

Kertesi, Gábor; Kézdi, Gábor

**Working Paper**

## Ethnic segregation between Hungarian schools: Long-run trends and geographic distribution

Budapest Working Papers on the Labour Market, No. BWP - 2012/8

**Provided in Cooperation with:**

Institute of Economics, Centre for Economic and Regional Studies, Hungarian Academy of Sciences

*Suggested Citation:* Kertesi, Gábor; Kézdi, Gábor (2012) : Ethnic segregation between Hungarian schools: Long-run trends and geographic distribution, Budapest Working Papers on the Labour Market, No. BWP - 2012/8, ISBN 978-615-5024-34-9, Hungarian Academy of Sciences, Institute of Economics, Centre for Economic and Regional Studies, Budapest

This Version is available at:

<https://hdl.handle.net/10419/108476>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



---

**BUDAPEST WORKING PAPERS ON THE LABOUR MARKET**  
**BWP – 2012/8**

**Ethnic segregation between Hungarian schools:  
Long-run trends and geographic distribution**

GÁBOR KERTESI – GÁBOR KÉZDI

---

INSTITUTE OF ECONOMICS, RESEARCH CENTRE FOR ECONOMIC AND REGIONAL STUDIES  
HUNGARIAN ACADEMY OF SCIENCES  
DEPARTMENT OF HUMAN RESOURCES, CORVINUS UNIVERSITY OF BUDAPEST  
BUDAPEST, 2012

Budapest Working Papers On The Labour Market  
BWP – 2012/8

Institute of Economics, Research Centre for Economic and Regional Studies,  
Hungarian Academy of Sciences  
Department of Human Resources, Corvinus University of Budapest

Ethnic segregation between Hungarian schools:  
Long-run trends and geographic distribution

Authors:

Gábor Kertesi  
senior research fellow  
Institute of Economics,  
Hungarian Academy of Sciences RCERS  
email: kertesi@econ.core.hu

Gábor Kézdi  
associate professor  
Central European University  
senior research fellow  
Institute of Economics,  
Hungarian Academy of Sciences RCERS  
email: kezdig@ceu.hu

November 2012

ISBN 978 615 5024 34 9  
ISSN 1785 3788

# **Ethnic segregation between Hungarian schools: Long-run trends and geographic distribution**

Gábor Kertesi – Gábor Kézdi

## **Abstract**

Using all of the available data on the ethnic composition of Hungarian primary schools, this paper documents the degree of between-school segregation of Roma versus non-Roma students between 1980 and 2011. We calculate the measures of segregation within school catchment areas as well as within micro-regions and within the larger municipalities (towns and cities). Catchment areas are clusters of villages, towns and cities that are closed in terms of student commuting, and they are defined by us using the observed commuting patterns. Our results show that ethnic segregation between Hungarian schools increased substantially between 1980 and 2011. Segregation appears to have decreased between 2006 and 2008 and increased again afterwards, but the noise in the data prevents us from drawing firm conclusions. In the cross section, school segregation is positively associated with the size of the educational market and the share of Roma students, similar to the results from U.S. metropolitan areas. These relationships strengthened over time in Hungary, and the change in segregation is associated with changes in the number of schools and the share of Roma students.

JEL classification: H75, I21, J15

**Keywords:** School segregation, Roma minority, disadvantaged students

**Acknowledgement:** We thank Melinda Tir for her assistance with data management, Tímea Molnár, Péter Dívós and Agnes Szabo-Morvai for their earlier work on programming and László Göndör for his help with the maps. We thank Gábor Bernáth, János Zolnay, the editor and the referee for their thoughtful comments. All the remaining errors are ours. Individual research grants from the Institute of Economics of the Hungarian Academy of Sciences, RCERS are gratefully acknowledged.

# **Iskolák közötti etnikai szegregáció a magyar iskolarendszerben: földrajzi eloszlás és hosszú távú trendek**

Kertesi Gábor – Kézdi Gábor

## **Összefoglaló**

A tanulmány a magyar általános iskolákról intézményszinten rendelkezésre álló összes etnikai adat felhasználásával iskolák közötti etnikai (roma / nem roma) szegregációs indexeket számol az ország valamennyi városára, kistérségére és iskolai ingázási vonzáskörzetére az 1980 és 2011 közötti időszak több évére: 1980-ra, 1989-re, 1992-re, illetve a 2006 és 2011 közötti periódus minden egyes évére. Iskolai ingázási körzeteknek neveztük el azon települések együttesét, amelyek több százezernyi általános iskolába járó tanuló megfigyelt iskolai ingázási kapcsolatai alapján többé-kevésbé zártnak tekinthetők. Az adatok tanúsága szerint a roma tanulók iskolák közötti elkülönülése igen jelentős mértékben nőtt 1980 óta, és különösképpen a rendszerváltást követő két évtizedben. Ami a legfrissebb adatokat illeti: noha a 2006-2008 közötti időszakban a szegregáció jól érzékelhető csökkenését, a 2008 után években pedig a szegregáció ismételt növekedését tapasztalhattuk, adathiányok miatt robusztus megállapításokat az utolsó öt év fejleményeire nem tehetünk. Keresztmetszeti összehasonlításban az iskolák közötti szegregáció mértéke leginkább azokon a helyeken magas, ahol az intézményszámban mért oktatási piac mérete nagy és a roma tanulók aránya magas. Ez a megfigyelés összhangban van az Egyesült Államokbeli tapasztalatokkal, ahol fekete bőrű tanulók nagyvárosi övezetekben mért iskolai szegregációja szintén ott magas, ahol bőséges iskolakínálat áll rendelkezésre és a fekete bőrű tanulók számaránya is magas. Ez az összefüggés nemcsak keresztmetszetben, hanem a változásokat tekintve is érvényes: a szegregáció ott nőtt meg leginkább, ahol az iskolák száma és a roma tanulók aránya növekedett.

**Tárgyszavak:** iskolai szegregáció, roma kisebbség, hátrányos helyzetű tanulók

**JEL kódok:** H75, I21, J15

## 1. INTRODUCTION

Over ten percent of the Hungarian students in primary schools are Roma. The typical Roma students come from substantially poorer families and have lower achievement than the typical non-Roma students (Kertesi and Kézdi, 2011). The extent to which Roma and non-Roma students study in the same schools can have serious consequences for ethnic differences in achievement and other outcomes as well as for the integrity of Hungarian society.

Using all of the available comprehensive data on the ethnic composition of Hungarian primary schools, this paper documents the degree of between-school segregation of Roma versus non-Roma students between 1980 and 2011. We show the long-run trends and the geographic distribution, and we estimate regressions to uncover the associations between segregation and other characteristics of the areas, which are identified from the cross-section and from the long-differenced panel of the areas for which school segregation is defined.

It is necessary to have some institutional knowledge of the Hungarian school system to understand school segregation. We are interested in the primary schools that cover grades 1 through 8 (these include some secondary schools that cover grades 5 through 8). Importantly, and similar to other countries in the region, Hungary is characterized by the dominance of state-owned primary schools, and parents are free to choose schools for their children. On top of the enrollment from within their own district, which is defined by the municipality, schools can admit children living outside of the district. The total enrollment in schools is determined by their capacity, the level of demand from within and from outside of their district and the allocation decision by the municipality.

We estimate the degree of segregation within three types of geographic area: the 174 micro-regions, the larger school catchment areas (clusters of villages, towns and cities that are closed in terms of student commuting in the 2000s and have two schools or more) and the larger municipalities (towns and cities with two or more schools). Our preferred unit of measurement is the catchment area because it represents the area that is the most relevant for school choice. In a sense, micro-regions are too large: school segregation within micro-regions is likely to be heavily influenced by the residential patterns across towns and villages. The towns and cities are too small: measuring segregation within their administrative boundaries misses potentially important commuting from and to villages in their agglomeration. The school catchment areas are not administratively registered units; they are defined by commuting possibilities. A contribution of our paper is to define the boundaries of those areas using the actual commuting patterns of all 6<sup>th</sup> graders observed in three different years.

Our preferred measure of segregation is the index of segregation (also known as the isolation index, see Clotfelter, 2004), but we also show results using the more traditional index of dissimilarity. There is no data from between 1992 and 2006, and the missing data decreases the reliability of the post-2006 figures. Aside from our best estimates, we also show conservative lower and upper bounds. We show time series of the average level of segregation and maps for its geographic distribution. Finally, we show cross-sectional and long-differenced regressions for partial correlations of the between-school segregation with the size of the educational market, the average school size and the fraction of Roma students.

Our results indicate that school segregation, on average, is moderate in Hungary. The mean of the index of segregation is approximately 0.2 in the geographic areas covered by our analysis and is approximately 0.3 in the areas around the three largest cities. Note that Hungarian schools are characterized by fixed assignment to groups within schools (“classes”). Within-school between-class segregation may therefore be as important for inter-ethnic contact as between-school segregation. Unfortunately, our data does not allow for calculating indices within-school ethnic segregation. But it allows for looking at the segregation of students whose mother has 8 grades of education or less, both between schools and within schools. On average, the level of their within-school segregation is about 40 per cent on top of the level of their between-school segregation (details of the calculations are available from the authors upon request). This suggests that the level of ethnic segregation, if measured across classes instead of schools, is likely to be about 40 per cent higher than the level of ethnic segregation across schools (0.28 instead of 0.20 on average, and 0.4 instead of 0.3 in the areas around the largest cities).

The data also show that, on average, the level of school segregation within Hungarian towns increased substantially between 1980 and 2011. According to our benchmark estimates, between-school segregation appears to have decreased between 2006 and 2008 and increased again afterwards. However, the trends after 2006 cannot be robustly identified due to severe data limitations. In the cross-sectional regressions, school segregation is positively associated with the size of the educational market and the share of the ethnic minority, similar to results from U.S. metropolitan areas, and these relationships strengthened over time. In the regressions estimated in long differences, the change in segregation is also associated with these factors, but the associations are weaker except for the change in the size of the Roma minority

The rest of the paper is organized as follows. Section 2 introduces the data, section 3 defines the effective catchment areas of schools, and section 4 introduces the measures of segregation. Section 5 shows the average levels of segregation and its times series, and section 6 shows its

geographic distribution. Section 7 shows the regression results, and section 8 concludes the paper.

## **2. DATA AND METHODS**

The level of school segregation for a particular area is measured using the total number of students and the fraction of Roma students in each school within the area. We use two sources that cover the population of Hungarian primary schools. Before 1992, all schools filled out a compulsory questionnaire that contained, among other things, the total number of students and the number of Roma students in the school. The number of Roma students was based on counts by classes carried out by teachers. We have data from the years 1980, 1989 and 1992. The reporting on Roma students was discontinued after 1992.

The data on the fraction of Roma students are available from 2006 in the Hungarian National Assessment of Basic Competences (NABC; Országos Kompetenciamérés in Hungarian). The NABC is a standards-based assessment, with tests on reading and mathematical literacy in grades 6 and 8 in primary schools (grades 4 and 8 in 2006 and 2007). The NABC became standardized in 2006, and we use data from 2006 through 2011 for our analysis. Aside from testing the students, the NABC collects additional data on students and schools. School-level data are provided by the school principals. The testing takes place in May of each year, and school-level data are collected at the same time. Among other things, the school-level data contains information on the number of students and the school principal's estimate of the fraction of Roma students in the school. These estimates are likely to contain significantly more noise than the figures from 1992 and before, but we have no reason to believe that they are biased (the estimates figures were not used for targeting any policy measure and they were not published).

The information is collected from each school site, i.e., each unit of the school with a separate address. This level of data collection is important because in some towns, the schools as administrative units comprise units at multiple locations, sometimes far from each other. Throughout the entire study, we use the word “school” to denote the school site and use the word “institution” for the level of administrative organization that can contain more than one school site.<sup>1</sup>

---

<sup>1</sup> With very few exceptions, institutions were single-address schools before the early 1990s, so the data from between 1980 and 1992 are at the school and the institutional level at the same time.



Our analysis contains data on the population of Hungarian schools that teach primary school students, i.e., students in grades 1 through 8.<sup>2</sup> Of these schools, the NABC covers all that had students in grade 4 or grade 8 in 2006 and 2007, and all schools that had students in grades 6 and 8 from 2008 onwards. Coverage by the NABC is limited because it misses the institutions that teach students with special educational needs (S.E.N. students) except in 2006. Another source of bias is that the information on the fraction of Roma students is missing in some schools that do participate in the NABC. In addition to the problem of S.E.N. students, therefore, nonresponse is another cause of missing data.

Missing data can bias the segregation indices. Suppose, for example, that the schools in which the principal fails to provide information have no Roma students at all. In that case, our measures overestimate exposure and therefore underestimate segregation because the missing schools have exposure levels below the average. In theory, it is also possible that the schools with missing data have an ethnic composition that is very close to the town-level average. In that case, our measure of segregation would be biased upwards. Similarly, missing data can bias the estimates of the size of the Roma student population. If the schools with no information all have zero Roma students, the true share of Roma students among all students is lower than the estimate. If, instead, all of the schools with missing information are all-Roma schools, the true fraction of Roma students is higher than the estimates. Note that the bias is different for the segregation measures (a measure of dispersion) and the overall share of Roma students (a mean).

*Table 1* shows the prevalence of missing data. The table shows the number of institutions from the administrative files (KIR-STAT), the number of institutions in the NABC data, the number of schools in the NABC data (recall that we define a school as a facility with a separate mailing address; some institutions have more than one school), and the number of schools with valid data. Administrative sources (KIR-STAT) have information on the number of students at the institution level but not at the school level as we define it. KIR-STAT has no information on the ethnic composition of schools.

---

<sup>2</sup> Traditionally, secondary schools would start with grade 9. In the early 1990s, some secondary schools began to recruit students in the lower grades and have incoming classes in grade 7 or as early as grade 5. These secondary schools are concentrated in the largest cities, most of them in Budapest. See Horn (2012) for a more detailed discussion. Our data cover all students in grades 1 through 8 including those enrolled in secondary schools. For simplicity, we call these institutions primary schools as well.

Table 1.

**The number of institutions and schools in Hungary in the administrative data and the NABC data, 2006-2011.**

	Number of institutions		Number of school sites	
	All ( from KIR-Stat)	In the NABC data	In the NABC data	In the NABC data with non-missing fraction of Roma students
2006	3334	3267	3966	3444
2007	3247	3048	3420	2883
2008	2693	2465	3130	2885
2009	2541	2371	3097	2858
2010	2481	2307	3060	2792
2011	2454	2278	2925	2763

Note: “Schools” are defined by their physical location (address); “institutions” can contain more than one school. We consider primary schools (and their institutions) to be the schools that teach students from grade 1 through grade 8. KIR-STAT is the administrative register for all educational institutions in Hungary. NABC (Országos Kompetenciamérés; the National Assessment of Basic Competences) is the national standard based assessment, with tests on reading and mathematics for grades 6 and 8 (grades 4 and 8 in 2006 and 2007). Students with special educational needs do not participate in the assessment, except in 2006. The school-level data in NABC cover all schools with at least one student who participated in the assessment.

Table 1 shows that both of the missing schools in the NABC data (and thus the missing information on all students) and the missing information on the Roma students in the NABC data are potentially important. We address the first problem by linking the schools through time and imputing student numbers from KIR-STAT. We address the problem of the missing Roma data in three alternative ways. The benchmark imputation is our best estimate. We complement the benchmark with an imputation that leads to the lowest possible value for the segregation index and one that leads to the highest possible value. Similarly, we compute the lower and upper bound estimates for the fraction of Roma students.<sup>3</sup> In most of the analysis, we focus on the results using the benchmark imputation, but we show the results with the alternative missing data treatments as well when they are important.

<sup>3</sup> The benchmark procedure uses the data from previous and subsequent years for the schools that do not experience large changes in total student numbers. Approximately 30 schools are still missing data in each year after this procedure. The imputation that leads to the lowest possible value of the segregation index uses the area-level average fraction of Roma students for the missing data (all initially missing data, including those that were filled in with our best estimate in the benchmark procedure). The imputation that leads to the highest value of the index of segregation imputes zero or one for the missing fraction of Roma students in a way that leaves the overall fraction of Roma students unchanged, up to indivisibility issues (it assigns the value of one to the smaller schools and zero to the larger schools following the observed relationship in the non-missing data). The imputation that leads to the lowest (highest) fraction of Roma students is simply zero (100 percent).

### 3. DEFINING CATCHMENT AREAS

School choice results in the extensive commuting of students between their residence and school. In this setting, the natural geographic unit for studying school segregation is the smallest area that covers all of the schools that are available to the students living in the area. In other words, it is the smallest area that is closed in terms of potential commuting. School segregation measured within larger units is influenced by residential patterns that commuting cannot overcome; school segregation measured within smaller units misses schools that should be considered.

In this section, we define the effective catchment areas of primary schools. Our smallest geographic units of observation are the municipalities (villages, towns and cities; there are over 3000 municipalities in Hungary). A catchment area can consist of a single municipality and a single school, more than one municipality and a single school, or multiple municipalities and/or multiple schools. Ideally, all students who live in a catchment area go to a school within the area, and nobody from outside the area goes to the schools within the area. The goal is to partition Hungary into a complete collection of disjoint areas. Ideally, these areas should not be too large. Areas that are too large would not only work against the purpose of the exercise (by making area-level analysis difficult) but would go against spirit of the definition (very few schools would be available for any particular student within the area).

We used individual data collected from the NABC data for the students' residence and the location of their schools for three years. We created a directed and weighted graph using the individual data on commuting connections. Municipalities (villages, towns and cities) are the nodes (vertices), and the numbers of students commuting between the nodes are the links (edges). The direction of the link is from the node of residence to the node of the school, and the weights are the number of commuters. The largest weights in this graph are on the links that connect the nodes to themselves (loops): these are the students whose school and residence is within the same municipality.

Catchment areas are a partition of the set of all municipalities (villages, towns and cities): every municipality belongs to one and only one catchment area. In the language of graph theory, catchment areas are the connected components in the entire graph. Connected components are defined for undirected (symmetric) and unweighted graphs: graphs that indicate whether two nodes are connected or not without any further information. For this problem, the original graph can be transformed into an undirected and unweighted graph with the help of a threshold value: two nodes are connected if and only if the number of students commuting between them

exceeds a threshold level in any direction. Given the undirected and unweighted graph, the breadth-first-search algorithm finds all of the connected components in the graph and thus creates a partition of the set of all municipalities.<sup>4</sup>

The data on students' residence come from administrative records of all sixth-graders from three years, 2008, 2009 and 2010. The overall number of observations is 304,125. Simple coding errors or administrative mistakes could create apparent links between two municipalities with no links. The probability of such events is never zero, but the same event is unlikely to happen twice. For this reason, we have chosen two for the threshold value used to transform the weighted into the unweighted graph: nodes are considered to be connected if the data imply that more than one student is commuting in any direction between them.<sup>5</sup>

It turns out, however, that this benchmark graph has one giant component and many tiny ones. The graph contains 99 components; out of these, 96 have 13 or fewer nodes (the distribution is, of course, very skewed). Of the remaining three components, one has 44 nodes, one has 229 nodes, and the largest has 2669 nodes.<sup>6</sup> The giant component contains Budapest and most cities from all regions of Hungary. This partition is clearly useless for any practical analysis. Therefore, we created an alternative partition: we broke the largest three components into smaller clusters by increasing the threshold value for links to 5 students per year on average (a total of 15 students for the three years) or at least 20 per cent of the originating node (the municipality of residence).<sup>7</sup> The resulting partition contains 1055 catchment areas. The largest area contains 71 municipalities, and it covers the Budapest agglomeration. The other large areas contain large cities and their agglomerations.<sup>8</sup>

---

<sup>4</sup>See, for example, [http://en.wikipedia.org/wiki/Breadth-first\\_search](http://en.wikipedia.org/wiki/Breadth-first_search) for a detailed description of the algorithm.

<sup>5</sup> If a municipality has no school and it sends one student only in these three years to any other municipality, that link is preserved. Similarly, the links that were below the threshold value of one student were preserved if they represented over 20 percent of all students from the sending municipality. Municipalities without schools that are not connected to any other municipality in the data were linked to the nearest neighboring municipality that has a school (using geographic coordinates).

<sup>6</sup> The emergence of a giant component is a classic result in graph theory: if links are created randomly, almost all nodes are connected with a high probability if the number of links exceeds a threshold value.

<sup>7</sup> Similarly to the previous step, links were preserved even if they were below the threshold if a municipality has no school and it sends its students to one and only one other municipality. Municipalities without schools that are not connected to any other municipality in the data were linked to the nearest neighboring municipality that has a school (using geographic coordinates).

<sup>8</sup> The threshold values used in the new partition are obviously ad-hoc, but the results represent an intuitively compelling partition and any attempt to break the giant component would require assumptions of this kind.

Table 2 shows the most important summary statistics on the catchment areas.<sup>9</sup> Not surprisingly, the size distribution is skewed, and the areas with the largest number of municipalities are even larger in terms of student population because they contain the largest cities.

Table 2.

**Summary statistics on the catchment areas. Number of municipalities and number of primary schools and students (information from schools is averaged over 2006 through 2011)**

Size of catchment area (number of municipalities)	Number of catchment areas	Average number of municipalities	Average number of primary schools	Average number of primary school students
1	624	1.0	1.2	232
2 to 4	297	2.7	2.1	408
5 to 9	74	6.3	7.1	1,885
10 to 19	37	12.9	9.4	2,192
20 to 49	20	30.5	25.4	6,102
50 to 71	3	60.0	224.9	65,045
Total	1055	3.0	3.3	782

#### 4. MEASURING SCHOOL SEGREGATION

Following the literature (e.g., Clotfelter, 2004), we measure segregation with the help of the following three indices: exposure of non-Roma students to Roma students (ENR), exposure of Roma students to non-Roma students (ERN), and the standardized version of these indices, referred to here as the segregation index (S). For completeness, we also look at the more traditional but theoretically less attractive index of dissimilarity (D). When we calculate the extent of exposure or segregation, we look at schools within a catchment area (or, alternatively, a micro-region or a town or city). To define and interpret these indices, we work with the following notation. Index  $i$  denotes the schools, and index  $j$  denotes the areas (these are the areas that contain the schools; students may reside outside the areas, see our discussion later).

$I_j$  is the number of schools in area  $j$ ,

$N_{ij}$  is the number of students in school  $i$  in area  $j$ ,

<sup>9</sup> Additional data on the catchment areas, including the set of municipalities in them and further data on students, is available from the authors upon request.

$N_j$  is the number of students in area  $j$ ,

$R_{ij}$  is the number of Roma students in school  $i$  in area  $j$ ,

$R_j$  is the number of Roma students in area  $j$ ,

$r_{ij}$  is the fraction of the Roma students among all students in school  $i$  in area  $j$ ,

$r_j$  is the fraction of the Roma students among all students in area  $j$ ,

$(1 - r_{ij})$  is the fraction of the non-Roma students among all students in school  $i$  in area  $j$ ,

$(1 - r_j)$  is the fraction of the non-Roma students among all students in area  $j$ ,

Index  $ENR_j$  measures the exposure of an average (a randomly chosen) non-Roma student in area  $j$  to the possibility of meeting Roma students.  $ENR_j$  is equal to the fraction of Roma students in each school averaged over schools, where the average is taken with weights that are equal to the share of non-Roma students in the school in all non-Roma students in the area. Formally,

$$ENR_j = \sum_{i=1}^{I_j} r_{ij} \frac{N_{ij} - R_{ij}}{N_j - R_j}, \quad \text{so that} \quad 0 \leq ENR_j \leq r_j$$

The minimum value of the exposure index is zero: in this case, no contact is possible between Roma and non-Roma students within the schools because the schools are either all-non-Roma (when  $r_{ij}=0$ ) or all-Roma (when  $N_{ij}-R_{ij}=0$ ). The maximum value of exposure is when the fraction of minority students in each school is equal to the fraction in the area:  $r_{ij}=r_j$  for all  $i$  in  $j$ . For  $ENR_j$  to make sense, we need  $0 < r_j < 1$ , i.e., there must be both Roma and non-Roma students in area  $j$ . This condition is satisfied in all of the areas that we consider.

The exposure of Roma students to non-Roma students ( $ERN_j$ ) is analogous: it measures the exposure of an average (a randomly chosen) Roma student in area  $j$  to the possibility of meeting non-Roma students.  $ERN_j$  is equal to the fraction of non-Roma students in each school averaged over schools, where the average is taken with weights that are equal to the share of the school in the Roma student population of the area. Formally,

$$ERN_j = \sum_{i=1}^{I_j} (1 - r_{ij}) \frac{R_{ij}}{R_j}, \quad \text{so that} \quad 0 \leq ERN_j \leq 1 - r_j$$

The minimum value of this exposure index is zero, also, and  $ERN_j = 0$  exactly when  $ENR_j = 0$ . This value indicates that no contact is possible between Roma and non-Roma students within the schools because the schools are either all-Roma ( $1-r_{ij}=0$ ) or all-non-Roma ( $r_{ij}=0$ ). The maximum value of Roma exposure occurs when the fraction of non-Roma students in each

school is equal to the fraction in the area:  $1-r_{ij}=1-r_j$  for all  $i$  in  $j$ . The two indices are intimately related:

$$ERN_j = \frac{1-r_j}{r_j} ENR_j$$

Despite their intuitive content, the exposure indices are rarely used. Their values depend on the overall fraction of minority students in the area, which poses a severe constraint on their use in comparing segregation across time or areas. The segregation index is intended to solve this problem. The index of segregation is a normalized version of the exposure indices, and thus it retains their information content, albeit in a less intuitive way. The normalization amounts to comparing exposure to its attainable maximum; there is also a reversal of sign so that the higher levels of the index indicate higher levels of segregation (less exposure). Intuitively, the segregation index shows the fraction of contact possibilities that are made impossible by segregation. Formally,

$$S_j = \frac{r_j - ENR_j}{r_j} = \frac{(1-r_j) - ERN_j}{1-r_j}, \quad \text{so that} \quad 0 \leq S_j \leq 1$$

The maximum value of the index is one: segregation is at its maximum when the exposure is zero. The minimum value is zero: it is attained at maximum exposure, which is when the fraction of Roma students is the same in every school.

An alternative measure of segregation is the index of dissimilarity. Defined from the viewpoint of Roma students, and with many schools in mind, this index can be interpreted as the percentage of non-Roma students that would have to move to different schools to have schools with the same fraction of Roma students within the area. Formally, the index of dissimilarity is defined as

$$D_j = \frac{1}{2} \sum_{i=1}^{I_j} \left| \frac{R_{ij}}{R_j} - \frac{N_{ij} - R_{ij}}{N_j - R_j} \right|, \quad \text{so that} \quad 0 \leq D_j \leq 1$$

Similar to the index of segregation defined above, a value of 1 would denote complete segregation, and a value of 0 would denote equal distribution across schools. In any other case, the index of dissimilarity is, in general, not equal to the index of segregation. The index of dissimilarity is a more traditional measure than the index of segregation, but it lacks the latter's theoretical relationship to exposure. For that reason, the index of segregation is a more useful measure that is used in the new literature on school segregation (Clotfelter, 1999).

## 5. TRENDS IN SCHOOL SEGREGATION IN HUNGARY, 1980-2011

We measure the ethnic composition of primary schools and segregation between schools in years 1980, 1989, 1992 and yearly between 2006 and 2011. Recall that the data in 1980, 1989 and 1992 are high quality, that there are no data from between 1992 and 2006, and that the data starting with 2006 are of lower quality, characterized by many schools without information on the fraction of Roma students. For that reason, from 2006 onwards, we show the conservative lower and upper bound estimates of both the overall share of Roma students and the index of segregation in addition to our best estimates. We define segregation within three geographic areas: catchment areas, micro-regions and municipalities (villages, towns and cities). Naturally, between-school segregation is defined for the areas with two schools or more. We restricted the analysis to areas that had two schools or more in each year of observation. This criterion was fulfilled by 175 out of the 1055 catchment areas and 140 towns or cities of the over 3000 municipalities (towns, cities and villages).

*Table 3* shows the averages of the segregation indices in 1980 and 2011 in the three geographic areas. The averages shown in the table are weighted by the distribution of students.

*Table 3.*

**The ethnic composition and ethnic segregation of primary schools in Hungary, 1989 and 2011, in catchment areas as well as in micro-regions and larger municipalities (towns and cities) and including the number of students, the fraction of Roma students, the indices of exposure, the index of segregation and the index of dissimilarity.**

Average values	Larger catchment areas		Micro-regions		Towns and cities	
	1980	2011	1980	2011	1980	2011
Average number of students	5,153	3,324	6,668	4,235	4,723	3,139
Fraction of Roma students	0.05	0.11	0.06	0.13	0.03	0.08
Exposure of non-Roma students to Roma students	0.04	0.08	0.05	0.10	0.03	0.07
Exposure of Roma students to non-Roma students	0.86	0.69	0.85	0.68	0.90	0.75
Index of segregation	0.09	0.22	0.09	0.22	0.07	0.19
Index of dissimilarity	0.48	0.53	0.47	0.53	0.47	0.51
Number of observations	175	175	174	174	140	140

Note: Average values (using the benchmark imputations from 2006 onwards) weighted by the number of students (except for the average number of students, which is unweighted).



The first row of Table 3 shows the number of students. The most important information here is the uniform decline in the number of students by 35 percent. The second row shows the fraction of Roma students. The figures show a strong increase: the fraction of Roma students in Hungarian primary schools more than doubled between 1980 and 2011. A small part of their increasing share is due to the increased participation of Roma students in primary school education, but a large part is due to demographics.

The catchment areas shown in this table refer to areas that had two or more schools during the 1980 to 2011 period and thus do not cover the smallest catchment areas, which have only one school. In the part of Hungary that is covered by these two-or-more-school catchment areas, the share of Roma students was 5 percent in 1980 and increased to 11 percent by 2011. The micro-regions cover the entire country, and thus the figures in the corresponding columns refer to the overall fraction of Roma students in Hungary. From a 6 percent level in 1980, the share of Roma students in primary schools (grades 1 through 8) increased to 13 percent by 2011. The corresponding figures in the larger municipalities (towns and cities with two or more schools) are 3 percent in 1980 and 8 percent in 2011. The lower levels in the larger catchment areas and the even lower levels in the larger municipalities show that the Roma population is overrepresented in the smaller villages and that the degree of overrepresentation did not decrease over time.

The exposure of non-Roma students to Roma students increased, but at a slower pace than the increase in the share of Roma students, the theoretical maximum of the exposure index. Mirroring this trend, the exposure of Roma students to non-Roma students decreased significantly, more than the decreasing share of non-Roma students would imply. Taken together, the trends in the indices imply an increasing trend in the segregation index.

Between-school segregation increased substantially in Hungary between 1980 and 2011. Taking the average of the catchment areas, the relevant geographic units in the system of free school choice in Hungary, the index of segregation increased from 9 percent to 22 percent. The intuitive content of these figures is that the chance of contact between Roma students with non-Roma schoolmates decreased from 91 percent of its theoretical maximum in 1980 to 78 percent of the maximum level by 2011.

*Table 4* shows the number of students, the share of Roma students and the index of segregation for the catchment areas around the six largest Hungarian cities in 1980 and 2011. Similar to the national trends, these areas experienced a large drop of 30 to 40 percent in the number of students. Again, similar to the national trends, the share of Roma students increased substantially in each area. The levels differ considerably, but the trends are rather similar except

for the catchment area of Budapest where the increase was more than three-fold, from 2 percent to 7 percent in. The highest share, both in 1980 and in 2011, was in the catchment area of the northern city Miskolc, while the lowest share was in the catchment area of the southern city Szeged.

Ethnic segregation increased considerably in most but not all of the catchment areas. The index of segregation increased almost threefold in the areas of Budapest, Miskolc and Debrecen. Segregation increased by a smaller amount in the Pécs and Győr areas, and it decreased substantially in the Szeged area.

*Table 4.*

**Ethnic composition and ethnic segregation of the primary schools in the catchment areas around the largest Hungarian cities, 1989 and 2011, including the number of students, the fraction of Roma students, and the index of segregation.**

	Budapest		Miskolc		Debrecen	
	1980	2011	1980	2011	1980	2011
Number of students	237,896	165,931	35,255	20,818	28,280	20,361
Fraction of Roma students	0.02	0.07	0.09	0.20	0.02	0.04
Index of segregation	0.06	0.26	0.13	0.36	0.09	0.26
Number of municipalities	71		33		7	

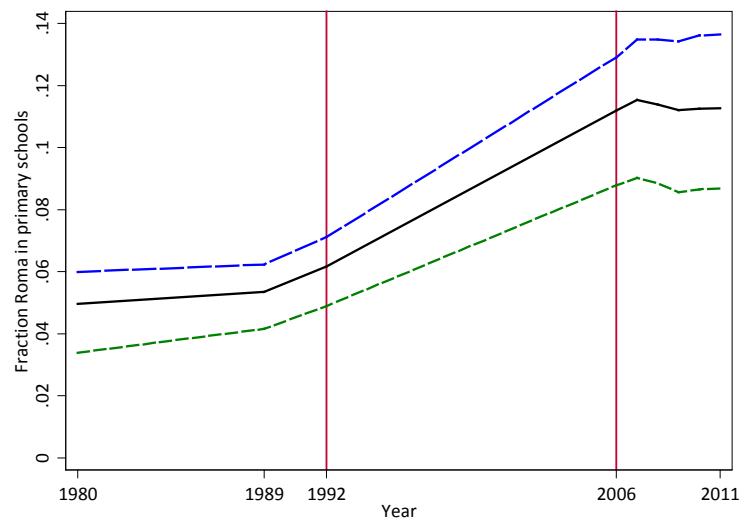
	Pécs		Szeged		Győr	
	1980	2011	1980	2011	1980	2011
Number of students	24,020.0	15,489.0	20,178.0	14,311.0	19,736.0	13,316.0
Fraction of Roma students	0.04	0.08	0.02	0.03	0.02	0.04
Index of segregation	0.13	0.16	0.16	0.05	0.06	0.13
Number of municipalities	55		12		37	

The level of segregation in 1980 could be considered to be low; the level in 2011 is moderate. The U.S. metropolitan areas that are characterized by the school segregation of African Americans and whites similar to the levels documented for large Hungarian areas include San Diego (0.28), Phoenix (0.31) or Los Angeles (0.33). These are not among the most segregated U.S. cities: the segregation index in New York City is 0.45; it is 0.57 in Chicago; and the most segregated metropolitan area is that of Detroit (0.71, see Clotfelter, 1999, p. 494).

Figure 1 shows the times series of the fraction of Roma students as estimated using the benchmark imputation procedure. Figure 5 in the Appendix shows the same time series together with the conservative lower and upper bounds. Recall that the bounds represent the most conservative imputations for the missing data. While we cannot rule out any figure within the bounds, our benchmark estimates use available information in a careful way and are thus likely to be close to the true figures. The post-2006 data are also noisier, although that noise is unlikely to have a significant effect on the aggregate figures. According to the benchmark results, the increase in the fraction of Roma students was concentrated in the 1989 to 2007 period, and the increase stopped afterwards. When one looks at the intervals between the lower and upper bounds in Figure 5, the apparent trend break is lost in the overall degree of uncertainty.

*Figure 1.*

**The time series of the fraction of Roma students in primary schools in larger catchment areas (solid line), micro-regions (long dashed line) and larger municipalities (towns and cities; dashed line). Hungary, 1980 through 2011.**

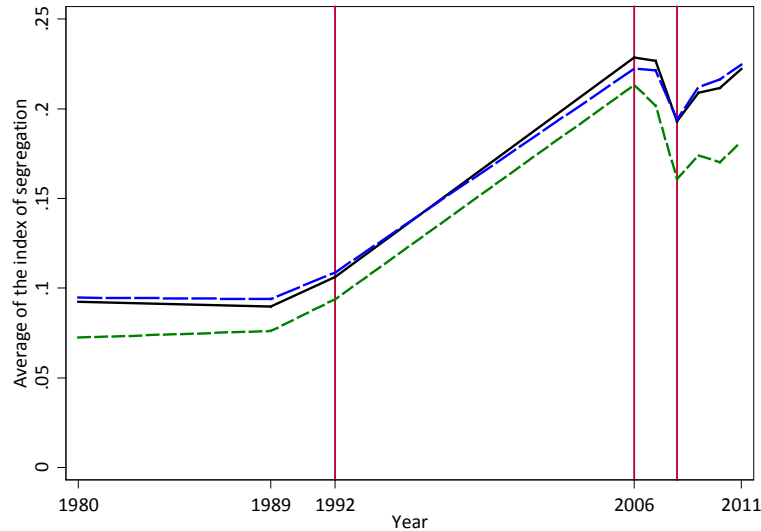


Note: The fraction of Roma students after 2006 is based using our benchmark imputations for missing data.

Figure 2 shows the time series for the index of segregation, and Figure 6 in the Appendix shows the uncertainty interval for our calculations using the lower and upper bound imputations for the missing data in 2006-2011. The figures show the time series of the index of segregation from 1980 through 2011 averaged over the geographic areas (catchment areas, micro-regions and larger municipalities).

Figure 2.

**The time series of the index of ethnic segregation between primary schools, including the average of the index (weighted by number of students) in larger catchment areas (solid line), micro-regions (long dashed line) and larger municipalities (towns and cities; dashed line). Hungary, 1980 through 2011.**



Note: The index after 2006 is based using our benchmark imputations for the missing data.

According to Figure 2, between-school segregation by ethnic lines stayed constant between 1980 and 1989 but began to increase afterwards. By 2006, between-school segregation reached a value that is more than double the 1989 level. This increase is large and is also robust to the imputation method that we chose for the missing data. Our best estimate for the index shows a significant decline in between-school segregation between 2006 and 2008, and the decrease appears to be driven by the larger municipalities. The slope of the decreasing trend is comparable to the slope of the previous increase, resulting in a small drop because of the short time interval.

The trend breaks in the time series coincide with trends in the desegregation initiatives of the government of Hungary. A law introduced in 2004 banned segregation based on race, ethnicity and social background and divided the burden of proof between the plaintiffs and the defendants. In the following years, advocacies and offices of the central government pressured some of the towns and cities to close down segregated schools. By anecdotal evidence, these central government activities came to a halt after 2008. The link between desegregation in larger municipalities and the observed patterns of segregation is further supported by the fact that the trend breaks are largest for the largest municipalities, from 0.21 to 0.16. The drop was

smaller, from 0.23 to 0.19, in the catchment areas that included not only the towns and cities but also some of the surrounding villages. This finding is consistent with the larger municipalities implementing desegregation within their administrative boundaries without the other parts of their catchment area following suit. Furthermore, some of the largest drops between 2006 and 2008 are observed in the cities that implemented changes in the composition of their schools as a result of desegregation plans (including, for example, Szeged, shown in Table 4). This evidence suggests that the observed trend breaks could be real.

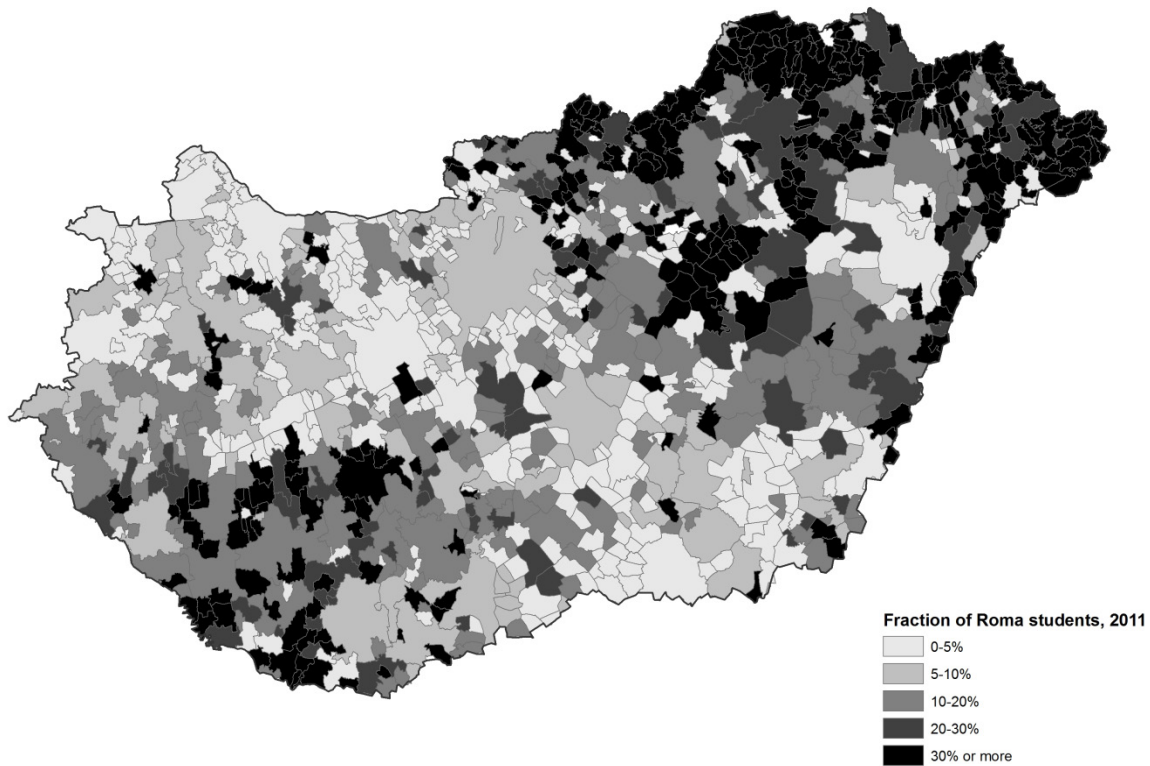
However, the trend breaks also coincide with the apparent breaks in the time series of the share of Roma students, which is harder to understand. This trend implies that the estimated breaks in the segregation indices could be spurious. Indeed, while the large increase between 1992 and 2006 is robust to the imputation method used after 2006, the trend breaks after 2006 are not robust at all. Similar to the Roma share series, the benchmark estimates are surrounded by a very wide interval of possible values between the conservative lower and upper bound, shown in Figure 6. As a result, the coincidence of the trend breaks with the desegregation activities could be completely spurious. Evidently, the missing information in the NABC data simply prevents us from identifying trends after 2006.

## **6. THE GEOGRAPHIC DISTRIBUTION OF SCHOOL SEGREGATION**

The Roma population is distributed unevenly in Hungary. Using all data available up to 1993, Kertesi and Kézdi (1998) presented detailed maps on the geographic distribution of the Roma population in Hungary. Using school-level information in a system characterized by school choice and the widespread commuting of students, we can present analogous maps at the level of the catchment areas for the 2000s. Figure 3 shows a map of Hungary divided into catchment areas (1055 clusters of villages, towns and cities) with the fraction of Roma students for all areas in 2011.

*Figure 3.*

**The share of Roma students in primary schools in all catchment areas of Hungary in 2011.**

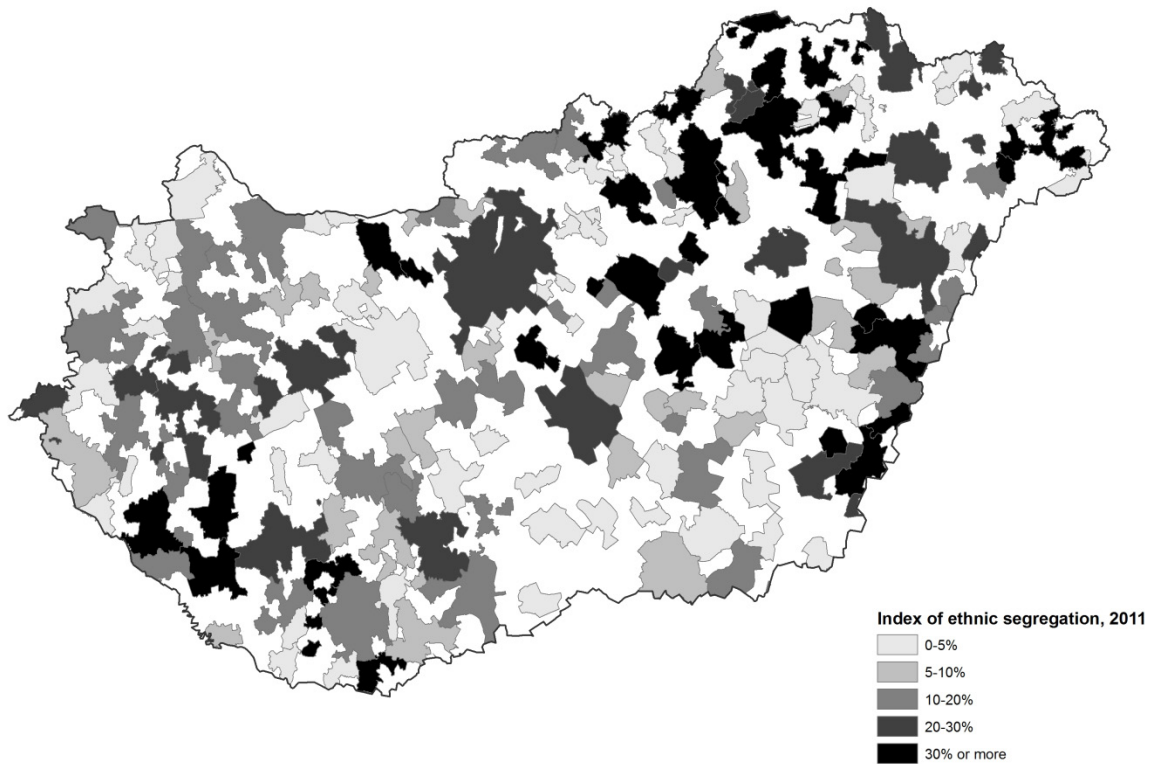


Note: Catchment areas are clusters of villages, towns and cities that are close in terms of student commuting. We defined these areas using the observed commuting patterns. The calculations are based on our benchmark imputations for the missing data.

Between-school segregation is defined for the larger catchment areas that have two or more schools. Throughout this paper, we consider the catchment areas that had two or more schools throughout the entire period. Figure 4 shows the map of the 175 largest catchment areas and presents the index of segregation in these areas.

Figure 4.

**Ethnic segregation between primary schools in the larger catchment areas of Hungary in 2011.**



Note: Catchment areas are clusters of villages, towns and cities that are close in terms of student commuting. We defined these areas using the observed commuting patterns. The calculations are based on our benchmark imputations for the missing data.

Comparing the two maps suggests two patterns. The geographic distribution of school segregation is quite similar to the geographic distribution of the Roma students. This similarity suggests a positive and potentially quite strong relationship between the share of Roma students in the area and the level of ethnic segregation between primary schools. However, this correspondence is far from being perfect. The areas around Budapest, Pécs and Győr, for example, are characterized by relatively strong segregation but a low fraction of Roma students. This finding suggests that other mechanisms can also be important and that the size of the area is likely to be related to these mechanisms. In the following section, we present regression results that show some more systematic evidence for these types of associations.

## **7. SCHOOL SEGREGATION AND THE SIZE OF THE EDUCATIONAL MARKET, SCHOOLS AND THE ROMA POPULATION**

In our final analysis, we show regression results with the index of segregation being the left hand-side variable and the size of the area (number of schools), the average size of the schools and the size of the Roma minority (fraction of Roma students) on the right hand-side. We first present the results from the cross-sectional regressions for 1980 and 2011. These regressions show cross-sectional associations: whether, in a given point in time, the areas that are larger, have larger schools or have a higher fraction of Roma students in the schools are characterized by higher or lower levels of school segregation.<sup>1011</sup>

The results are shown in

The fraction of Roma students in the area is the strongest predictor of school segregation, with increasing magnitude over time and across geographic units (being the strongest predictor within towns and cities). Towns and cities that had a one percentage point higher fraction of Roma students in their schools were characterized by a 0.75 percentage point higher index of segregation. In terms of standardized coefficients, the towns and cities with a fraction of Roma students that is higher by one standard deviation (0.1) were characterized by a half of a standard deviation (0.14) higher index of segregation on average, holding the number of schools and the average school size constant.

Table 5, and the summary statistics are in Table 7 in the Appendix. The number of schools in the area was positively associated with school segregation in 2011, while the association was substantially weaker in 1980. The change is also statistically significant. In 2011, the standard deviation of the log number of schools was between 0.8 and 1.0 depending on the geographic area definition (Table 7); the areas that are larger by one standard deviation were characterized by a one tenth of a standard deviation higher index of segregation on average, holding ethnic composition and average school size constant. The average size of the schools is negatively, albeit weakly, correlated with the segregation between schools, with no clear pattern across years or definitions of the geographic area.

---

<sup>10</sup> Apart from missing information from some schools after 2006, our data represent the population of schools. We show standard errors nevertheless, because we interpret our regressions as models that try to uncover more general tendencies in educational markets that are characterized by the properties of the Hungarian educational markets in the observed years.

<sup>11</sup> Note that the Budapest agglomeration is an outlier in terms of size, and it experienced larger than average increase in both the share of Roma students and the index of segregation. Nevertheless, the estimated coefficients are very similar when we exclude Budapest.



The fraction of Roma students in the area is the strongest predictor of school segregation, with increasing magnitude over time and across geographic units (being the strongest predictor within towns and cities). Towns and cities that had a one percentage point higher fraction of Roma students in their schools were characterized by a 0.75 percentage point higher index of segregation. In terms of standardized coefficients, the towns and cities with a fraction of Roma students that is higher by one standard deviation (0.1) were characterized by a half of a standard deviation (0.14) higher index of segregation on average, holding the number of schools and the average school size constant.<sup>12</sup>

Table 5.

**School segregation and the size of the educational market, schools and the Roma population. Cross-sectional regressions for selected years.**

Dependent variable: S (index of Segregation)	Larger catchment areas		Micro-regions		Larger municipalities	
	1980	2011	1980	2011	1980	2011
Log number of schools	0.022 [2.45]*	0.055 [4.98]**	0.020 [1.81]	0.066 [7.10]**	0.021 [2.06]*	0.062 [8.84]**
Log average school size	-0.024 [1.51]	-0.022 [0.58]	-0.032 [2.19]*	-0.067 [2.17]*	-0.056 [2.09]*	-0.036 [0.84]
Fraction of Roma students	0.439 [4.27]**	0.661 [6.86]**	0.247 [3.00]**	0.563 [8.40]**	0.624 [2.53]*	0.747 [6.06]**
Constant	0.142 [1.41]	0.076 [0.36]	0.200 [2.23]*	0.288 [1.80]	0.343 [2.02]*	0.121 [0.49]
Number of observations	175	175	174	174	140	140
R-squared	0.12	0.30	0.10	0.35	0.11	0.42

Robust t-statistics in brackets. \* significant at the 5 percent level. \*\* significant at the 1 percent level. Observations are weighted by the square root of the number of students in the area.

<sup>12</sup> These results are similar to the regression results of Clotfelter (1999; p. 501). In particular, the magnitudes of all three partial correlations are similar to our estimates. His regression has the log number of students as opposed to the log number of schools and the log average size of the school districts as opposed to the log average size of the schools. Of course, his measure of segregation is between African American and white students. Our results are very similar if we include the log number of students instead of the log number of schools.

Table 8 in the Appendix shows the regression results for all years for the larger catchment areas. The results suggest that the large increase in the coefficients took place between 1992 and 2006, and the years after 2006 are characterized by further increases, with ups and downs without any clear pattern.

After the cross-sectional regressions, we turn to the regressions estimated in long differences: changes between 1980 and 2011.

Table 6 shows the results, and Table 9 in the Appendix has the appropriate summary statistics. Table 10 and Table 11 show the corresponding results separately for the communist period (1980 to 1989) and the post-communist period (1989 to 2011).

*Table 6.*

**Changes in school segregation and changes in the size of the educational market, schools and the Roma population from 1980 to 2011. Regression results.**

Dependent variable: $S_{2011} - S_{1980}$ (change in the index of segregation)	Larger catchment areas	Micro- regions	Larger municipalities
Log change in number of schools	0.170 [3.23]**	0.116 [2.43]*	0.018 [0.42]
Log change in average school size	0.068 [1.30]	-0.01 [0.19]	-0.059 [1.08]
Change in fraction of Roma students	0.605 [4.31]**	0.792 [7.39]**	0.839 [3.84]**
Constant	0.098 [3.01]**	0.057 [2.05]*	-0.016 [0.56]
Number of observations	175	174	140
R-squared	0.17	0.23	0.14

Robust t-statistics in brackets. \* significant at the 5 percent level. \*\* significant at the 1 percent level. Observations are weighted by the square root of the number of students in the area.

The results from these regressions show the extent to which the areas that experienced larger-than-average increases in the number of schools, school size or the fraction of Roma students tend to be characterized by larger-than-average increases in school segregation. When interpreting the results, one must keep in mind that, on average, school segregation increased, the number of schools decreased (except in the larger municipalities), the average school size decreased (especially in the larger municipalities) and the fraction of Roma students increased during the observed period. These trends were the most pronounced during the post-communist period (1989 to 2011). On average, there were no significant shifts before 1989, but the variation in changes was substantial even then, so that interesting associations can be identified.

The results are qualitatively similar to the cross-sectional associations measured in 2011. An increase (decline) in the number of schools by 10 percent is associated with an increase (decline) in the index of school segregation by one to two percentage points in the larger catchment areas and the micro-regions. These magnitudes are actually stronger than the cross-sectional estimates in 2011: a one standard deviation (0.35 to 0.42) higher increase in the log number of schools is associated with an approximately one third of a standard deviation (0.14 to 0.16) increase in segregation. No association is present within the larger municipalities. The changes in the average school size are not associated with changes in segregation, holding the number of schools and the ethnic composition constant. Similar to the cross-sectional results, the change in the fraction of Roma students is the strongest predictor of changes in school segregation. The magnitudes are similar to the cross-sectional associations (a one standard deviation increase in the fraction of Roma students is associated with a half of a standard deviation increase in segregation).

## **8. CONCLUSIONS**

In this paper, we documented the degree of between-school segregation of Roma versus non-Roma students between 1980 and 2011. We showed the long-run trends and geographic distributions as well as the regression estimates of some robust associations.

An important contribution of our paper was the definition of school catchment areas: clusters of villages, towns and cities that are closed in terms of student commuting in the 2000s. This geographic aggregation allows school segregation to be analyzed at the level of the smallest and most relevant geographic area. The use of the catchment areas also allows school-level information to be used to estimate figures for the people living in those areas, such as the share of the Roma minority. Those data are available upon request from the authors.

From a theoretical point of view, our most interesting results are the regression estimates. These estimates show that the size of the educational markets (defined as the number of schools) is strongly and positively associated with between-school segregation. This association is consistent with the notion that school choice and selective commuting are among the most important mechanisms behind segregation, and the size of the market increases differentiation between schools, therefore providing a higher incentive to commute. This explanation is, however, not the only possible one. The fraction of Roma students in the area is an even stronger predictor of segregation. Explaining this association could be even harder. However,

both associations are robust in the sense that they are identified from the cross-section as well as from the long differences, and analogous results for both are found in the U.S. as well.

From a policy perspective, another interesting finding is the coincidence of an apparent trend break in segregation between 2006 and 2008, coincident with the timing of the most intensive desegregation campaigns. Unfortunately, the quality of the data does not allow for a robust analysis here. Improving the data quality by implementing the full coverage of schools is necessary for fine analysis of the effects of desegregation policies and other aspects of school segregation in Hungary.

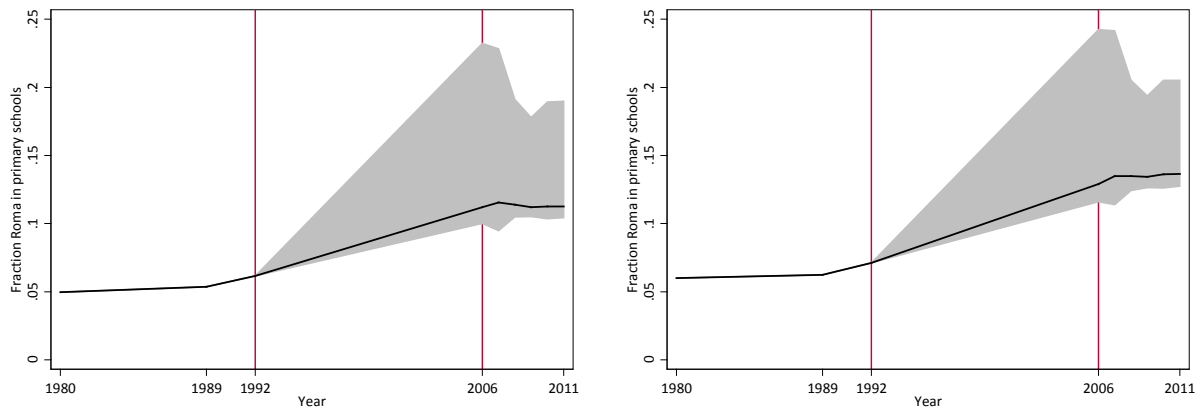
## REFERENCES

- Clotfelter, Charles T (1999): Public school segregation in metropolitan areas. *Land Economics*. Vol. 75. No. 4. November.
- Clotfelter, Charles T (2004): *After Brown. The rise and retreat of school desegregation*. Princeton University Press. Princeton and Oxford.
- Horn, Daniel (2012), “Early selection in Hungary. A possible cause of high educational inequality.” Mimeo.  
[http://mta.academia.edu/DanielHorn/Papers/1646306/Early\\_Selection\\_in\\_Hungary - A possible cause of high educational inequality](http://mta.academia.edu/DanielHorn/Papers/1646306/Early_Selection_in_Hungary_-_A_possible_cause_of_high_educational_inequality)
- Kertesi, Gábor and Gábor Kézdi (1998), *A cigány népesség Magyarországon. Dokumentáció és adattár*. Socio-typo, Budapest.
- Kertesi, Gábor and Gábor Kézdi (2011), “The Roma/non-Roma test score gap in Hungary.” *American Economic Review* Vol. 101. No. 3. 2011. 519-525.

## APPENDIX: ADDITIONAL FIGURES AND TABLES

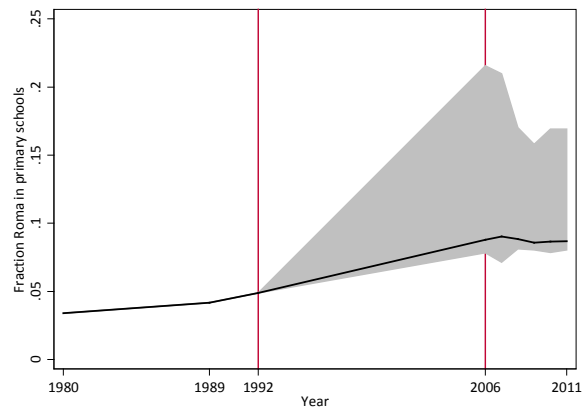
Figure 5.

**The time series of the fraction of Roma students primary schools in larger catchment areas (panel A), micro-regions (panel B) and larger municipalities (towns and cities; panel B). Hungary, 1980 through 2011.**



Panel A. Larger catchment areas

Panel B. Micro-regions

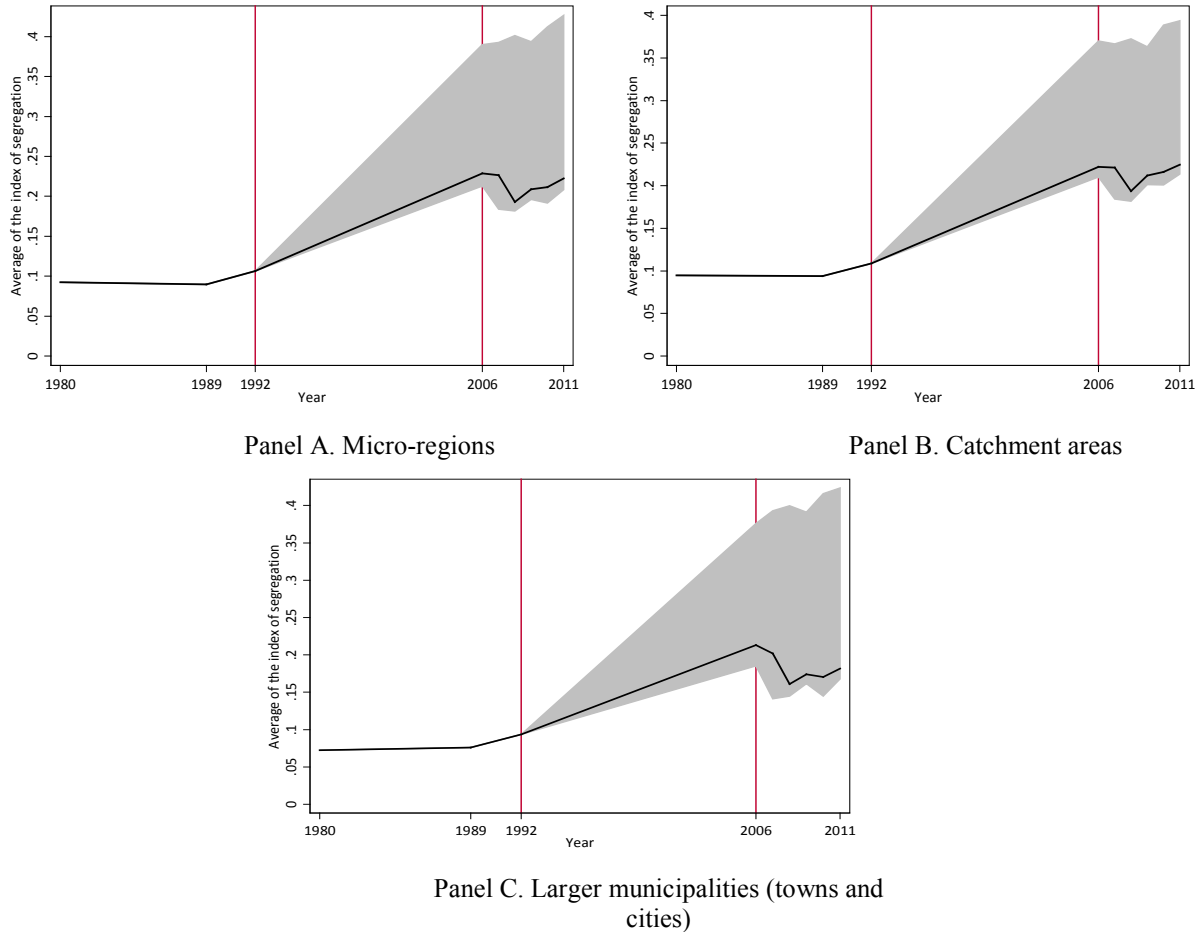


Panel C. Larger municipalities (towns and cities)

Note: the lines are based on our benchmark imputations for missing data after 2006. Grey area: conservative lower and upper bounds using alternative imputations.

Figure 6.

**The time series of the index of ethnic segregation between primary schools. Average of the index (weighted by number of students) in micro-regions (panel A), catchment areas (panel B) and larger municipalities (towns and cities; panel B). Hungary, 1980 through 2011.**



Note: the lines are based on our benchmark imputations for missing data after 2006. Grey area: conservative lower and upper bounds using alternative imputations.

Table 7.

**Summary statistics for school segregation and the size of the educational market, schools and the Roma population, corresponding to the regressions in**

**The fraction of Roma students in the area is the strongest predictor of school segregation, with increasing magnitude over time and across geographic units (being the strongest predictor within towns and cities). Towns and cities that had a one percentage point higher fraction of Roma students in their schools were characterized by a 0.75 percentage point higher index of segregation. In terms of standardized coefficients, the towns and cities with a fraction of Roma students that is higher by one standard deviation (0.1) were characterized by a half of a standard deviation (0.14) higher index of segregation on average, holding the number of schools and the average school size constant.**

Table 5.

	Larger catchment areas		Micro-regions		Larger municipalities	
	1980	2011	1980	2011	1980	2011
Mean						
S (index of segregation)	0.08	0.17	0.10	0.20	0.06	0.12
Log number of schools	2.05	1.83	8.43	7.93	7.81	7.41
Log average school size	5.71	5.41	5.60	5.33	6.15	5.60
Fraction of Roma students	0.08	0.18	0.08	0.17	0.06	0.12
Standard deviation						
S (index of segregation)	0.09	0.17	0.08	0.14	0.11	0.14
Log number of schools	0.91	0.93	0.72	0.80	0.90	0.89
Log average school size	0.44	0.31	0.41	0.27	0.39	0.28
Fraction of Roma students	0.06	0.14	0.06	0.14	0.05	0.10
Number of observations	175	175	174	174	140	140

Table 8.

**School segregation and the size of the educational market, schools and the Roma population. Cross-sectional regressions for all years for the larger catchment areas.**

Dependent variable: S (index of Segregation)	Larger catchment areas								
	1980	1989	1992	2006	2007	2008	2009	2010	2011
Log number of schools	0.022	0.022	0.025	0.039	0.041	0.038	0.038	0.042	0.055
	[2.45]*	[3.00]* *	[3.47]* *	[3.38]* *	[3.37]* *	[4.05]* *	[4.30]* *	[3.89]* *	[4.98]* *
Log average school size	-0.024	-0.010	-0.006	0.073	0.065	0.051	0.050	0.027	-0.022
	[1.51]	[0.75]	[0.39]	[2.87]* *	[2.46]*	[1.66]	[0.99]	[0.57]	[0.58]
Fraction of Roma students	0.439	0.464	0.511	0.555	0.620	0.635	0.595	0.617	0.661
	[4.27]* *	[4.54]* *	[5.36]* *	[6.12]* *	[6.45]* *	[6.15]* *	[5.52]* *	[5.66]* *	[6.86]* *
Constant	0.142	0.055	0.019	-0.379	-0.354	-0.296	-0.274	-0.156	0.076
	[1.41]	[0.62]	[0.22]	[2.95]* *	[2.53]*	[1.75]	[1.01]	[0.60]	[0.36]
Number of observations	175	175	175	175	175	175	175	175	175
R-squared	0.12	0.19	0.22	0.24	0.26	0.27	0.20	0.23	0.30

Robust t-statistics in brackets. \* significant at 5 per cent. \*\* significant at 1 per cent. Observations are weighted by the square root of the number of students in the area.



Table 9.

**Summary statistics of the change in school segregation, the size of the educational market, the size of schools and the size of the Roma population. 1980 to 2011; corresponding to the regressions in**

**Table 6.**

	Larger catchment areas	Micro-regions	Larger municipalities
Mean			
Log change in S (index of segregation)	0.09	0.10	0.06
Log change in number of schools	-0.22	-0.23	0.15
Log change in average school size	-0.30	-0.27	-0.55
Change in fraction of Roma students	0.10	0.09	0.06
Standard deviation			
Log change in S (index of segregation)	0.16	0.14	0.16
Log change in number of schools	0.42	0.35	0.35
Log change in average school size	0.41	0.30	0.39
Change in fraction of Roma students	0.10	0.09	0.07
Number of observations	175	174	140

Table 10.

**Changes in school segregation and changes in the size of the educational market, schools and the Roma population. 1980 to 1989 and 1989 to 2011. Regression results.**

Dependent variable:	1980 to 1989			1989 to 2011		
	Larger catchment areas	Micro-regions	Larger municipalities	Larger catchment areas	Micro-regions	Larger municipalities
DS (change in the index of Segregation)						
Log change in number of schools	0.085 [2.30]*	0.002 [0.05]	0.112 [1.24]	0.171 [2.84]**	0.095 [2.28]*	-0.013 [0.26]
Log change in avg. school size	-0.001 [0.02]	-0.068 [1.86]	-0.022 [0.26]	0.061 [1.04]	-0.028 [0.56]	-0.072 [1.30]
Change in fraction of Roma students	0.612 [1.29]	1.194 [2.16]*	1.708 [1.73]	0.564 [4.24]**	0.69 [7.22]**	0.759 [4.27]**
Constant	-0.009 [1.42]	-0.007 [1.05]	-0.019 [1.33]	0.109 [3.32]**	0.06 [2.51]*	-0.005 [0.17]
Number of observations	175	174	140	175	174	140
R-squared	0.06	0.08	0.13	0.19	0.25	0.19

Robust t-statistics in brackets. \* significant at 5 per cent. \*\* significant at 1 per cent. Observations are weighted by the square root of the number of students in the area.

Table 11.

**Summary statistics of the change in school segregation, the size of the educational market, the size of schools and the size of the Roma population. 1980 to 1989 and 1989 to 2011.**

	1980 to 1989			1989 to 2011		
	Larger catchment areas	Micro-regions	Larger municipalities	Larger catchment areas	Micro-regions	Larger municipalities
Mean						
Log change in S	-0.01	-0.01	-0.01	0.10	0.11	0.07
Log change in number of schools	-0.03	-0.06	0.03	-0.19	-0.17	0.12
Log change in avg. school size	0.02	0.04	0.04	-0.32	-0.31	-0.59
Change in fraction of Roma students	0.00	0.00	0.00	0.10	0.09	0.06
Standard deviation						
Log change in S	0.07	0.07	0.10	0.15	0.11	0.12
Log change in number of schools	0.19	0.15	0.22	0.38	0.30	0.32
Log change in avg. school size	0.18	0.14	0.23	0.41	0.28	0.34
Change in fraction of Roma students	0.02	0.01	0.03	0.09	0.09	0.07
Number of observations	175	174	140	175	174	140