

Adamecz-Völgyi, Anna; Scharle, Ágota

**Working Paper**

## Books or babies? The incapacitation effect of schooling on minority women

GLO Discussion Paper, No. 474

**Provided in Cooperation with:**

Global Labor Organization (GLO)

*Suggested Citation:* Adamecz-Völgyi, Anna; Scharle, Ágota (2020) : Books or babies? The incapacitation effect of schooling on minority women, GLO Discussion Paper, No. 474, Global Labor Organization (GLO), Essen

This Version is available at:

<https://hdl.handle.net/10419/213878>

**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.*

# Books or babies? The incapacitation effect of schooling on minority women\*

Anna Adamecz-Völgyi<sup>†</sup>      Ágota Scharle<sup>‡</sup>

February 13, 2020

## Abstract

This paper examines the effects of an increase in the compulsory school leaving age on the teenage fertility of Roma women, a disadvantaged ethnic minority in Hungary. We use a regression discontinuity design identification strategy and show that the reform decreased the probability of teenage motherhood among Roma women by 13.4-26.0% and delayed motherhood by two years. We separate the incapacitation and human capital effects of education on fertility by exploiting a database that covers live births, miscarriages, abortions and still births, and contains information on the time of conception precise to the week. We find that longer schooling decreases the probability of getting pregnant during the school year but not during summer and Christmas breaks, which suggests that the estimated effects are generated mostly through the incapacitation channel.

Keywords: Education, Compulsory school leaving age, Teenage fertility, Disadvantaged ethnic minorities, Regression discontinuity design  
JEL: J13, C21, I26

## 1 Introduction

Teenage fertility disrupts human capital accumulation and is one of the most important sources of intergenerational poverty transmission (Bonell, 2004). There is compelling evidence on the negative health, social, and economic consequences of teenage childbearing, such as lower educational attainment and labour market attachment rates, poorer soft skills, higher rates of benefit receipt, and infant mortality (Klepinger et al.,

---

\*The present document has been produced using the 2001 and 2011 Hungarian Census and the Vital Statistics Database files of the Hungarian Central Statistical Office. The calculations and the conclusions within the document are the intellectual product of Anna Adamecz-Völgyi and Ágota Scharle as authors. The authors gratefully acknowledge financial support from the Hungarian National Scientific Research Program (OTKA), Grant no. 112792. We are thankful for the data access to the Institute of Economics of the Hungarian Academy of Sciences, to Gábor Kézdi, Tibor Czeglédi and the research room team of the Institute of Economics at the Hungarian Central Statistical Office. We are grateful to Gábor Kézdi, Hessel Oosterbeek, Róbert Lieli, Gábor Kertesi, Dorottya Szikra, Judit Durst, Balázs Váradi, Balázs Kapitány, Ildikó Husz, Béla Janky and Nikki Shure for their useful comments and suggestions. We thank Flóra Samu her great work as research assistant on a previous version of this paper.

<sup>†</sup>Corresponding author. Department of Social Science, UCL Institute of Education, 20 Bedford Way London, WC1H0AL and Budapest Institute for Policy Analysis 84. Dohány utca, 1074 Budapest. E-mail: a.adamecz-volgyi@ucl.ac.uk. ORCID: 0000-0001-6408-5830.

<sup>‡</sup>Budapest Institute for Policy Analysis 84. Dohány utca, 1074 Budapest. E-mail: agota.scharle@budapestinstitute.eu. ORCID: 0000-0003-4838-8764

1999; Chevalier and Viitanen, 2003; Fletcher and Wolfe, 2009; Fletcher and Padrón, 2016; Wilson, 2017). Negative impacts have been found to carry on to the next generation as well: Navarro Paniagua and Walker (2012) show that daughters of teenage mothers are less likely to have post compulsory education and more likely to become teenage mothers themselves.

Due to its high opportunity costs, the prevalence of teenage motherhood has been declining in most developed countries. In the 2000s, the share of women aged 15-19 giving birth per year was between 0.5-2.5% in most Western-European countries<sup>1</sup>. However, teenage motherhood is still very common among disadvantaged ethnic minorities living in developed countries. Examples include Mexican women in the US, women of Pakistani and Bangladeshi origin in the UK, Turkish women in Belgium and France, and Roma women throughout Europe. In Hungary, more than a quarter of Roma women have their first child by their 18th, and almost half of them have their first child by their 20th birthday.

Bean and Swicegood (1985) argue that fertility differences between minority and majority women are generated by social and economic exclusion. The opportunity costs of early childbearing tend to be lower for minority women because their perceived and actual future economic opportunities are constrained, independently of their fertility patterns. Lopoo (2004) finds that disadvantaged teenagers with working mothers (which is assumed to improve girls' perceived labour market prospects) are 18% less likely to give birth compared to their peers who have non-working mothers. Wolfe et al. (2007) propose that girls' fertility decisions depend on their perceptions of the consequences of childbearing for future income and relationships. They show that negative consequences observed in the older generation reduce the teen fertility rate. Kearney and Levine (2012) conclude that young women choose motherhood instead of investing in the development of their own human capital because "they feel they have little chance of advancement". Kearney and Levine (2014) provide evidence that women of low social and economic status are more likely to give birth at a young age if they live in a US state with a higher level of income inequality.

The majority of earlier studies focus on the fertility effects of education in general, without making a distinction between minority or majority ethnic groups. Their identification strategy is usually based on changes in education policy (school entry rules: McCrary and Royer (2011); Tan (2017); length of school day: Berthelon and Kruger (2011); length of schooling: Black, Devereux, and Salvanes (2008); Silles (2011); Cygan-Rehm and Maeder (2013); Wilson (2017); Geruso and Royer (2018); expansion of post-compulsory education: James and Vujčić (2019)). Most papers find that having more education does reduce the probability of early childbearing and/or delays motherhood. There is much less evidence on the potential impact of education on teenage motherhood among ethnic minority women in particular. The only such study we have found so far is Bifulco, Lopoo, and Oh (2015) who examine the effects of school desegregation in the US and find that it has not reduced the fertility of black teenagers. They conclude that, although school desegregation has generated benefits for black students on several domains, decreasing early childbearing is not one of them. This finding is consistent with the above explanation: desegregation tackled the quality of education available to black girls, but did not address expected labour market prospects.

The impact of education on fertility may work via two main channels: incapacitation or human capital investment. While the incapacitation effect can only influence conception during the period of education (extended by some exogenous event), the human capital effect may materialise during or after this period as well. Incapacitation<sup>2</sup> refers to the possibility that teenagers may not have the desire, time, or opportunity to have a child while they are in school. This does not necessarily imply that teenagers are less sexually active

---

<sup>1</sup>Source: Eurostat

<sup>2</sup>The incapacitation effect is sometimes referred to as the incarceration effect (for example, in Black, Devereux, and Salvanes (2008)). We deliberately chose not to use this expression for two reasons. First, it alludes to the physical isolation of young women as a potential solution to reducing teenage motherhood. Second, it implicitly suggests that being in school reduces the likelihood of sexual intercourse, which has not been documented and may not be true at all.

while in school. It might also happen that they have easier access to contraception, morning-after pills or abortion.

The human capital effect is commonly described as an increase in education that increases the expected wage, which in turn increases the opportunity cost of having a teen birth. Human capital development might also bring non-pecuniary benefits such as increased awareness of contraception and/or the potential negative consequences of early motherhood, or improved mating strategies. Most of the literature distinguishes between incapacitation and human capital channels by looking for effects on fertility beyond the age directly affected by the reform. Investigating reforms raising the compulsory school leaving age, [Black et al. \(2008\)](#), [Cygan-Rehm and Maeder \(2013\)](#) and [Silles \(2011\)](#) find evidence on both incapacitation and human capital effects. [Geruso and Royer \(2018\)](#) follows a different approach: assuming that all pregnancies lasted for 9 months, they deduce the approximate conception time of babies and look at the effects of extended schooling on the probability of pregnancies conceived before, during, or after the new schooling period. They only find statistically significant effects during the new schooling period and interpret this as evidence for incapacitation effects. [Berthelon and Kruger \(2011\)](#) concentrate on the incapacitation effect of education and show that longer school days reduce the probability of motherhood and criminal behavior among teenage girls aged 15-19.

Our contribution to the literature is threefold. First, we examine how extended compulsory education affects teenage childbearing among ethnic minority women characterized by poor labour market opportunities and low opportunity cost of motherhood. We look at Roma women, who make up the largest ethnic minority in Hungary, accounting for about 5-7% of the total population, or 500,000-700,000 people. Unlike in Western Europe, the vast majority of Roma communities in Hungary settled down by the end of the 19th century and travellers practically disappeared by the early 1990s ([Crowe, 1995](#); [Kemény and Janky, 2005](#)). School attendance among the Roma steadily increased after the 1960s and by the late 1990s, the Roma-non Roma gap in children completing the 8-year elementary school was reduced to a negligible level ([Hajdú et al., 2014](#)). However, Roma have continued to experience social exclusion. Belonging to the Roma minority is strongly correlated with poverty, social exclusion, and long term unemployment. Roma children are more likely to attend segregated schools, which are often of lower quality ([Kemény and Janky, 2005](#); [Ladányi and Szelényi, 2002](#); [Kertesi, 2005](#); [Gábos et al., 2006](#)). The Roma employment gap is especially large for women ([FRA, 2014](#)). Returns to education rose sharply during the 1990s, but labour market opportunities for the low educated and the Roma declined. Teenage fertility of non-Roma women is just as low in Hungary as in most Western-European countries, while among Roma women, it is comparable to levels measured in developing countries such as the Congo or Kenya ([Janky, 2007](#); [UNFPA, 2013](#)).

We estimate the intention-to-treat (ITT) effects of longer schooling by exploiting the reform of compulsory schooling introduced in 1996. The reform increased the compulsory school leaving (CSL) age from 16 to 18, applicable to those starting elementary school in September 1998. Our identification strategy is based on the elementary school enrollment rule that allows us to set up a fuzzy regression discontinuity design (RDD) identification strategy around a cutoff date of birth. We find that the reform decreased the probability of teenage motherhood among Roma women by 3.5-6.8 percentage points (13.4-26.0%), depending on the empirical method we use, and caused a two-year delay in motherhood. We find no effects among non-Roma women.

Our second contribution to the literature is that we exploit a database of all known pregnancies, including live births, abortions, fetal losses, and still births, linked to a large subsample of Roma women in the 2011 Hungarian Census. The data allow us to reconstruct the conception time of pregnancies precise to the week and directly investigate the effect of the reform on the probability of getting pregnant, having an abortion, and ending a pregnancy with an abortion. We find a negative effect on the probability of having an abortion and on the probability of ending a pregnancy with an abortion (conditional on getting pregnant). This last

result is estimated on the small subsample of pregnant Roma women and is not significant; however, the large coefficient might suggest that the reform decreased the probability of unwanted pregnancies more.

Third, we separate the incapacitation and human capital effects of education by showing that the reform decreased the probability of getting pregnant during the school year but not during summer and Christmas breaks, indicating that the effect is generated mainly by the incapacitation channel. This result is in line with the conclusion of [Geruso and Royer \(2018\)](#) and consistent with the explanation that the poor labour market prospects of ethnic minority women keep the opportunity cost of childbearing low. In terms of showing that being in school has a contemporaneous incapacitation effect, our results are analogous to those of [Berthelon and Kruger \(2011\)](#) as mentioned above and [Jacob and Lefgren \(2003\)](#) who show that teenagers are less likely to engage in criminal behavior on schooldays.

The rest of the paper is structured as follows. Section 2 describes the data and the fertility pattern of Roma adolescents. Section 3 summarizes the reform and the institutional background. The identification strategy and empirical methods are detailed in Section 4. Section 5 presents the results, Section 6 reports several robustness checks while Section 7 concludes.

## 2 Data and the teenage fertility of Roma women

### 2.1 Data

This paper uses three data sources: the 2001 Hungarian Census, the 2011 Hungarian Census, and the Vital Statistics database. As it is detailed in Section 4, we estimate the effects of the reform in a fuzzy RDD framework, where compliance with the elementary school enrollment rule generates a discontinuity in the probability of being exposed to the new CSL age scheme around a cutoff date of birth. Due to the enrollment rule, those born on June 1, 1991, or later, were more likely to start school after the reform than those born before this date. We use the 2001 Hungarian Census to demonstrate the compliance to the enrollment rule, i.e. the jump in the probability of being treated around the cutoff, June 1, 1991. The 2001 Hungarian Census was implemented in the spring of 2001 when the relevant cohort was 9-10 years old. It records the birth year and month, and, for those in school, it registers which grade of school they were attending at that time. We do not observe the age at which children started school, and thus whether they were affected by the reform. We infer this information from the grades students are in 2001 and show in Section 3.2 that there is indeed a jump in the probability of being treated around the cutoff.

Similarly to the earlier Census, the 2011 Hungarian Census also contains information about the entire Hungarian population. It observes the relevant cohort in October 2011, when aged 19-20. In addition to the birth year and months of individuals, it includes their exact date of birth as well, along with information on gender, ethnicity, and the number of successfully completed grades in school. It registers the year and months of birth of the first five children of women, independent from whether the children live with their mothers or not. We use the 2011 Hungarian Census data to estimate whether Roma women became mothers during their teenage years.

The Vital Statistics database covers all pregnancy-ending medical events. It records all live births, still births, abortions, and miscarriages. The data come from surveys completed by women in the hospital at the time of the event. The Vital Statistics registers the mother's day of birth, the date of the event, how far along the pregnancy was at the time of the event (measured by week), the gender of the child in the case of giving birth, and previous pregnancy history.

We are interested in understanding how school attendance affects Roma girls' fertility decisions. As teenagers usually drop out of school after having a child, we concentrate on their first live births, and, their pregnancy-ending events before having their first child. The Vital Statistics database does not have ethnic

markers, while the 2011 Hungarian Census does. Therefore, we link all pregnancy-ending events of childless women, and all first live births to the 2011 Hungarian Census, which will herein be referred to as “linked data”.

Our linking procedure is based on the common variables in the two data sets: exact date of birth of the mother, place of residence, and, in the case of live births, the year and month of the event. We can link two observations from the two data sets together if they are not duplicates, i.e. they belong to exactly one woman born on one particular day who lives in one particular settlement. As a consequence, we can only use data of those living in settlements with less than 50,000 inhabitants for the linking procedure. From this subsample, we are able to link 40% of pregnancy-ending events to the 2011 Hungarian Census.

The linking procedure has to fulfill two requirements:

- Linking a certain pregnancy event to the 2011 Hungarian Census has to be random, i.e. it should not be correlated with individual characteristics related to fertility or education; and
- Linking a certain pregnancy event to the 2011 Hungarian Census cannot be correlated with being born right below or above the cutoff, June 1, 1991.

Tables 9-11 in Appendix A show how these two requirements were fulfilled. The records of live births can be linked the most reliably to the 2011 Hungarian Census because the year and month of birth-giving are included in both databases. However, the linking of other pregnancy events seems to be related to both personal characteristics and the reform in that:

- The events of younger and more educated women are more likely to be linked; and
- The events of those born after June 1, 1991. are less likely to be linked.

Although our linking procedure is not random, the magnitude of the potential bias is small. We estimate the effect of the reform on outcome variables available both in the Census and the linked data to consider the size of the potential bias, and we find very similar results (see Table 5). Furthermore, controlling for year and month of birth, settlement size, and county fixed effects reduces the systematic relationship between the intention-to-treat status of women and the linking of their Vital Statistics events substantially (see Table 11 in Appendix A). Thus, we reproduce our main estimations using these additional control variables as well and they do not change our results (see Table 8).

We construct the following outcome variables:

- *Probability of giving birth by age 16-20*, from both the 2011 Hungarian Census and the linked data. We define age at first birth-giving as a continuous variable. Age categories are defined as  $(0,age]$ . For example, “having the first child by age 16” means giving birth for the first time at age  $(0;16]$ ; either before, or exactly on, the day of the mother’s sixteenth birthday. Giving birth one day later, therefore, is captured as “having the first child by age 17”.
- *Probability of getting pregnant by age 18*. From the 2011 Hungarian Census, we calculate the approximate conception time assuming that all pregnancies lasted for 9 months. Using this assumption, we construct the time of conception of live births by monthly precision. The linked data have information on the week of pregnancy at the time of all pregnancy-ending events so we can reconstruct the conception time of all linked pregnancies by weekly precision.
- *Probability of having an abortion* from the linked data.

- *Probability of having an abortion conditional on getting pregnant*, from the linked data.
- *Probability of getting pregnant during school years*. To test the incapacitation effect of education, we construct binary variables to capture whether the calculated conception time falls on a date during the school year or during a school break. In Hungary, the school year starts on 1 September and lasts until the middle of June with a two-week Christmas break in the second half of December. As in the 2011 Hungarian Census we see conception time by monthly precision, we can pin down the school year as either Sept - May (loosing two weeks of school time in June) or as Sept - June (loosing two weeks of summer breaks in June). Both options lead to similar results. In the linked data where conception time is captured precise to the week, we define school years as from 1 Sept to 14 June, excluding the two weeks of Christmas breaks between 15 Dec - 31 Dec.
- *Probability of getting pregnant in summer breaks*. Again, in the 2011 Hungarian Census we can define summer breaks either as June-Aug or July-Aug and both lead to similar results, while in the linked data we define them as 15 June - 31 Aug.
- *Probability of getting pregnant in Christmas breaks*. We capture the two-week Christmas breaks only in the linked data and we define them as 15 Dec - 31 Dec.

## 2.2 Identifying Roma in the 2001 and 2011 Hungarian Census

Both the 2001 and the 2011 Hungarian Census rely on individuals' self-identification of Roma ethnicity, and both suffer from an underreporting problem to different extents. The 2011 Hungarian Census registers almost 50% more Roma people than the 2001 Census for three reasons. First, it explicitly asks respondents whether they have dual-nationality identities, whereas the 2001 Hungarian Census only allows three choices of answers to the question about national identity. The 2011 Hungarian Census first has a question about national identity, set up similarly to that of the 2001 Hungarian Census. Then, it explicitly asks the respondents whether they feel that they belong to any other nationality or ethnicity group, in addition to the one which was identified in the first question. Compared to the previous census, this new method allows for a substantially improved identification of the Roma population (Messing, 2011). Second, civil organizations campaigned in 2011 for the Roma to reveal their ethnic identity. In some 2011 Hungarian Census tracts, census takers who themselves identified as Roma were employed (Institute, 2013). Unfortunately, the details of these initiatives are not documented publicly. According to statistics on the "We belong here!" (Ide tartozunk!) campaign, 1,046 Roma survey takers were employed. However, it is not known which census tract they worked in, or how many questionnaires they registered (Data source: Open Society Foundation)<sup>3</sup>. Lastly, there must have been demographic changes in the Roma population between 2001 and 2011, including population growth and geographical migration. However, we have no information on the magnitude of these phenomena.

As a result of these three factors, the 2001 Hungarian Census reported a total of 217,097 Roma persons, while the 2011 Hungarian Census reported 315,525 Roma people<sup>4</sup>. In spite of this improvement, demographers estimate that the 2011 Hungarian Census still identifies only half of the total number of the Roma population (Habicsek, 2007). We know from a regular household survey recording Roma identity both as reported by the respondent and as assessed by the surveyor that individuals identifying themselves as Roma are more likely to live in smaller settlements, to be less educated, to be unemployed, and to live on lower and less stable income, than those identified as Roma by others but not by themselves (Tárki, 2013). Those

<sup>3</sup>The total number of Census tract units is over 47,000 (Data source: Hungarian Statistical Office).

<sup>4</sup>Own estimation from the 2001 and 2011 Hungarian Censuses.



Table 1: The prevalence of teenage childbearing before the reform (cohorts born in 1988-1990)

Probability of giving birth by age ...	All women	Roma women	Non-Roma women	Ethnicity unknown
16	0.007 (0.000)	0.065 (0.003)	0.004 (0.000)	0.007 (0.085)
17	0.017 (0.000)	0.150 (0.004)	0.010 (0.000)	0.019 (0.136)
18	0.032 (0.000)	0.260 (0.005)	0.020 (0.001)	0.038 (0.192)
19	0.052 (0.001)	0.379 (0.006)	0.034 (0.001)	0.063 (0.243)
20	0.073 (0.001)	0.472 (0.006)	0.052 (0.001)	0.090 (0.286)
No. of obs.	169,375	8,020	154,055	7,300

Data source: estimation based on the 2011 Hungarian Census data. Standard errors of the group means are in parenthesis. Age is a continuous variable; “by age 16”, for example, means giving birth either before, or exactly on, the mother’s 16th birthday.

who identify themselves as Roma in the Census thus belong to the lower half of the income distribution of the Roma in Hungary.

We can estimate the consistent effects of the reform on Roma women if a declaration of being Roma is not related to the reform. This issue is discussed in Section 4.

### 2.3 Teenage fertility in Hungary

Table 1 summarizes the prevalence of teenage motherhood before the reform. The probability of having the first child by age 18 is low on average (3.2%) and among non-Roma women (2.0%); however, it is very high among Roma women (26.0%). Almost half (47.2%) of Roma women become a mother by age 20.

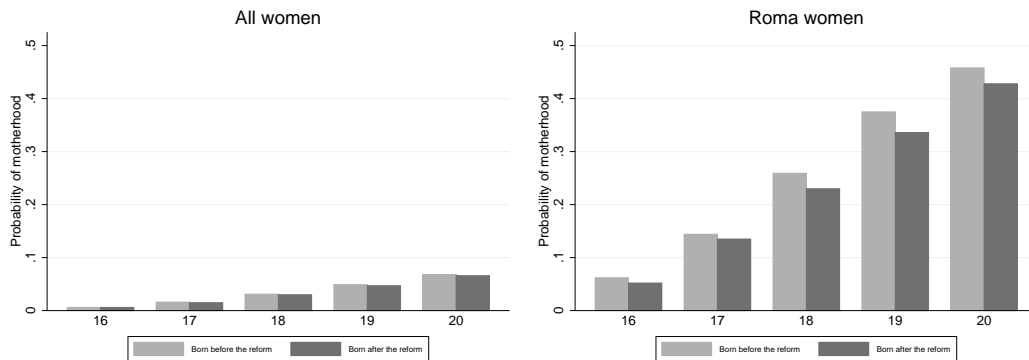
Teenage fertility has been historically high among Roma women in Hungary, but there is little evidence to explain it. Roma communities themselves are heterogeneous with respect to their fertility patterns (Janky, 2005). In some local Roma communities, a 14-year-old girl or boy is treated as an adult and most children of adolescent Roma mothers are born in stable relationships. Out of Roma mothers born in 1990 who gave birth by age 18, 81% were either married or lived with a long term partner in 2011<sup>5</sup>. Teenage fertility has been decreasing in Hungary even among Roma women, but it has remained stable in some groups of the poor (Szikra, 2010) and in some marginalized Roma communities (Durst, 2007).

Figure 1 presents the probability of teenage motherhood among those born right below versus right above the cutoff. The raw data suggest that Roma women born above the cutoff, who were more likely to go to school after the reform, were less likely to have their first child by age 18-20, than Roma women born below the cutoff. On the sample of all women, we do not see a difference.

<sup>5</sup>Own calculation from the 2011 Census.



Figure 1: The probability of having the first child by age 16-20 before and after the reform



Data source: own estimation from the 2011 Hungarian Census. Born before the reform: women born at most 180 days before June 1, 1991. Born after the reform: women born at most 180 days after June 1, 1991. No. of observations: All women: 29,275 and 29,353; Roma women: 1,381 and 1,394.

### 3 Institutional background and the reform

#### 3.1 Roma students in the Hungarian education system

The educational outcomes of Roma students, both male and female, lag behind in several aspects. They are more likely to go to lower-quality schools and to repeat grades in both elementary and secondary school. If they are admitted to secondary school they are more likely to go to lower-level secondary schools than their non-Roma peers. The achievement gap in standardized reading and math test scores between Roma and non-Roma students is comparable to the size of the black to white test score gaps of the United States in the 1980s (Kertesi and Kézdi, 2011, 2014). About 13-50% of this gap comes from the fact that Roma students do not have access to high-quality education, and the remainder is accounted for by differences in social background (Kertesi and Kézdi, 2014). The Hungarian education system is rated as one of the worst among the OECD countries in terms of offsetting social disadvantages. According to the 2012 Program for International Student Assessment (PISA) study, family background explains one of the largest shares of the variance in mathematics test results in Hungary among the OECD countries (OECD, 2014). With free elementary school choice and early tracking, the Hungarian education system is highly segregative with respect to disadvantaged students in general, and with respect to Roma students in particular (Kertesi and Kézdi, 2009).

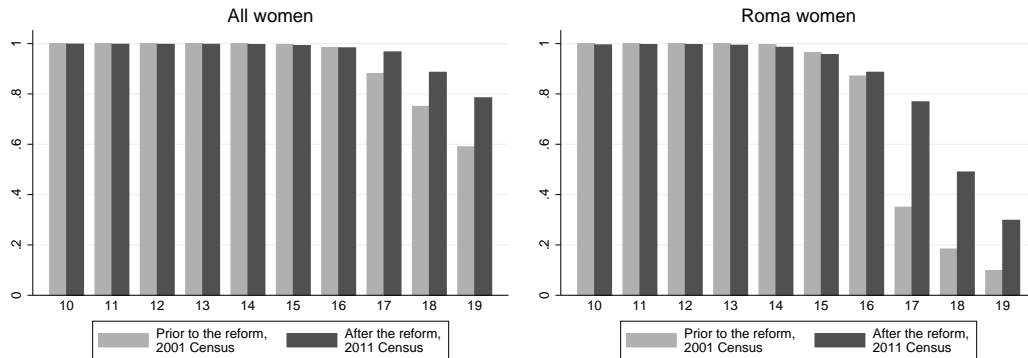
#### 3.2 Compulsory education and the reform

Before the reform, students had to stay in school until the end of the academic year in which they reached age 16. The reform, the Public Education Act (1996), extended compulsory school attendance by two years, from age 16 to age 18 and first applied to those starting school in 1998. The regulation allowed few exceptions and sanctions on noncompliance were relatively harsh. Schools were not allowed to expel pupils below the school leaving age on any account, and the obligation of school attendance could be lifted only in special cases after age 16, such as getting married or having a child. Schools were supposed to keep a record of all absences and a ministerial decree required them to notify parents after the first unjustified absence (11/1994. (VI. 8.) MKM decree). If the pupil missed another class, the school had to inform

the parents of the consequences via the municipal child welfare agency and convince them to fulfill their parental responsibilities. If a child missed 50 classes (unjustified), the school director had to inform the district notary, who could fine the parents of up to 50 thousand forints (about the monthly net minimum wage at the time). Parents not letting their children attend school for long periods could be imprisoned for up to five years (Kazuska, 2012). There is no data on the actual use of sanctions by municipal notaries, but some qualitative evidence suggests that sanctions were applied unevenly, often depending on the discretion of the notary (Mártonfi, 2011a).

To the best of our knowledge, no administrative data are published on the number of students who should legally be still in school but they are not. The Public Education Statistics of the Public Education Information System, which is the administrative school census, captures who is enrolled in school, but it does not register ethnicity. The only full coverage data source on schooling status and ethnicity at the same time is the Hungarian Census. Figure 2 shows the share of all women and Roma women in school in the 2001 Census, prior to the reform, as well as in the 2011 Census, after the reform, by age. Prior to the reform, 96.4% of 15-year-old Roma women were still in school right before reaching the actual CSL age of 16 that was in place at that time.<sup>6</sup> At age 17, only 35% of Roma women were still in school in 2001.

Figure 2: The share of women still in school before and after the reform, by age



Source: 2001 and 2011 Hungarian Census. No. of observations: 20,630 in 2011 and 33,657 in 2011.

After the reform, in 2011, the share of Roma women still in school at age 17 (i.e. before reaching the new CSL age of 18), was 76.9%. For comparison, among all women, the share of those still in school at age 17 was 96.7% in 2011. While the enforcement of the new CSL age might not have been perfect, it still contributed to increasing the share of Roma women in school at age 17 from 35.0% in 2001 to 76.9% in 2011 (Figure 2). Due to imperfect enforcement, we can only estimate the lower bound of the effects of the reform. Furthermore, as having a child could have been used as a reason to drop out of school early, some women might have faced extra incentives towards childbearing.

The reform first applied to students who enrolled in elementary school in the 1998/99 academic year, in September 1998. This cohort was already aware of the increase in compulsory schooling at age 6. Although the reform included other elements as well, increased CSL age was the only measure leading to a sharp difference between those entering elementary school in September 1997 versus in September 1998. The

<sup>6</sup>Source: own estimation from the 2001 Census. Compulsory schooling obligation can be fulfilled in homeschooling as well. However, even those in homeschooling belong to a school and are supposed to be covered by education statistics as those in school. Homeschooling is rare; in the 2013/14 academic year, the share of those in homeschooling was 0.68% (Education Office of the Ministry of Human Resources, 2014). Due to lack of data, very little evidence is available about homeschooling in general.

new legislation also laid down how to adapt the structure of secondary schools to accommodate the new CSL age by forcing all programs to offer at least four grades, and thus, last until at least age 18. This adaptation process already began in the 1998/1999 academic year and by the time the first affected cohort entered secondary education in 2006, it had been completed for half a decade.

Elementary school has 8 grades. According to the elementary school enrollment rule, that time students were expected to enter the first grade of elementary school at age 6 if they were born before June 1, but only a year later at age 7 if they were born on June 1 or after. Thus, compliance to the enrollment rule creates a discontinuity in the probability of starting school in one academic year versus the next: those born before June 1, 1991, were more likely to start elementary school under the old CSL age regime of age 6, while those born on June 1, 1991, or later, were more likely to start elementary school under the new CSL age legislation of age 7 and to stay in school two years longer.

Compliance to the enrollment rule is not perfect. On top of the rule, elementary school enrollment is a decision made during preschool, jointly by parents, preschool teachers, and if needed, pedagogical and psychological counselors who are employed by public Pedagogical Service Centers. In this period, attending preschool was mandatory from age 5. When it is due to make a decision about entering elementary school, preschool teachers have to provide an official opinion about whether a child is ready to enter school. If there are any doubts, based on the inquiry of preschool teachers, the local Pedagogical Service Center completes a “school readiness examination”.

Table 2: Average compliance to the school enrollment rule (all students)

	Share of early starters	Share of comp- liers	Share of late starters	No. of school starters
Compliance by Academic Years, share of those starting elementary school at the given year (Total=100%)				
1997/1998	0.02	0.80	0.18	127,214
1998/1999	0.02	0.78	0.20	125,875
1999/2000	0.01	0.78	0.21	121,424
Compliance by Cohorts, share of cohort size (Total=100%)				
June 90-May 91	0.02	0.79	0.19	129,489
June 91-May 92	0.02	0.78	0.20	126,294

Data source: aggregate data from the Public Education Information System - Public Education Statistics (*KIR-STAT* in Hungarian). Individual-level data is not available for this period. “Early starters” refers to those entering elementary school before reaching age 6 by May 31. “Compliers” refers to those entering elementary school according to the age rule of compulsory schooling. “Late starters” refers to those entering elementary school a year later than what is determined by the age rule.

On average, compliance with the enrollment rule is 78-80% (Table 2). Some parents have a preference for delaying enrollment, especially, but not exclusively, if their child was born right below the cutoff, because they think that being relatively old in the class is better than being relatively young. According to administrative data, about 18-20% of a cohort start elementary school later than when they are supposed to based on their date of birth (late starters). Those born above the cutoff, in June-Dec, also might start school earlier (early starters), but only if they completed at least one year in preschool that was compulsory at that time before school enrollment and passed the school readiness examination. Early school start is rare though, less than 2% on average (see Table 2). Roma students, in particular, are prone to start school later, some by even two years: partly because they tend to start the one-year compulsory preschool later than their peers and partly because due to their low socio-economic background they lag behind in terms of cognitive and non-cognitive skills and would not pass the school readiness examination (Kende and Illés, 2007). It

is not uncommon that Roma and/or disadvantaged students delay school enrollment to age 8. However, delaying school enrollment to that age (very late starters) is usually a consequence of special individual circumstances, like having special education needs or long term illness (Kende and Illés, 2007), and it is largely independent of the enrollment rule.

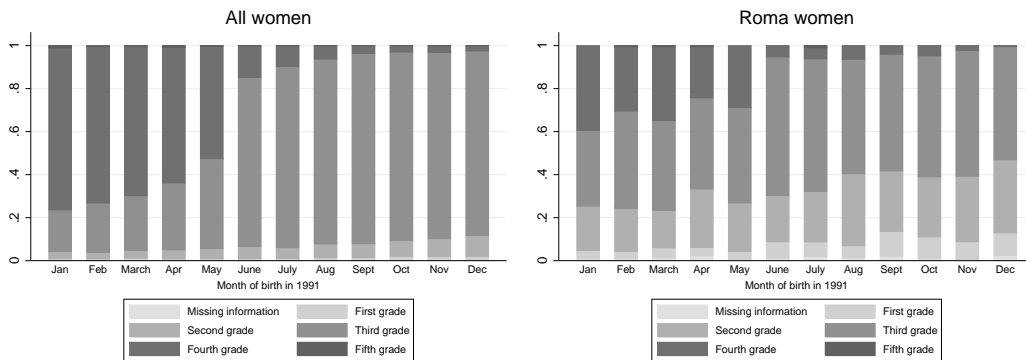
While no administrative data are available on the school enrollment of Roma students, the Census provides information on whether someone is in school, and if so, which grade she attends, at the time of the data collection. However, as grade repetition is possible and is not registered in the Census<sup>7</sup>, we can not directly check compliance to the enrollment rule. For the sample of Roma women, the only thing we can show is that there is a jump in the probability of starting school after the reform around the cutoff but we can not pin down the exact magnitude of this jump. Consequently, we will only estimate ITT effects around the cutoff.

Theoretically, those who started school after the reform, in 1998 or later, should have been in the third or a lower grade, while those who started school before the reform, in 1997 or earlier, should have been in the fourth or in a higher grade in the Spring of 2001, if they had not repeated grades. About 20% of Roma students are expected to be grade repeaters though (see Table 13 in Appendix A); thus, the only thing we know with certainty is that if someone is in the fourth or in a higher grade in the 2001 Census she must have started school already in 1997 or earlier, before the reform. If someone is in the third or in a lower grade, it might mean that they either started school in 1998 or later, after the reform, or they started school before the reform and repeated grades (potentially multiple times). Figure 3 shows the distribution of female students born in 1991 across grades in 2001 by month of birth. The high share of those still in the first or second grade among Roma women confirms the expected high prevalence of late/very late starters and grade repetitions on both sides of the cutoff. What is clear though is that Roma women born below the cutoff are more likely to be in the fourth grade and less likely to be in the third grade than Roma women born above the cutoff, while the difference around the cutoff in the probability of being in the first or second grade is small (and if tested, insignificant), showing that the share of grade repeaters and very late starters is about the same on both sides of the cutoff. The fact that those born below the cutoff are more likely to be in the fourth grade and thus more likely to have started school before the reform, than those born above the cutoff, shows that there is indeed a jump in the probability of being exposed to the reform around the cutoff in 1991.

---

<sup>7</sup>In the aggregate administrative data of the Public Education Statistics of the Public Education Information System database, the share of grade repeaters in the first grades of elementary school is low, about 1.4-4.0% (Table 12 in the Appendix), and stayed stable in this period (National Institute of Public Education, 2006). Such aggregate administrative data are however not available by ethnicity. The only data on grade repetition among Roma students come from the Hungarian Life Course Survey (HLCS). The HLCS is a cohort study that follows 10,000 young people who were in Grade 8 in the spring of 2006 (Hajdú, 2015). As Table 13 in Appendix A demonstrates, 5% of non-Roma and 20% of Roma students repeated grade at least once.

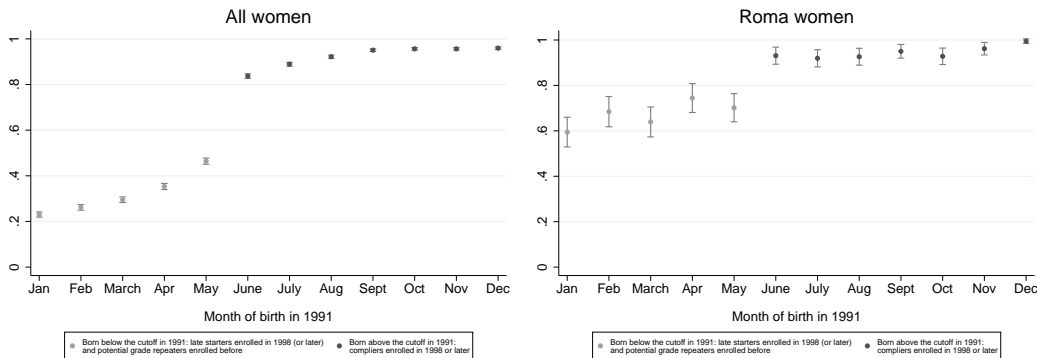
Figure 3: The distribution of students across school grades in the Spring of 2001, women born in 1991



Data Source: own estimation from the 2001 Hungarian Census. No. of observations: 60,302 and 2,313

We use the probability of being at most in the third grade in the Spring of 2001 to proxy the probability of starting school after the reform<sup>8</sup> (Figure 4). Again, those in the fourth grade surely started school before the reform while those in at most the third grade either started school after the reform or they started school before the reform and repeated grades. Thus, in the worst case scenario, we underestimate the size of the jump in the probability of being exposed to the reform. This is not a problem for our purpose, however: all we need to estimate the ITT effects of the reform is to demonstrate that there is a significant jump in the probability of exposure around the cutoff and we do not use the size of the jump. Among all women born in 1991, the probability of being at most in the third grade jumps from 47% to 85%, while among Roma women, it jumps from 70% to 94% around the cutoff (see Figure 4).

Figure 4: The probability of being at most in the third grade in the Spring of 2001



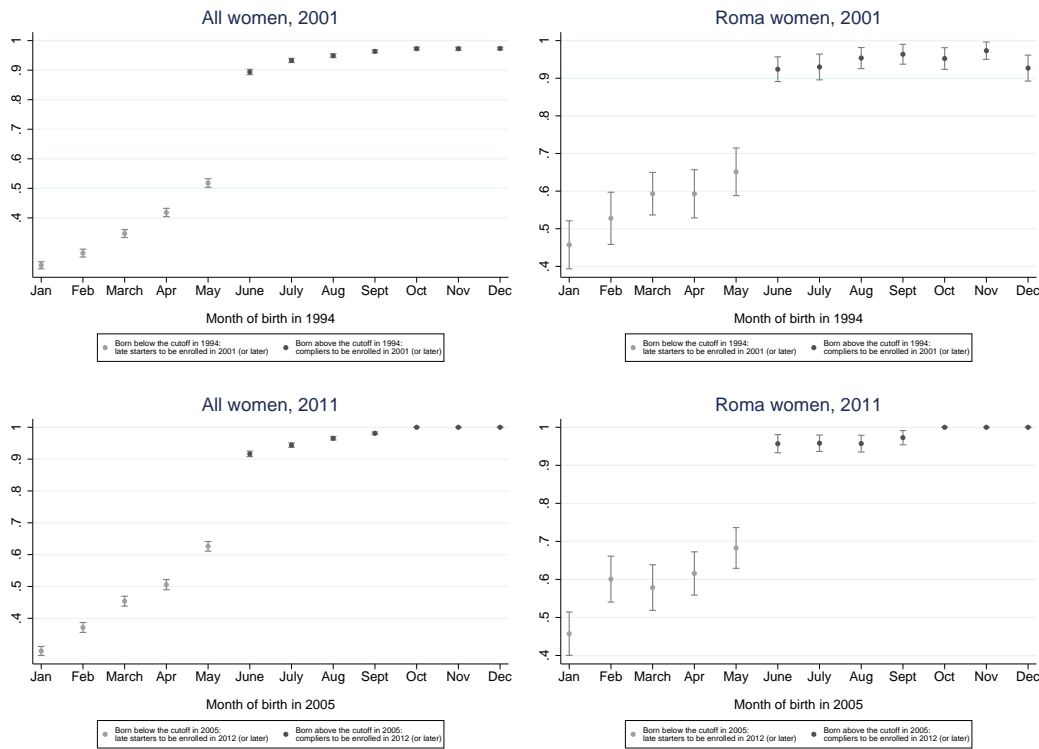
The average probability of starting school in 1998, plotted with the 95% confidence intervals of the means. Data Source: own estimation from the 2001 Hungarian Census. 0 on the x-axis refers to being born in June 1991. No of individual observations: 60,302 and 2,313.

To have a rough idea about how large the jump would be if we did not have to deal with the problem of

<sup>8</sup>Note that we could also choose the opposite approach and plot the probability of being at least in the fourth grade in 2001 as those in the fourth grade surely started school before the reform, even if they repeated grades. In that case, we would see the same drop in the probability of starting school before the reform around the cutoff. As  $p(\text{being at most in the third grade}) = 1 - (\text{being at least in the fourth grade})$ , it does not matter which approach we use. We find that plotting the probability of being treated is a more intuitive approach.

unobserved grade repetitions, we can look at children who are not in school yet at age 6 in the Censuses. If they were born below the cutoff, they are going to be late starters as they should have already been in school; if they were born above the cutoff, they are compliers as they should not be in school yet<sup>9</sup>. The relevant cohorts for this comparison consist of those born in 1994 in the case of the 2001 Hungarian Census, and those born in 2005 in the case of the 2011 Hungarian Census<sup>10</sup>. Figure 5 shows that in the two Censuses, among Roma women born in June-May, 46-68% , while among those born in June-December, 92-100% of pupils are not in school yet at age 6. The jump in the probability of not being in school yet at age 6 around the cutoff is about 27-28 percentage points among Roma women, which is only slightly higher than to the jump we have found above for those born in 1991 (24 pp). This result is in line with our finding that there is no jump in the probability of being in a lower class (first or second) at age 9/10 among those born in 1991 around the cutoff; thus, the probability of starting school at age 8 and repeating grades are similar on the two sides of the cutoff.

Figure 5: Compliance to the enrollment rule: the probability of not being in school yet at age 6 in 2001 and 2011, women born in 1994 and 2005



The average probability of not being in school yet in the Spring of 2001 and in the Autumn of 2011, plotted with the 95% confidence intervals of the means. Data Source: own estimation from the 2001 and 2011 Hungarian Census. 0 on the x-axis refers to being born in June 1991. No. of individual observations: All women: TBA; Roma women: 2,635 and 3,407.

<sup>9</sup>They also could be late starters if they delayed school enrollment to age 8. However, for our purposes, for those born above the cutoff being a complier or a late starter makes no difference; both would be considered as compliers to the reform.

<sup>10</sup>The difference between 2001-1994=7 and 2011-2005=6 comes from the fact that the 2001 Census was taken in the spring while the 2011 Census was taken in the autumn.

Table 3: Compliers, never- and always takers according to the Rubin-framework

	Children born below the cutoff, $Z=0$	Children born above the cutoff, $Z=1$
Children started school under the old scheme in 1997, $D=0$	compliers ( $D(0) = 0$ )	never-takers (early starters, $D(1) = 0$ )
Children started school under the new scheme in 1998, $D=1$	always-takers (late starters, $D(0) = 1$ )	compliers ( $D(1) = 1$ )

Adapted from [Adamecz-Völgyi \(2018\)](#).

We do not have information on how grade repetition might have affected the enforcement of CSL age. There should be no problem with those who started school in 1998: even if they repeated a grade later, they entered a new class where everybody else was exposed to the higher CSL age, too. The question is what happened to those who started in 1997, before the reform, repeated a grade and moved to a class where everybody else was supposed to stay in school until age 18. Practically, those students who were aware of the fact that they could have dropped out at age 16, should have had the possibility to do so if they wanted to. On the other hand, some students who just went with the flow might have stayed in school longer, together with their peers in the repeating class. As some women born below the cutoff might have shifted to the intention-to-treat group due to grade repetitions, we might underestimate the real effect around the cutoff.

## 4 Identification strategy and empirical methods

### 4.1 Identification strategy<sup>11</sup>

As detailed in Section 3.2, our identification strategy is based on compliance with the school enrollment rule, which creates a discontinuity in the probability of starting elementary school after the reform at the date of birth of June 1, 1991. This discontinuity allows us to identify the ITT effects of the reform using an RDD strategy, with being born on June 1, 1991, as the cutoff.

As our data does not allow us to estimate the exact size of the jump in the probability of being treated around the cutoff for Roma women, this analysis will only estimate the ITT effects of the reform. This is a fuzzy RDD setup where compliance with the enrollment rule is endogenous. Being born right after, rather than before, the cutoff is used as an instrument for starting school after the reform. We identify the ITT effects according to the potential outcome framework of [Rubin \(2005\)](#). We define a binary instrumental variable  $Z$  that captures whether individual  $i$  was born above or below the cutoff as  $Z = 1(\text{born on or after June 1, 1991})$ , and the actual treatment indicator  $D$  that captures whether individual  $i$  started school under the new legislation as  $D = 1(\text{started school under the new scheme in 1998})$ . The potential treatment indicators  $D(0)$  and  $D(1)$  capture whether individual  $i$  entered school under the new CSL age legislation scheme, conditional on  $Z$ . Using this notation, the actual treatment indicator can be expressed as  $D = Z * D(1) + (1 - Z) * D(0)$ . Table 3 summarizes the compliers, never- and always-takers according to the Rubin potential outcome framework. The size of the jump in the probability of starting school under the new CSL age scheme is  $P(D(1) = 1) - P(D(0) = 1)$ , which is estimated to be 0.24 among Roma women (Figure 4).

Our identification strategy is based on five assumptions. First, we assume that the instrument is exogenous: whether someone was born on the left or the right side of the cutoff is random. As the reform was introduced five years after the birth of the relevant cohort, it could not have led to a manipulation of birth. The question whether children born in distinct months of a year might have systematically different

<sup>11</sup>Our identification strategy draws heavily on [Adamecz-Völgyi \(2018\)](#).



outcomes might be still valid (see [Fan et al. \(2014\)](#) and [Buckles and Hungerman \(2013\)](#)). In general, the literature is concerned with whether children born in the winter months are inherently different from those born in the spring. There seems to be some evidence that less educated women are more likely to give birth in the winter, when the environment to a newborn baby is also less favorable than in the spring. Our identification strategy compares children born before and after June 1. Even though the question of whether those born in May vs. in June are intrinsically different has not been discussed in the literature yet, Section 6.3 will relax the exogeneity assumption of the date of birth as a robustness check.

The second is the exclusion restriction: being born right before or after the cutoff affects fertility decisions exclusively through the reform. This assumption is far from being trivial. Independently from the reform, those born in May, if compliant with the enrollment rule, are supposed to start school a year earlier than those born in June ([Angrist and Krueger, 1991](#); [Altwicker-Hámori and Köllő, 2012](#)). Thus, it is an empirical question whether fertility decisions of those born in May are different from those born in June in years when no reform occurred. Furthermore, the fact that those born in May are likely to start school a year earlier than those born in June reduces the “net” impact of the reform on the compliers to one year. We conduct two robustness checks to see whether the exclusion restriction is a reasonable assumption in this case. First, we do not find significant differences in the probability of teenage motherhood around the same cutoff at June 1 in the year before and after the reform; we hypothesize that the potential effects of starting school at a younger age and spending more time in school probably balance each other out (see Section 6.2). Second, we directly control for any potential impacts around the same cutoff in other years in Section 6.3.

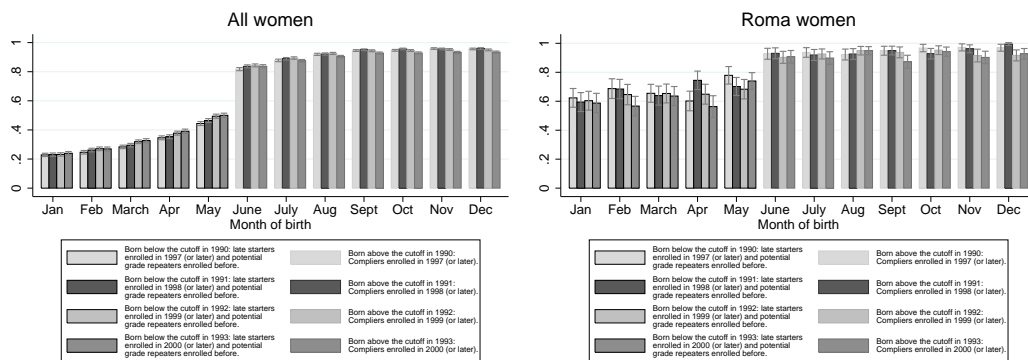
The third assumption is that our instrumental variable is continuous and no defiers exist. We assume that the reform

- did not induce anyone who was born above the cutoff to purposefully start school a year earlier to prevent spending two additional years in education, and
- did not induce anyone who was born below the cutoff to purposefully start school a year later in order to be exposed to the reform.

We assume that the timing of school start was determined by the child’s development and their parents general preferences about the best age of starting school in all cases, i.e., both for those born below the cutoff who started school a year later (late starters or always-takers), and for those born above the cutoff who started school a year earlier (early starters or never-takers). This seems plausible considering that dislike for school typically stems from lack of information on the benefits of education and/or a very steep discount rate of future incomes, which is likely correlated with lack of information on government reforms and a short planning horizon. However, due to lack of precise information on school enrollment, we cannot fully exclude the possibility of some degree of manipulation. If there were Roma parents who manipulated school enrollment to save their children from two extra years in school, some women in the intention-to-treat group belonged to the control group and led us to underestimate the effect of the reform. Similarly, if there were parents who manipulated school enrollment to seek exposure to the reform, we might overestimate the effect of the reform.

According to aggregate administrative data, the share of late starters and early starters was stable on average in this period both both across cohorts measured by date of birth, and measured by academic years (see Table 14 in Appendix A). Again, such data are not available specifically for the Roma and we must rely on the 2001 Hungarian Census. Figure 6 shows compliance to the enrollment rule across four cohorts born in 1990-1993 around the June 1 cutoff in the same fashion as Figure 4 did for those born in 1991. The data do not provide a clear picture though. The share of compliers on the right side of the cutoff is fairly stable across these years so there are no signs of parents being more likely to send their children born in 1991 to

Figure 6: The probability of starting school at age 7 (or later) in 1997-2000



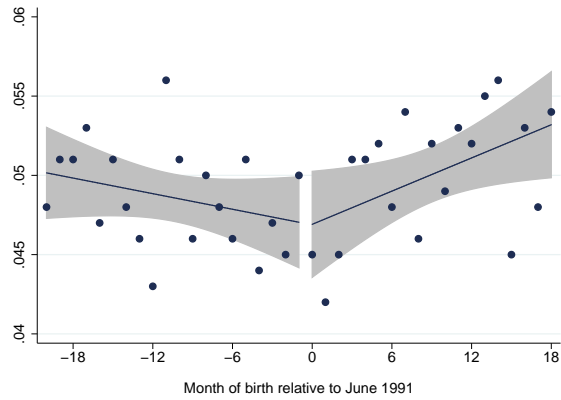
The average probability of starting school at age 7 or later in 1997-2000 by month of birth, plotted with the 95% confidence intervals of the means. Data Source: own estimation from the 2001 Hungarian Census. 0 on the x-axis refers to being born in June 1991. No of individual observations: All women: 61,453, 61,068, 58,455, 56,363; Roma women: 2,308, 2,313, 2,282, and 2,386.

school earlier in order to avoid the extra two years of schooling. On the left side of the cutoff, there are some months with significant differences. April seems to be the most problematic, when the probability of starting school late (and/or being a grade repeater) is significantly higher in 1991 than in 1990 and in 1992, which might mean that some parents wanted to send their kids to school late explicitly in order to make sure that they are forced to stay in school two years longer. However, April is the only month when we observe this pattern, and, the difference in April between 1993 and 1994 is again significant although nothing happened between these two years, suggesting the all this might just be noise in the data. As the share of late starters is fairly stable in Jan-March, we run a robustness check in Subsection 6.4 where we exclude individuals born right around the cutoff, in May-July, from the estimation sample (donut hole RDD), and show that even if there were defiers, they do not drive the results.

The fourth identification assumption is that the fact whether one reports herself as Roma or not is independent of the reform. As detailed in Section 2.2, ethnicity is self-reported, and roughly every second Roma person does not report himself or herself as Roma in the 2011 Hungarian Census. Theoretically, the effects of longer schooling on human capital accumulation could impact self-assessment in two ways. If human capital development induces individuals to be more conscious about their identity, those exposed to the reform could be more likely to reveal their Roma ethnicity, meaning we overestimate the effect of the reform. However, this works in an opposite fashion if human capital development induces people to hide their minority status. There is some evidence for the latter in relation to the black population in the US (Fryer and Torelli, 2010). If this happens, our results would have a downward bias. To test the validity of the assumption that whether one reports herself as Roma or not is independent of the reform we compare the share of women identifying themselves as Roma below and above the cutoff. Figure 7 shows that we find no significant discontinuity in the share of Roma women at the cutoff. The same is true for the number of Roma women around the cutoff (see Figure 8).

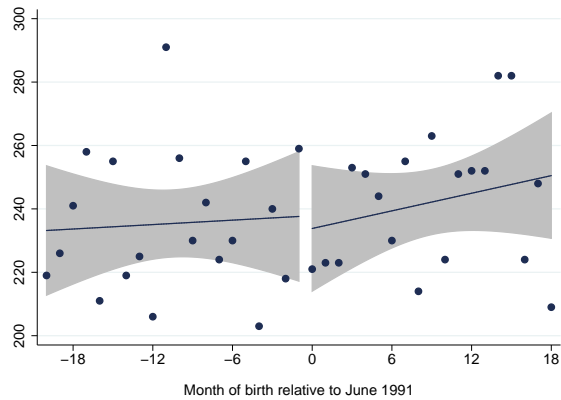
Fifth, in addition to Roma self-identification being independent of the reform, we also assume its independence from teenage motherhood. In addition to the reform, ethnic identification may be affected by individual circumstances as well. Kézdi and Simonovits (2016) examine the causal effects of economic hardship on Roma identification of adolescents in Hungary. They build on a theory that “individuals are more likely to categorize themselves as members of a group if they perceive themselves to be more similar to the other members of that group”. They show that Roma adolescents are more likely to identify themselves as

Figure 7: The share of Roma women by month of birth, 1991



Data source: own estimation from the 2011 Hungarian Census. 0 on the x axis refers to being born in June 1991. Linear regression lines estimated separately below and above the cutoff, plotted with 95% confidence intervals. No. of individual observations: 263,298.

Figure 8: The number of Roma women by month of birth, 1991



Data source: own estimation from the 2011 Hungarian Census. 0 on the x axis refers to being born in June 1991. Linear regression lines estimated separately below and above the cutoff, plotted with 95% confidence intervals. No. of individual observations: 16,667.

Roma if their family experienced economic hardship since being Roma is associated with poverty. It would be problematic if a similar mechanism occurred with respect to teenage childbearing, and that teenage mothers were more likely to identify as being Roma. Kézdi and Simonovits (2016) do not examine the causal effect of early childbearing on Roma identification. However, in their regressions of Roma identification on economic hardship, they do control for having a child. They do not find a robust significant correlation between having a child in teenage years and identifying as Roma.

We cannot test the causal relationship between teenage motherhood and Roma identification using our data. However, this paper concludes that increasing the CSL age decreased the probability of teenage pregnancy and motherhood among Roma women. If becoming a teenage mother made women more likely to identify as Roma, and increased CSL age decreases the probability of teenage motherhood among *some* women, it is possible that we underestimate the effect that the reform had on Roma women.

## 4.2 Empirical methods

Similarly to Adamecz-Völgyi (2018), we estimate the ITT effects of the reform around the cutoff using both a nonparametric and a parametric estimation strategy. Nonparametric estimates are generated by estimating weighted local linear regressions on both sides of the cutoff, within a certain bandwidth (Hahn et al., 2001; Imbens and Lemieux, 2008). For simplicity, weights are computed by applying a rectangular kernel function to the distance from each observation to the cutoff in terms of day of birth. This is the standard method of RDD estimation as it has excellent properties in estimating the difference of two conditional expectations evaluated at the boundary points of the cut-off (Cheng et al., 1997).

The following local linear models are estimated within a certain bandwidth:

$$y_i = \alpha_{NP} + \beta_{NP} * itt_i + \gamma_{NP} * x_i + \delta_{NP} * x_i * itt_i + \varepsilon_i$$

where

$y_i$  is the outcome variable;

$itt_i$  is the intention-to-treat variable, which is 1 if individual  $i$  was born on June 1, 1991, or later, and 0 otherwise;

$x_i$  is the running variable, number of days in date of birth before or after June 1, 1991 (and 0 if individual  $i$  was born on June 1, 1991);

$x_i * itt_i$  is an interaction term of  $x_i$  and  $itt_i$ , allowing for the local linear function to be different on the two sides of the cutoff; and

$\beta_{NP}$  captures the ITT effect of the reform estimated using our nonparametric approach.

We follow a conservative strategy and use the strictest procedure to set the bandwidth of the local linear regressions: the optimal bandwidth routine of Calonico et al. (2014a), abbreviated as CCT for the remainder of the paper, along with the 50-150% versions of the optimal bandwidths as robustness checks. The optimal bandwidths set by the method are 150-200 days wide below and above the cutoff, depending on the outcome variable and the sample.

A parametric approach is used as one of the robustness checks on the sample of individuals born in 1980-1993 using 4th order global polynomial models. The estimated parametric models are the followings:

$$y_i = \alpha_p + \beta_p * itt_i + f(x_i, itt_i) + u_i$$

where

$f(x_i, itt_i)$  is a 4th-order polynomial function of the running variable, which is different on the two sides of the cutoff; and  $\beta_P$  captures the ITT effect of the reform estimated using our parametric approach.

There are two reasons for complementing the nonparametric analysis with parametric models. First, they can accommodate additional control variables. In particular, birth month fixed effects are used to capture the impacts that any potential monthly seasonality has on the characteristics of the child, and thus relaxes the assumption of exogenous variation in month of birth, and birth year fixed effects are used to capture the potential effects of business cycles. An interesting feature of the Census data (and the Hungarian health system) is that the day of the week matters with respect to the probability of being born. That is, a child is more likely to be born Tuesday through Friday than Saturday through Monday, and this probability difference is weakly related to the educational status of the mother<sup>12</sup>. This paper will not document this phenomenon. However, because June 1 in 1991 fell on a Saturday, day of the week fixed effects are also included in the parametric models to control for this pattern.

Second, we use two databases that are linked together. Linking observations related to one woman between the two data sets is not random. However, the systematic relationship between the linking procedure and being born after the cutoff disappears after controlling for additional individual characteristics (see Section 2.1). We do not control for these characteristics in our nonparametric RDD strategy, only in our parametric approach. Finding the same results in both approaches supports that any potential bias coming from non-random linking must be small.

### 4.3 The number of completed school years around the cutoff

We know from the qualitative studies of Mártonfi (2011a,b) that most Roma (and non-Roma) students were compliant to the legislation and showed up in school at older ages as required. As detailed in Section 2, the Hungarian Census does not allow us to directly observe how long students stayed in school, we only observe the number of successfully completed school grades. The data do not capture unfinished, incomplete or repeated years in school. Figure 9 shows the share of women staying in school after completing 5, 6, ..., 12 grades in school. Non-Roma women born right above the cutoff have completed more grades by 2011 than non-Roma women born below the cutoff, especially at and above Grade 10. For Roma women, the improvement seems to be more widespread due to grade repetitions. On average, Roma women born above the cutoff are 3-4 percentage point more likely to stay in school after completing grades 8-11 and 2-3 percentage points more likely to stay in school after completing grade 12-14 than Roma women born below the cutoff.

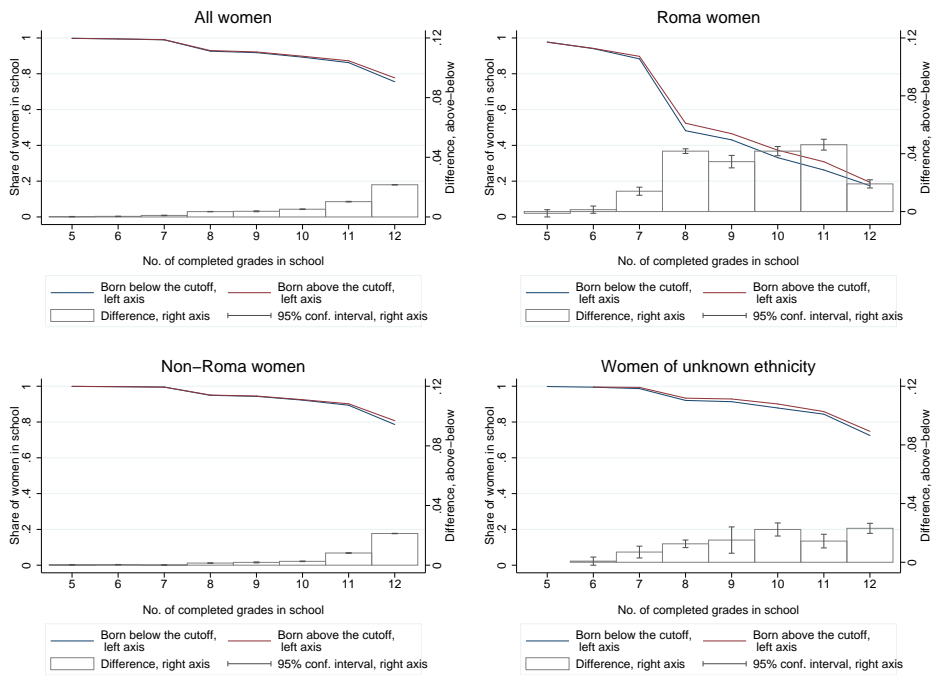
### 4.4 Teenage fertility around the cutoff

Figure 10 shows a similar picture by comparing the share of childless women by age among those born right below and above the cutoff. There is no difference between the share of childless women around the cutoff among non-Roma women. However, such differences are observed among the group of Roma women. The gap in their entry into motherhood begins after reaching age 17, it peaks between ages 18-19, and starts closing by age 20.

---

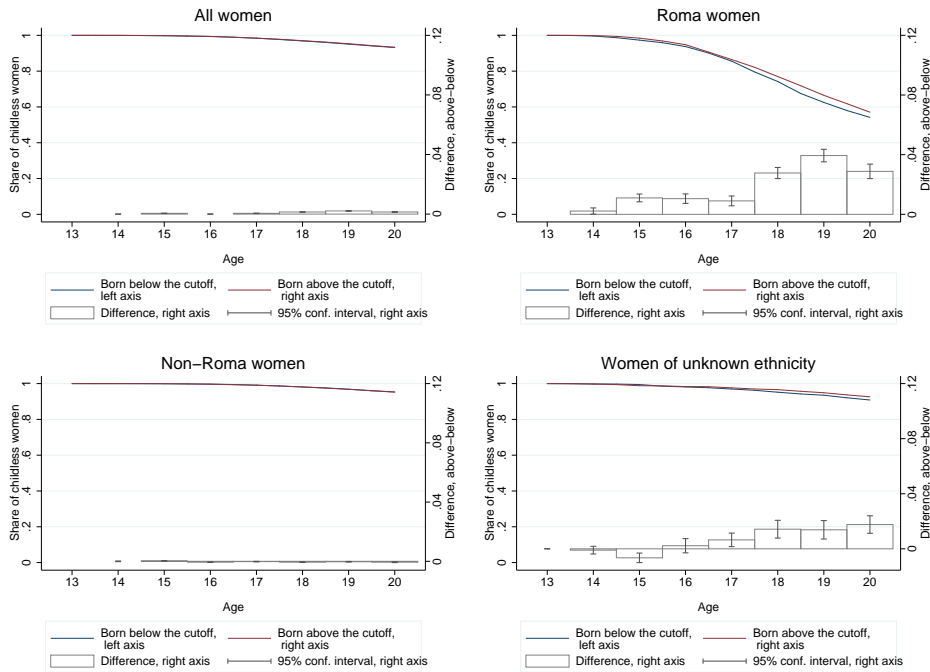
<sup>12</sup>Own estimation from the 2011 Census.

Figure 9: The share of women staying in school after completing 5, 6, ..., 12 grades



Kaplan-Meier survival functions with respect to the No. of successfully completed years in school. Data source: own estimation from the 2011 Hungarian Census. Born before the cutoff (control group): born at most 180 days before June 1, 1991. Born after the cutoff (intention-to-treat group): born at most 180 days after June 1, 1991. Number of observations: 56,628, 2,775, 53,634 and 2,219, respectively.

Figure 10: The share of childless women by age



Kaplan-Meier survival functions with respect to giving birth, by age. Age is a continuous variable; giving birth by age 16, for example, means giving birth either before, or exactly on, the mother's 16th birthday. Data source: own estimation from the 2011 Hungarian Census. Born before the cutoff (control group): born at most 180 days before June 1, 1991. Born after the cutoff (intention-to-treat group): born at most 180 days after June 1, 1991. Number of observations: 56,628, 2,775, 53,634 and 2,219, respectively.



Table 4: Effects on the probability of motherhood

Probability of having the first child by age ...	All women	Roma women	Non-Roma women	Ethnicity Unknown
16	0.001 (0.002) [50,005] {152.2}	0.009 (0.020) [2,298] {148.2}	0.001 (0.001) [51,884] {173.6}	-0.003 (0.011) [2,486] {201.3}
17	-0.002 (0.001) [60,137] {184.3}	-0.027 (0.020) [3,000] {194.8}	0.000 (0.002) [53,321] {178.4}	-0.007 (0.011) [2,510] {203.0}
18	-0.004 (0.002) [65,502] {201.2}	-0.068*** (0.022) [2,906] {188.3}	0.000 (0.003) [56,700] {190.8}	-0.02** (0.008) [2,545] {206.7}
19	-0.008 (0.004) [61,070] {187.9}	-0.067** (0.023) [3,015] {195.7}	-0.001 (0.004) [61,000] {205.3}	-0.01 (0.013) [2,100] {168.6}
20	-0.003 (0.004) [51,619] {157.2}	-0.018 (0.030) [2,633] {169.8}	-0.001 (0.004) [44,611] {148.0}	-0.044** (0.017) [2,438] {197.7}

Linear probability models estimated by local linear regressions with rectangular kernel bandwidths set using the bandwidth optimization routine of (Calonico et al., 2014a). Age is a continuous variable; giving birth by age 16, for example, means giving birth either before, or exactly on, the mother's 16th birthday. Each reported coefficient comes from a separate regression. The coefficients on the treatment dummy variable are reported in the table. Control variables: linear functions of the running variable separately below and above the cutoff. Robust standard errors clustered by birth year and month are in parentheses, observation numbers in brackets, bandwidth measured in days in braces. \*\*\* p<0.01, \*\*p<0.05, \*p<0.1.

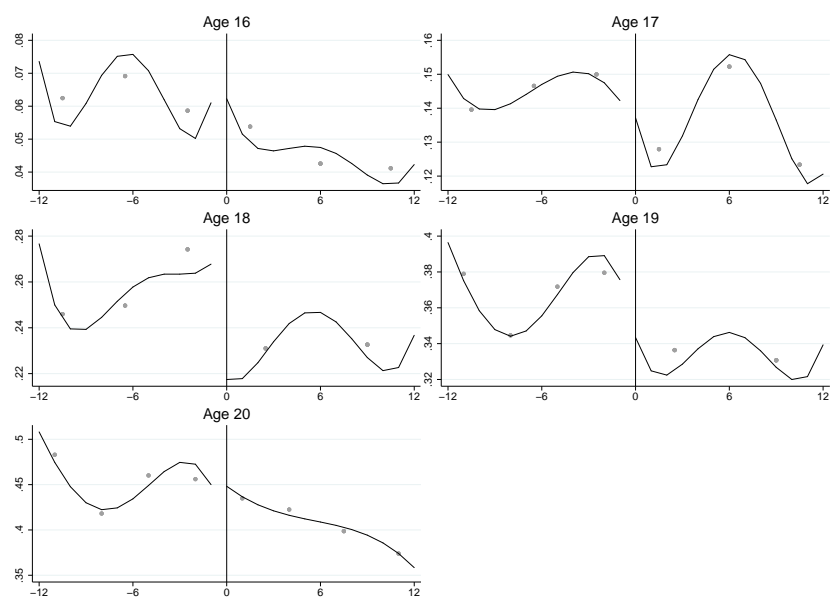
## 5 Estimation results

### 5.1 The effects of the CSL age increase on teenage motherhood

Table 4 (and Figure 11) presents the ITT effects of the reform on the probability of having the first child by ages 16-20. The reform has a significant negative effect on the probability of first birth-giving by age 18 among Roma women. Considering that the share of Roma women who give birth by age 18 is at 26.0% in the control group (see Table 1), the estimated 6.8 percentage point effect is quite large,  $6.8/26=26\%$ . The effect is temporary and vanishes by age 20.

Our results coincide with the phenomenon also seen in Figure 10. The gap in motherhood in the treated and control groups of Roma women starts to open after the CSL age of 16, and fades out after age 19, when the CSL age is non-binding anymore. This pattern suggests that the reform delayed motherhood by 2 years among Roma adolescents.

Figure 11: The probability of motherhood by age 16-20 around the cutoff, Roma Women



Regression-discontinuity (RD) plots on monthly data using the rdplot command of Stata (Calonico et al., 2014b), Roma women only. Age is a continuous variable; giving birth by age 16, for example, means giving birth either before, or exactly on, the mother's 16th birthday. Local linear regressions fit separately below and above the cutoff on birth year and month averages, using the rule-of-thumb bandwidth of lpoly in Stata with 90% confidence intervals. Month of birth 0 represents June 1991. No. of individual observations: 4,780.

## 5.2 The mechanics of the incapacitation effect

The effects materialise during the period directly effected by the reform, i.e. below or just above age 18, which suggests that at least some of the impact among Roma women is due to the incapacitation effect of education. Our data allows us to examine the conception pattern of these pregnancies so we try to shed light on how such an incapacitation effect is working in practice.

As detailed in Section 2, we have two methods for identifying the conception time of pregnancies: at monthly precision from the 2011 Hungarian Census data, and at weekly precision on a subsample of the 2011 Hungarian Census data, linked individually to the Vital Statistics database (linked data). Both methods support that the reform decreased the probability of getting pregnant during school year only, when adolescents had to be physically in school, and it had no significant effect on the probability of getting pregnant during summer breaks (Table 5).

Results estimated on the linked data show that the probability of getting pregnant decreased by 6.5 percentage points (Table 5, Block B). We do not see evidence for increased demand for abortions (Block C). In fact, along with less pregnancies, the probability of having an abortion decreased by 2.9 percentage points.

Furthermore, the linked data suggest that the probability of “unwanted” pregnancies (i.e. those ending with abortion) decreased a little more than the probability of “wanted” (i.e. kept through full term) pregnancies. Although the difference is not significant between the two coefficients, the probability of getting pregnant decreased a bit more than the probability of giving birth. Also, the probability of choosing an abortion to end a pregnancy declined by 8.9 percentage points (Block C). While this estimate is large, it is not significant; potentially, because it is estimated on the small sample of Roma women who did get pregnant by age 18 ( $n=442$ ).

Lastly, the linked data offers a possibility to pin the academic year further down to examine the probability of getting pregnant during Christmas breaks separately (Block G). Similarly to summer breaks (Block F), we see small and insignificant coefficients (decrease by 0.8 percentage points). Finding no change in the probability of getting pregnant during school breaks supports the hypothesis that the effect is generated mainly by the incapacitation channel.

## 6 Robustness checks

In this section we provide robustness checks to support our earlier results.

### 6.1 Alternative bandwidth choices

In this subsection we show that our results are not sensitive to bandwidth choice. We do the same local linear estimation as before, using the 50-150% versions of the optimal bandwidth set by the CCT routine. As Table 6 shows, our results do not depend on bandwidth choice in magnitude.

### 6.2 Effects around the same cutoffs in 1990 and 1992

One might be worried that we simply measure the effects of starting elementary school at different ages (a little after age 6 in the control group vs. age 7 in the treated group) instead of measuring the effects of the CSL age increase. To show that this is unlikely to be the case, we replicate our results around the same cutoffs in the year before and after the reform. Table 7 presents our results. The real, 1991 cutoff is the only one producing significant impacts.

Table 5: The mechanics of the incapacitation effect (Roma women)

Outcome	Census data	Linked data
A. Probability of giving birth by age 18	-0.066** (0.022) [2,788]	-0.056*** (0.014) [2,327]
B. Probability of getting pregnant	-0.066** (0.019) [2,788]	-0.065** (0.021) [2,327]
Time of conceptions defined: Census data: assuming 9-month pregnancies Linked data: based on information on the week of conception		
C. Probability of abortion	-	-0.029* (0.016) [2,327]
D. Probability of ending pregnancy with abortion	-	-0.089 (0.077) [442]
E. Probability of getting pregnant during school years	-0.054** (0.024) [2,788]	-0.054** (0.023) [2,327]
School years defined: Census data: Sept - May Linked data: 1 Sept - 14 June, excluding 15 Dec - 31 Dec		
F. Probability of getting pregnant during summer breaks	-0.011 (0.015) [2,788]	-0.003 (0.009) [2,327]
Summer breaks defined: Census data: June - Aug Linked data: 15 June - 31 Aug		
G. Probability of getting pregnant during Christmas breaks	-	-0.008 (0.007) [2,327]
Christmas breaks defined: Linked data: 15 Dec - 31 Dec		

Linear probability models estimated by local linear regressions with rectangular kernel using 180-day bandwidths. Each reported coefficient comes from a separate regression. The coefficients on the intention-to-treat variable are reported in the table. ITT group: born on June 1, 1991, or within 180 days after. Control group: born at most 180 days before June 1, 1991. Control variables: linear functions of the running variable separately below and above the cutoff. Age is a continuous variable; giving birth by age 18 means giving birth either before, or exactly on, the mother's 18th birthday. Robust standard errors clustered by birth year and month are in parentheses, observation numbers in brackets. \*\*\* p<0.01, \*\*p<0.05, \*p<0.1.  
#If summer breaks were defined as July-Aug on the Census data, the respective ITT coefficient would be 0.003 (se=0.006).

Table 6: The effects of the reform using alternative bandwidths, Roma women

Version of the CCT bandwidth	Probability of					
	giving birth by age 18 (Census data)	giving birth by age 18 (Linked data)	getting pregnant by age 18 (Linked data)	getting pregnant during school years (Linked data)	getting pregnant during summer breaks (Linked data)	getting pregnant during Christmas breaks (Linked data)
50%	-0.042 (0.029)	-0.023 (0.021)	-0.051*** (0.016)	-0.001 (0.017)	-0.008 (0.008)	-0.013* (0.006)
75%	-0.051* (0.025)	-0.048*** (0.009)	-0.041* (0.017)	-0.026 (0.015)	0.006 (0.009)	-0.012 (0.007)
100%	-0.068*** (0.022) [2,906] {188.3}	-0.051*** (0.012) [2,855] {221.8}	-0.054** (0.018) [2,580] {191.8}	-0.043** (0.017) [2,664] {205.9}	0.010 (0.009) [4,325] {329.2}	-0.007 (0.006) [2,933] {226.4}
125%	-0.048** (0.022)	-0.041** (0.014)	-0.062*** (0.018)	-0.051** (0.019)	0.009 (0.009)	-0.008 (0.006)
150%	-0.050** (0.022)	-0.034** (0.016)	-0.045** (0.020)	-0.038* (0.020)	0.008 (0.008)	-0.007 (0.005)

Linear probability models estimated by local linear regressions with rectangular kernel bandwidths set using the bandwidth optimization routine of (Calonico et al., 2014b). Each reported coefficient comes from a separate regression. Age is a continuous variable; giving birth by age 18 means giving birth either before, or exactly on, the mother's 18th birthday. Robust standard errors clustered by birth year and month are in parentheses, observation numbers in brackets, bandwidth in days in braces. \*\*\* p<0.01, \*\*p<0.05, \*p<0.1.

Table 7: The effects around cutoffs in 1990-1992 (Roma women)

Cutoff year	Probability of ...					
	giving birth by age 18 (Census data)	giving birth by age 18 (Linked data)	getting pregnant by age 18 (Linked data)	getting pregnant during school years (Linked data)	getting pregnant during summer breaks (Linked data)	getting pregnant during Christmas breaks (Linked data)
1990	0.037 (0.029)	-0.012 (0.013)	0.005 (0.018)	-0.007 (0.022)	0.007 (0.011)	0.005 (0.003)
1991	-0.066** (0.022)	-0.056*** (0.003)	-0.065** (0.011)	-0.054** (0.039)	-0.003 (0.711)	-0.008 (0.278)
1992	0.056 (0.040)	0.033 (0.033)	-0.005 (0.030)	0.009 (0.023)	-0.020 (0.018)	0.005 (0.004)

Linear probability models estimated by local linear regressions with rectangular kernel using 180-day bandwidths. Each reported coefficient comes from a separate regression. The coefficients on the intention-to-treat variable are reported in the table. ITT group: born on June 1, 1991, or within 180 days after. Control group: born at most 180 days before June 1, 1991. Control variables: linear functions of the running variable separately below and above the cutoff. Age is a continuous variable; giving birth by age 18 means giving birth either before, or exactly on, the mother's 18th birthday. Robust standard errors clustered by birth year and month are in parentheses. \*\*\* p<0.01, \*\*p<0.05, \*p<0.1. Number of observations are around 2,000 in each case.

It might be surprising that we find no effect around the same cutoff in other years. Two things happen around cutoffs in other years: (1) those born in June start school at an older age, and (2) those born in June spend one year less in school before reaching CSL age. Theoretically, starting school at an older age affects teenage fertility in both directions. [Black et al. \(2011\)](#) argue that starting school when older may be beneficial for human capital development because older children are at a more advanced stage of their developmental life. In addition, social development may depend on a child's age relative to the class. If being older than one's peers was beneficial, starting school when older would be beneficial; however, it is not clear whether this is really the case. On the other hand, starting school when older is harmful if children learn more in school than in pre-school (or at home). Furthermore, parental investment in raising their children may depend on school starting age as well. [Black et al. \(2011\)](#) find that starting school at a slightly older age decreases teenage fertility in Norway; however, this effect is not robust across all specifications they use. In the Norwegian school system, schooling is compulsory for 9 years and not until a given age. Thus, contrary to the Hungarian case, those starting school later do not spend fewer years in school. In our case, it is likely that the effects of starting school when older and spending fewer years in school before reaching CSL age cancel each other out.

### 6.3 Controlling for monthly and yearly seasonality (parametric approach)

Although we find no significant effects around the cutoffs in other years, the estimated coefficients are not significant zeros. One may also worry whether the month of birth is really exogenous (see the dispute about this by [Buckles and Hungerman \(2013\)](#) vs. [Fan, Liu, and Chen \(2014\)](#)). Furthermore, as discussed in Section 2.1 and Section 4.1, we would like to provide an additional robustness check to our data-linking procedure as well. Therefore, we estimate flexible global polynomial models using a sample of Roma women born in 1980-1993, controlling for these individual characteristics.<sup>13</sup>

Our global polynomial models include the same intention-to-treat variable as in Subsection 4.2, *itt*, represented as a 1 if individual *i* was born on June 1, 1991, or after, and as a 0 otherwise, along with a 4th order function of the running variable separately below and above the cutoff, and year of birth, month of birth, day of the week of birth, settlement type, and county fixed effects. If there is a change in the prevalence of teenage pregnancies at any particular year or month of birth, we would be able to control for it with this specification, and would estimate the effect that is still remaining. As Table 8 shows that the effect of the reform in these models is significant and very similar in magnitude to the ones estimated by non-parametric regressions.

### 6.4 Donut hole RDD

Although, as we have argued, it is unlikely that Roma parents had been gaming the school enrollment system by sending their kids to school earlier/later in order to purposefully avoid/seek exposure to longer schooling, we cannot provide convincing empirical evidence against it due to the lack of administrative schooling data by ethnicity. Thus, we apply an empirical strategy that excludes children born right around the cutoff, where gaming the system might have been the easiest. This strategy is referred as a donut hole RDD in the literature [Kirdar et al. \(2018\)](#). We estimate a series of RDD's on the probability of motherhood by age 18 among Roma women, sequentially excluding those born within  $i=0-60$  day below and above the cutoff from the estimation

<sup>13</sup>The choice of the period 1980-1993 is arbitrary. We have tried out several other period lengths and they all give similar results. The only important consideration here is that those born in 1993 are already 18 years old at the time of the Census in October 2011; thus, we can not go further ahead in time. To the other direction, below 1991, we could go as many years as we want. The main consideration was to include a longer period to make room to the estimation of month of birth fixed effects.

Table 8: Effects after controlling for seasonality, Roma women

	Probability of ...					
	giving birth by age 18	giving birth by age 18	getting pregnant by age 18	getting pregnant during school years	getting pregnant during summer breaks	getting pregnant during Christmas breaks
	(Census data)	(Linked data)	(Linked data)	(Linked data)	(Linked data)	(Linked data)
Treatment effect	-0.053*** (0.019)	-0.035** (0.015)	-0.045** (0.018)	-0.047*** (0.015)	0.007 (0.011)	-0.004 (0.005)
No. of obs	35,386	29,521	29,521	29,521	29,521	29,521

Sample: Roma women born in 1980-1993. Intention-to-treat group: born on or after June 1, 1991. Control group: born before June 1, 1991. Control variables: 4th order polynomial function of the running and running\*intention-to-treat variable, year of birth fixed effect (FE), month of birth FE, day of the week of birth FE, settlement type FE, county FE. Age is a continuous variable; giving birth by age 18 means giving birth either before, or exactly on, the mother's 18th birthday. Robust standard errors clustered by birth year and month in parenthesis. Significance: \*10%, \*\*5%, \*\*\*1% .

sample using the originally estimated CCT bandwidth (188.3 days below and above the cutoff, as in Table 4). When  $i=0$ , we do not exclude anybody and get the same result as before (-0.068, Table 4). When  $i=1, 2, \dots, 60$ , we exclude everybody born in the +/- 1, 2, ..., 60-day interval around the cutoff. As shown in Figure 12, our estimates stay fairly stable between -.089 and -.037 (mean of estimates: -.067) in spite of excluding the hypothetical gamers. We conclude that even if some parents of children born right around the cutoff had played against the enrollment rule, not these observations are leading our results.

## 7 Discussion

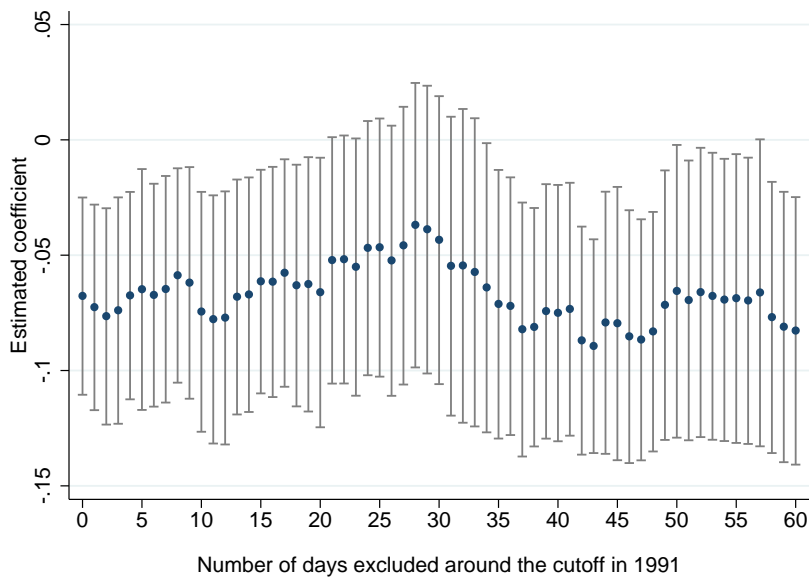
This paper looked at the ITT effects of increasing the compulsory school leaving age from 16 to 18 on teenage fertility using a fuzzy RDD identification strategy around a cutoff date of birth. We find that the reform decreased the probability of having the first child by age 18 among Roma women by 3.5-6.8 percentage points, or 13.4-26.0%. Our findings suggest that the effect of longer schooling among Roma women fades away by age 20; thus, the reform delayed motherhood by about two years.

We are the first in this literature to separate the incapacitation and human capital effects of education on fertility by exploiting a unique database that captures the conception time of pregnancies precise to the week. Our data cover all known pregnancies, including those ending with live birth, still birth, miscarriage and abortion. While [Geruso and Royer \(2018\)](#) infer the probability of conceptions from the estimated effect of CSL age on the prevalence of abortions, we directly show that higher CSL age decreased the probability of getting pregnant. We find that the reform decreased the probability of getting pregnant during the school year only, and not during summer or Christmas breaks, suggesting that the effect was generated mainly through the incapacitation channel. We also show that extended compulsory schooling decreased the probability of having an abortion among Roma women. We find suggestive evidence that those who did get pregnant became less likely to end their pregnancy with abortion indicating that extended schooling might have decreased the prevalence of unwanted pregnancies more.

The effects that we find on the teenage motherhood of Roma women fit the estimates of earlier studies focusing on the general population and are relatively large. The most comparable effects in the literature range between 5.8 – 11.5% ([Black et al., 2008](#)) and 35% ([Cygan-Rehm and Maeder, 2013](#)). In terms of the mechanisms, our findings are in line with [Geruso and Royer \(2018\)](#) who found that increasing the CSL age



Figure 12: Estimated effects on the probability of motherhood by age 18 around the cutoff if those born in the +/- 0-60-day interval around the cutoff are excluded from the sample (Donut hole RDD), Roma women



Each plotted coefficient is a separate RDD-estimate around the cutoff (1 June, 1991). Bandwidth is 188.3 days below and above the cutoff, set using the bandwidth optimization routine of [Calonico et al. \(2014b\)](#). 95% confidence intervals are plotted around all estimates based on robust standard errors clustered by month of birth. Roma women only. Age is a continuous variable; giving birth by age 18 means giving birth either before, or exactly on, the mother's 18th birthday. No. of observations: 2,906 when  $i=0$  and no one is excluded, and, 2,202 when  $i=60$  and when everyone is excluded who was born in the +/- 60-day interval around the cutoff.

in the UK in 1972 had an incapacitation effect only. The rest of the literature finds at least some evidence on human capital effects of increasing the CSL age; either during the teenage years, but still after the age that was directly affected by the reform (Black et al., 2008; Silles, 2011; Wilson, 2017), or even later in life (Cygan-Rehm and Maeder, 2013).

Looking at all women together, or non-Roma women in particular who make up over 90% of the sample, we do not find any effects on teenage fertility. This result is in contrast with earlier studies of similar reforms and is due mainly to the small bite of the Hungarian reform among non-Roma women. Most earlier studies look at reforms implemented between 1920 and 1972 that increased school participation rates substantially. By contrast, at the end of the 2000s in Hungary, the vast majority of non-Roma teenagers stayed in school until age 18 already before the reform. Furthermore, most earlier reforms increased the CSL age from 14 to 15 or from 15 to 16 while the Hungarian reform increased it from 16 to 18. It is likely that the effects of compulsory schooling on fertility are decreasing by age. For example, (Black et al., 2008) look at various extensions of compulsory schooling across US states and find that while increasing the CSL age to 16 and 17 significantly reduces the probability of teenage motherhood, increasing it further to 18 does not.

The large size of the effect that we find for Roma women implies that raising the school leaving age can be effective in reducing the incidence of teenage pregnancy among socially excluded women even if it does not affect the general population. We believe that our results have external validity for developed countries where disadvantaged ethnic minorities, characterized by high prevalence of teenage fertility, live in social exclusion. It is an important policy implication that the impact of educational interventions might be heterogeneous by ethnic background: looking only the average effects might hide socially and economically important impacts that are related to relatively small groups of ethnic minorities.

## References

- Adamecz-Völgyi A (2018) Increased compulsory school leaving age affects secondary school track choice and increases dropout rates in vocational training schools. Budapest Working Papers on the Labour Market 1801, Institute of Economics, Centre for Economic and Regional Studies, Hungarian Academy of Sciences
- Altwickler-Hámori S, Köllő J (2012) Whose Children Gain from Starting School Later? Evidence from Hungary. *Educational Research and Evaluation: An International Journal on Theory and Practice* 18(5):459–488
- Angrist JD, Krueger AB (1991) Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics* 106(4):979–1014
- Bean FD, Swicegood G (1985) *Mexican American Fertility Patterns*. University of Texas Press
- Berthelon ME, Kruger DI (2011) Risky behavior among youth: Incapacitation effects of school on adolescent motherhood and crime in Chile. *Journal of Public Economics* 95(1-2):41–53
- Berthoud R (2001) Teenage births to ethnic minority women. *Population Trends* (104):12–17
- Bifulco R, Lopoo LM, Oh SJ (2015) School desegregation and teenage fertility. *Educational Evaluation and Policy Analysis* 37(4):591–611
- Black SE, Devereux P, Salvanes KG (2008) Staying in the classroom and out of the maternity ward? The effect of compulsory schooling laws on teenage births. *Economic Journal* 118(530):1025–1054

- Black SE, Devereux PJ, Salvanes KG (2011) Too young to leave the nest? The effects of school starting age. *The Review of Economics and Statistics* 93(2):455–467
- Bonell C (2004) Why is teenage pregnancy conceptualized as a social problem? A review of quantitative research from the USA and UK. *Culture, Health & Sexuality* 6(3):255–272
- Buckles KS, Hungerman DM (2013) Season of birth and later outcomes: Old questions, new answers. *The review of economics and statistics* 95(3):711–724
- Calonico S, Cattaneo MD, Titiunik R (2014a) Robust data-driven inference in the regression-discontinuity design. *Stata Journal* 14(4)
- Calonico S, Cattaneo MD, Titiunik R (2014b) Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82:2295–2326
- Cheng MY, Fan J, Marron JS (1997) On automatic boundary corrections. *The Annals of Statistics* 25(4):1691–1708
- Chevalier A, Viitanen TK (2003) The long-run labour market consequences of teenage motherhood in Britain. *Journal of Population Economics* 16(2):323–343
- Crowe DM (1995) *A History of The Gypsies of Eastern Europe and Russia*. Palgrave Macmillan
- Cygan-Rehm K, Maeder M (2013) The effect of education on fertility: Evidence from a compulsory schooling reform. *Labour Economics* 25:35–48
- Durst J (2007) Több a kára, mint a haszna: születésszabályozás a "gettóban". *Demográfia* 50(1):74–103.
- Education Office of the Ministry of Human Resources (2014) Országos jelentés. a magántanulóvá nyilvánítással kapcsolatos eljárás szakmai ellenőrzése. Tech. rep., Oktatási Hivatal Tanügy-igazgatási és Ellenőrzési Osztály, Budapest
- Fan E, Liu JT, Chen YC (2014) Is the 'quarter of birth' endogenous? Evidence from one million siblings in Taiwan. Working Paper 20444, National Bureau of Economic Research
- Fletcher J, Padrón N (2016) The effects of teenage childbearing on adult soft skills development. *Journal of Population Economics* 29(3):883–910, DOI 10.1007/s00148-016-0589-9
- Fletcher JM, Wolfe BL (2009) Education and labor market consequences of teenage childbearing: Evidence using the timing of pregnancy outcomes and community fixed effects. *Journal of Human Resources*, University of Wisconsin Press 44(2)
- FRA (2014) Roma survey - Data in focus Poverty and employment: The situation of Roma in 11 EU Member States. Tech. rep., FRA – European Union Agency for Fundamental Rights
- Fryer RG, Torelli P (2010) An empirical analysis of 'acting white'. *Journal of Public Economics* 94(5):380–396
- Gábos A, Tátrai A, B Kis A, Szivos P (2006) Anyagi depriváció Magyarországon, 2009–2015. *Társadalmi Riport* 14(1):24
- Geruso M, Royer H (2018) The impact of education on family formation: Quasi-experimental evidence from the UK. Working Paper 24332, National Bureau of Economic Research

- Hablicsek L (2007) Kísérleti számítások a roma lakosság területi jellemzőinek alakulására és 2021-ig történő előrebecslésére. *Demográfia* 50(1):7–54.
- Hahn J, Todd P, Van der Klaauw W (2001) Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69(1):201–209
- Hajdú T (2015) How can young Roma achieve success in Hungary? Roma in secondary schools and in the labor market. Tech. rep., Társi Social Research Institute, Budapest
- Hajdú T, Kertesi G, Kézdi G (2014) Roma fiatalok a középiskolában. Beszámoló a TÁRKI Életpálya-felmérésének 2006 és 2012 közötti hullámaiból. In: Kolosi T, Tóth IG (eds) *Társadalmi Riórt*, 2014, TÁRKI, Budapest, pp 265–302
- Imbens G, Lemieux T (2008) Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2):615–635
- Institute B (2013) Evaluation of the Roma inclusion impact of two mainstream EU funded employment programmes. Tech. rep., Budapest Institute
- Jacob BA, Lefgren L (2003) Are idle hands the devil’s workshop? Incapacitation, concentration, and juvenile crime. *American Economic Review* 93(5):1560–1577
- James J, Vujić S (2019) From high school to the high chair: Education and fertility timing. *Economics of Education Review* 69:1–24, DOI 10.1016/j.econedurev.2018.12.002
- Janky B (2005) A cigány nők társadalmi helyzete és termékenységé. In: Nagy I, Pongrácz T, Tóth IG (eds) *Szerepváltozások. Jelentés a nők és a férfiak helyzetéről*, SZCSEM-Tárki, Budapest, pp 136–148
- Janky B (2007) A korai gyermekvállalást meghatározó tényezők a cigány nők körében. *Demográfia* 50(1):55–73.
- Kazuska M (2012) A tankötelezettség múltja, jelene és jövője. *Miskolci Jogi Szemle* VII.(1):128–142
- Kearney MS, Levine PB (2012) Why is the teen birth rate in the United States so high and why does it matter? *Journal of Economic Perspectives* 26(2):141–163
- Kearney MS, Levine PB (2014) Income Inequality and Early Nonmarital Childbearing. *Journal of Human Resources* 49(1):1–31
- Kemény I, Janky B (2005) Roma population of Hungary 1971-2003. *East European Monographs* (702):70
- Kende A, Illés A (2007) A rugalmas beiskolázás és az oktatási szakadék összefüggései. *Új Pedagógiai Szemle* 57(11):17–41
- Kertesi G (2005) A társadalom peremén - romák a munkaerőpiacon és az iskolában. Osiris Kiadó, Budapest
- Kertesi G, Kézdi G (2006) A hátrányos helyzetű és roma fiatalok eljuttatása az érettségéhez. Egy különösen nagy hosszú távú költségvetési nyereséget biztosító befektetés. *Budapest Working Papers on the Labour Market 2006/6*, Institute of Economics, Centre for Economic and Regional Studies, Hungarian Academy of Sciences
- Kertesi G, Kézdi G (2009) Roma és nem roma fiatalok középiskolai továbbtanulása. In: Fazekas K (ed) *Oktatás és foglalkoztatás*, MTA Közgazdaságtudományi Intézet, Budapest, pp 122–136

- Kertesi G, Kézdi G (2010) Iskolázatlan szülők gyermekei és roma fiatalok a közép-iskolában. Beszámoló az educatio életpálya-felvételének 2006 és 2009 közötti hullámaiból. Társadalmi Riport, Tárki Social Research Institute
- Kertesi G, Kézdi G (2011) The Roma/non-Roma test score gap in Hungary. *American Economic Review* 101(3):519–525
- Kertesi G, Kézdi G (2014) On the test score gap between Roma and non-Roma students in Hungary and its potential causes. Budapest Working Papers on the Labour Market 2014/1, Institute of Economics, Centre for Economic and Regional Studies, Hungarian Academy of Sciences
- Kézdi G, Simonovits G (2016) Economic hardship triggers identification with disadvantaged minorities. *The Journal of Politics* 78(3):882–892
- Kirdar MG, Dayioglu M, Kocç I (2018) The Effects of Compulsory-Schooling Laws on Teenage Marriage and Births in Turkey. *Journal of Human Capital* 12(4):640–668, DOI 10.1086/700076
- Klepinger D, Lundberg S, Plotnick R (1999) How Does Adolescent Fertility Affect the Human Capital and Wages of Young Women? *Journal of Human Resources* 34(3):421–448
- Ladányi J, Szelényi I (2002) Cigányok és szegények Magyarországon, Romániában és Bulgáriában. *Szociológiai Szemle* 2002(4):72–94
- Lee DS, Lemieux T (2010) Regression Discontinuity Designs in Economics. *Journal of Economic Literature* 48(2):281–355
- Lopoo LM (2004) The effect of maternal employment on teenage childbearing. *Journal of Population Economics* 17(4):681–702
- Mártonfi G (ed) (2011a) A 18 éves korra emelt tankötelezettség teljesülése és (mellék)hatásai. Oktatókutatató és Fejlesztő Intézet, Budapest
- Mártonfi G (ed) (2011b) Hány éves korig tartson a tankötelezettség? Válaszkísérlet egy rossz kérdésre: szakpolitikai javaslat. Oktatókutatató és Fejlesztő Intézet, Budapest
- McCrary J, Royer H (2011) The effect of female education on fertility and infant health: Evidence from school entry policies using exact date of birth. *The American Economic Review* 101(1):158–195
- Messing V (2011) Megjegyzések egy „roma” projekt margójára. *sociohu* 1(2)
- National Institute of Public Education (2006) Jelentés a magyar közoktatásról 2006. Oktatókutatató és Fejlesztő Intézet
- Navarro Paniagua M, Walker I (2012) The impact of teenage motherhood on the education and fertility of their children: Evidence for Europe. IZA Discussion Papers 6995, Institute for the Study of Labor (IZA)
- OECD (2014) PISA 2012 results in focus: What 15-year-olds know and what they can do with what they know. Tech. rep., OECD
- OECD (2015) Education Policy Outlook: Hungary. Tech. rep.
- Rubin DB (2005) Causal Inference Using Potential Outcomes. *Journal of the American Statistical Association* 100(469):322–331, DOI 10.1198/016214504000001880

- Silles MA (2011) The effect of schooling on teenage childbearing: Evidence using changes in compulsory education laws. *Journal of Population Economics* 24(2):761–777
- Szikra D (2010) Serdülőkori terhességek – egy jövőbeli kutatást megalapozó tanulmány és összefoglalás. MTA Gyermekszegénység elleni program
- Tan PL (2017) The impact of school entry laws on female education and teenage fertility. *Journal of Population Economics* 30(2):503–536
- Tárki (2013) Tárki Monitor jelentések 2012. egyenlőtlenség és polarizálódás a magyar társadalomban. Tech. rep., Tárki Social Research Institute
- UNFPA (2013) Adolescent Pregnancy: A Review of the Evidence. Tech. rep., United Nations Population Fund, New York
- Wilson T (2017) Compulsory education and teenage motherhood. Stirling Economics Discussion Paper 2017-01
- Wolfe B, Haveman R, Pence K, Schwabish JA, Schwabish JA (2007) Do youth nonmarital childbearing choices reflect income and relationship expectations? *Journal of Population Economics* 20(1):73–100

## A Appendix

Table 9: The balance of the linked data

	Mean, unlinked obser- vations	No. of unlinked obser- vations	Mean, linked obser- vations	No. of linked obser- vations	T-test p-value
All events					
Born after June 1, 1991	0.48	2,234	0.48	1,480	0.88
Age	16.71	2,234	16.69	1,480	0.48
Completed elementary school	0.76	2,234	0.83	1,480	0.00
Lives in a settlement with less than 10,000 inhabitants	0.67	2,234	0.72	1,480	0.00
Live births					
Born after June 1, 1991	0.50	924	0.50	701	0.94
Age	16.77	924	16.69	701	0.09
Completed elementary school	0.72	924	0.76	701	0.05
Lives in a settlement with less than 10,000 inhabitants	0.71	924	0.77	701	0.01
Abortions					
Born after June 1, 1991	0.47	1089	0.47	670	0.84
Age	16.65	1089	16.69	670	0.42
Completed elementary school	0.79	1089	0.89	670	0.00
Lives in a settlement with less than 10,000 inhabitants	0.64	1089	0.67	670	0.26
Miscarriages					
Born after June 1, 1991	0.48	203	0.47	106	0.85
Age	16.74	203	16.62	106	0.32
Completed elementary school	0.82	203	0.89	106	0.12
Lives in a settlement with less than 10,000 inhabitants	0.65	203	0.69	106	0.50
Stillbirths					
Born after June 1, 1991	0.44	18	0.67	3	0.50
Age	16.92	18	17.69	3	0.20
Completed elementary school	0.89	18	1.00	3	0.57
Lives in a settlement with less than 10,000 inhabitants	0.83	18	1.00	3	0.47

Sample: women born within 180 days before or after June 1, 1991. Women living in settlements with less than 50,000 inhabitants only.

Table 10: The effects of personal characteristics on Vital Statistics database events being linked to the Census

	All events	Live births	Abortions	Miscarriages	Stillbirths
Age at the event	-0.025*** (0.006)	-0.035** (0.013)	-0.022* (0.008)	-0.044 (0.028)	0.198 (0.140)
Educational status at the event	0.124*** (0.015)	0.094*** (0.024)	0.188*** (0.026)	0.210* (0.078)	0.195 (0.259)
Settlement size	-0.006** (0.002)	-0.001 (0.005)	-0.008 (0.004)	-0.003 (0.009)	0.011 (0.056)
No. of obs	3,714	1,625	1,756	309	21

Sample: women born within 180 days before or after June 1, 1991. Women living in settlements with less than 50,000 inhabitants only. Linear probability models, one regression per event type. Dependent variable: whether event  $i$  could be linked to the Census. Additional control variables: number of days of birth relative to June 1, 1991, county fixed effects. Robust standard errors clustered by month of birth are in parenthesis.

Table 11: Treatment effect on the probability of linking pregnancy events to the Census

Events in the Vital Statistics Database	Local linear model	Global polynomial model
All events	-0.049** (0.017) [3,702]	-0.009 (0.027) [18,999]
Live birth only	-0.026 (0.060) [1,625]	0.032 (0.048) [8,338]
Abortion only	-0.061** (0.025) [1,759]	-0.084* (0.042) [9,020]
Miscarriage only	-0.098 (0.091) [309]	-0.320** (0.092) [107]
Stillbirth only	0.113 (0.334) [21]	0.196 (0.390) [1,534]
Sample	Women born within 180 days before or after June 1, 1991	Women born in 1988-92
Control variables	Linear function of the running variable separately below and above the cutoff.	4th order function of the running variable separately below and above the cutoff; FE: year of birth, month of birth, day of the week at birth, county, settlement type.

Linear probability models. Left hand side variable: whether observation  $i$  in the Vital Statistics database could be linked to the Census. Women living in settlements with less than 50,000 inhabitants only. Robust standard errors clustered by birth year and month in parenthesis, observation numbers in brackets. A negative treatment effect indicates that the pregnancy and birth-related events of those born after June 1, 1991, are less likely to be linked to the Census.



Table 12: Aggregate data on grade repetition in Grades 1-3, % of students in grade

Academic year	Grade 1	Grade 2	Grade 3	CSL age
1995/1996	4.0	1.9	1.6	16
1996/1997	3.9	2.0	1.5	16
1997/1998	3.9	1.9	1.5	16
1998/1999	4.0	1.8	1.5	18
1999/2000	3.9	1.9	1.4	18
2000/2001*	4.2	1.9	1.4	18

\*The 2000/2001 is the academic year in which the 2001 Census is taken. Data Source: [National Institute of Public Education \(2006\)](#)

Table 13: Grade repetition in Grades 1-4, % of students in Grade

No. grade repetitions	Non-Roma students	Roma students	Roma girls	Roma boys
None	94.51	80.04	80.86	79.24
One	4.89	17.12	16.36	17.87
Two	0.40	2.20	1.86	2.53
More than two	0.02	0.37	0.37	0.36
No data	0.15	0.27	0.56	0.00
Total no. of obs.	8,930	1,092	538	554

Data Source: own estimation from the Hungarian Life Course Survey.

Table 14: Average compliance with the age rule of compulsory schooling, PES data

	Early starters (never- takers)	Compliers	Late starters (always- takers)	No. of students in Grade 1
Compliance by academic years, share of those starting elementary school at the given year				
1997/1998	0.02	0.80	0.18	127,214
1998/1999	0.02	0.78	0.20	125,875
1999/2000	0.01	0.78	0.21	121,424
Compliance by cohorts, share of cohort size				
Born between June 90-May 91	0.02	0.79	0.19	129,489
Born between June 91-May 92	0.02	0.78	0.20	126,294

Data Source: Public Education Statistics (PES) of the Public Education Information System (KIR-STAT). "Early starters" refers to those entering elementary school without reaching age 6 by May 31. "Compliers" refers to those entering elementary school according to the age rule of compulsory schooling at age 6. "Late starters" refers to those entering elementary school a year later than expected under the age rule.