% level in bold and dummy variables evaluated at their extreme values).

	$\partial P_0/\partial \mathbf{x}_i$	$\partial P_1/\partial \mathbf{x}_i$	$\partial P_2/\partial \mathbf{x}_i$	$\partial P_3/\partial \mathbf{x}_i$
<i>Individual characteristics</i>				
Gender (Female = 1)	-0.00	-0.03	0.01	0.04
Age	<b>-0.03</b>	<b>-0.01</b>	<b>0.03</b>	<b>0.01</b>
Generation (2 <sup>nd</sup> = 1)	-0.01	-0.05	-0.01	0.06
Schooling participation (yes = 1)	<b>-0.06</b>	<b>-0.29</b>	<b>0.03</b>	<b>0.31</b>
Turkish-dummy	0.00	0.01	0.00	-0.01
Moroccan-dummy	0.01	0.09	0.02	-0.11
Surinam-dummy	0.00	-0.01	0.00	0.02
Antillean-dummy	0.00	0.02	0.00	-0.02
<i>Household head characteristics</i>				
Gender household head (female =1)	0.00	0.04	0.01	-0.05
Age household head	0.00	0.00	0.00	0.00
Schooling attainment household head	<b>-0.04</b>	<b>-0.02</b>	<b>0.04</b>	<b>0.02</b>
Not speaking Dutch with their children	<b>0.04</b>	<b>0.01</b>	<b>-0.03</b>	<b>-0.02</b>
<i>Network variables</i>				
No member of a society	<b>0.03</b>	<b>0.01</b>	<b>-0.02</b>	<b>-0.01</b>
No contact with the Dutch	<b>0.04</b>	<b>0.02</b>	<b>-0.04</b>	<b>-0.02</b>
Gamma coefficient ( $\hat{\gamma}_i$ )	0.11	0.04	-0.10	-0.05

The characteristics of the household heads also provide intuitive appealing results. Age of the household heads is not important for the schooling attainment of the children, while being female has on a district level a negative impact on the schooling results of the children. A possible explanation may be that these household heads are single mothers, who do not have enough time to devote to their children for additional education.<sup>7</sup> The level of schooling attainment of the household head or the parental human capital is positively affecting schooling attainment of the children, with a magnitude which is in line with other research (Borjas, 1992).

Not being a member of a society (which could range from a sports club to a political society) and not having many Dutch friends has a negative influence on the level of schooling. This may be due to the influence of friends and the environment on human capital transfers. In the sports club e.g., an ethnic child could learn more elements of Dutch language than at home (or even at school). Being more intertwined with the Dutch community could also indicate a higher commitment to e.g. the Dutch schooling system.

<sup>7</sup>Around 85% of the female household heads in our data set are single.

Because we had ample information about the parents, we were able to construct language and network variables for the parents in order to analyze their effect on the children's schooling attainment. Especially the general proficiency of the parents in the Dutch language would seem highly influential in the child's educational attainment. On the other hand, it seems plausible that not the parental knowledge but the practice of the Dutch language enables their children to learn the Dutch language more quickly. Because they had no significant impact, these variables were omitted from the final estimations.

Returning to the identification problem, we have to admit that it is not possible to explain the impact of clustering on a structural basis without further assumptions. In this stage we are only interested in the (direction of) impact of clustering. However, a possible explanation of self selection is not as obvious as it seems. Because of a rather stringently controlled social housing market (the focus of the survey used), self selection at lower geographical scales is hardly an issue in the large cities in the Netherlands.<sup>8</sup> Only 9 percent of the houses in our data set were privately owned and 5 percent were owned by a private foundation. The prevailing policy within the social housing market until the mid 90s was that if somebody was eligible he was offered a house; one could decline this offer up to three times, after which one was placed at the end of the list again. Candidates could crudely indicate in which part of the city they wanted to live, but these housing offers did not necessarily have to be in these districts. In this framework, possible sources of selection behavior are more directed to choices whether to buy or rent a house, or living in a city or not, and less in which urban neighborhood one wants to live.

## 4 Discussion

This paper aimed at explaining differences in educational attainment between ethnic groups from a geographical perspective. Therefore, we introduced in Section 2 a scale-independent clustering index, which was developed from economic optimization theory. This index is able to measure excess clustering from population groups within a certain area. The hypothesis stated in the introduction was that excess clustering of ethnic groups could hamper their human capital transmission by means of negative network externalities. However, the results show that only that if the Dutch are clustered a positive impact is transmitted, while excess clustering of the other ethnic minorities do not result in a significant impact. Of course, this will also result in a relatively better schooling position for the Dutch, which will widen eventually the economic performance gap between the indigenous and ethnic minority groups.

This small impact of geographical clustering, especially at the neighborhood level, may be due to the highly controlled urban housing market in the Netherlands. However, as figures 1 and 2 indicate, there is also at the neighborhood level a wide variation in geographical clustering between neighborhoods. Furthermore, the neighborhoods one would expect to be clustered (like in the south-east part of Amsterdam) show indeed a high value of the gamma coefficient. So the conclusion is valid that on a neighborhood level children living in highly clustered neighborhoods are not doing significantly worse on school than children in less clustered areas. Furthermore, the results clearly show that the variation among the ethnic groups

---

<sup>8</sup>For the sake of illustration, more than 80 percent of the Turks and Moroccans in the Netherlands appear to make use of the social housing market (Bolt, 2001).

can be explained by a limited set of moderator variables, which is certainly not in support of Borjas' ethnic capital explained in the Introduction.

Although ethnic capital and geographical clustering does not seem to have a large impact on educational attainment, we see from our results that social interactions in general have a large influence on the educational level. The fact that an ethnic child has contacts with the Dutch and membership of a club or society are highly beneficial for his or her schooling results. In general, there may be the tendency that strictly bordered areas as our urban districts and neighborhoods will exert less influence on their inhabitants, because of increasingly better transportation and communication possibilities. This will severely change the spatial dimension, where transformation channels are less bounded by Euclidean distance.

In addition to network variables, educational attainment of a child is of course also influenced by parental human capital. Parental human capital is in this research proxied by the schooling attainment of the household head and a variable indicating whether the household head speaks Dutch with his or her child. The importance of the former variable is already indicated by Borjas (1992), while the latter variable is particularly examined by Chiswick (see e.g. Chiswick and Miller, 1999).

As already mentioned, due to the strictly controlled urban housing market in the Netherlands, individuals are less able to choose the place where to live, which reduces possible sources of self-selection in the dataset. However, we know that especially the Dutch population is more inclined to buy a house or to move out of the city (suburbanization). It is therefore very likely that there is a great amount of unobserved heterogeneity present in our data set, due to this source of self-selection. There are several extensions of our model possible in order to control for this self-selection. The most frequently used is to gather panel data. However, in this case we need a large time-span to capture housing mobility, in which we have to assume that our structural parameters remain constant. Especially in the case of ethnic minorities, which are in the middle of an integration or even assimilation process, this would seem a too harsh assumption. An alternative approach is to model the unobserved heterogeneity by a joint estimation of two correlated processes (see e.g. Van den Berg et al., 1994). In this case we may link educational attainment with the choice whether to buy or rent a house.

Finally, according to Manski (1993), it is very difficult to identify the underlying structural causes for group effects. If ethnic children have many contacts with Dutch children, it is beneficial for their schooling results, but this does not say anything about the reason. It could well be that these children have unobserved characteristics which make them perform well at school and also encourage them to have frequent contacts with Dutch children. Therefore, the question whether educational attainment is determined by social interactions, whether they are both results of an unknown factor, or even whether social interactions is the endogenous factor and educational attainment the cause, remains still unanswered.

**Acknowledgment:** *The authors would like to thank Raymond Florax, Pieter Gautier and participants of the International Symposium on Regional Development in Port Elizabeth and of the conference "Migration and Development" in Acquafredda di Maratea for their useful comments and remarks. The assistance of Arend Odé and Henk Scholten in providing the data is gratefully acknowledged.*

## References

- [1] Bartel, A.P., Where do the New U.S. Immigrant Live?, *Journal of Labor Economics*, vol. 7, pp. 371-391, 1989.
- [2] Bartels, C.P.A., *Economic Aspects of Regional Welfare, Income, Distribution and Unemployment*, Martinus Nijhoff, Leiden, 1977.
- [3] Becker, G.S., A Theory of Social Interactions, *Journal of Political Economy*, vol. 82, pp. 1063-1093, 1974.
- [4] Bénabou, R., Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance, *American Economic Review*, vol. 86, no. 3, pp. 584-609, 1996.
- [5] Bolt, G., *Wooncarriers van Turken en Marokkanen in Ruimtelijk Perspectief*, Netherlands Geographical Studies 281, Utrecht, 2001.
- [6] Borjas, G.J., Ethnic Capital and Intergenerational Mobility, *Quarterly Journal of Economics*, vol. 107, pp. 123-150, 1992.
- [7] Borjas, G.J., The Intergenerational Mobility of Immigrants, *Journal of Labor Economics*, vol. 11, pp. 113-135, 1993.
- [8] Borjas, G.J., Ethnicity, Neighborhoods, and Human-Capital Externalities, *American Economic Review*, vol. 85, pp. 365-390, 1995.
- [9] Carrington, W.J., E. Detragiache and T. Vishnawath, Migration with Endogenous Moving Costs, *American Economic Review*, vol. 86, pp. 909-930, 1996.
- [10] Chiswick, B.R. and P.W. Miller, Language Skills and Earnings among Legalized Aliens, *Journal of Population Economics*, vol. 12, pp. 63-89, 1999.
- [11] Deurloo, M.C. and S. Musterd, Ethnic Clusters in Amsterdam, 1994-96: A Micro-area Analysis, *Urban Studies*, vol. 35, pp. 385-396, 1998.
- [12] Durlauf, S.N., Spillovers, Stratification, and Inequality, *European Economic Review*, vol. 38, pp. 836-845, 1994.
- [13] Ellison, G. and E.L. Glaeser, Geographical Concentration, *Journal of Political Economy*, vol. 105, pp. 890-927, 1997.
- [14] Greene, W.H. , *Econometric Analysis*, Macmillan Publishing Company, New York, 1992.
- [15] Isard, W., *Location and Space-Economy: A General Theory Relating to Industrial Location, Market Areas, Land Use, Trade, and Urban Structure*, The Technology Press of Massachusetts Institute of Technology, Cambridge, 1956.

- [16] Kloosterman, R.J., J. van der Leun and J. Rath, Mixed Embeddedness: (In)formal Economic Activities and Immigrant Business in the Netherlands, *International Journal of Urban and Regional Research*, vol. 23, pp. 252-266, 1999.
- [17] LaLonde R.J. and R.H. Topel, Immigrant in the American Labor Market: Quality, Assimilation and Distributional Effects, *AEA Papers and Proceedings*, vol. 81, pp. 297-302, 1991.
- [18] Lucas, R.E., On the Mechanics of Economic Development, *Journal of Monetary Economics*, 22, pp. 3-42, 1988.
- [19] Maddala, G.S., *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge, 1983.
- [20] Manski, C.F., Identification of Endogenous Social Effects: The Reflection Problem, *Review of Economic Studies*, vol. 60, pp. 531-542, 1993.
- [21] Massey, D.S. and N.A. Denton, *American Apartheid: Segregation and the Making of the Underclass*, Harvard University Press, Cambridge, 1993.
- [22] McFadden, D., Conditional Logit Analysis of Qualitative Choice Behavior, in P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, 1973.
- [23] Tesser, P.T.M., C.S. van Praag, F.A. van Dugteren, L.J. Herweijer and H.C. van der Wouden, *Rapportage Minderheden: Concentratie en Segregatie*, Sociaal en Cultureel Planbureau, Rijswijk, 1995.
- [24] Tesser, P.T.M., J.G.F. Merens and C.S. van Praag, *Rapportage Minderheden 1999*, Sociaal en Cultureel Planbureau, Rijswijk, 1999.
- [25] Theil, H., *Linear Aggregation of Economic Relations*, North-Holland, Amsterdam, 1954.
- [26] Van Delft, H., C. Gorter and P. Nijkamp, In Search of Ethnic Entrepreneurship Opportunities in the City, *Environment and Planning C: Government and Policy*, vol. 18, pp. 429-451, 2000.
- [27] Van den Berg, G.J., M. Lindeboom and G. Ridder, Attrition in Longitudinal Panel Data and the Empirical Analysis of Dynamic Labour Market Behaviour, *Journal of Applied Econometrics*, vol. 9, pp. 421-435, 1994.
- [28] Van Kempen, R. and J. van Weesep, Segregation, Housing and Ethnicity in Dutch cities, *Tijdschrift voor Economische en Sociale Geografie*, vol. 88, pp. 188-195, 1997.
- [29] Wilson, W.J., *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*, University of Chicago Press, Chicago, 1987.