Institute of Education

Books or babies? Incapacitation and human capital effects of extended compulsory schooling on the teenage fertility of ethnic minority women

Anna Adamecz-Völgyi
Ágota Scharle

Department of Quantitative Social Science
Working Paper No. 19-01
Feb 2019

Disclaimer

Department of Quantitative Social Science, UCL Institute of Education, University College London,

20 Bedford Way, London WC1H 0AL, UK

Books or babies? Incapacitation and human capital effects of extended compulsory schooling on the teenage fertility of ethnic minority women

Anna Adamecz-Völgyi[1] and Ágota Scharle[2]

Abstract

This paper examines the effects of an increase in the compulsory school-leaving (CSL) age on the teenage fertility of Roma women, a disadvantaged ethnic minority in Hungary. We use a regression discontinuity design (RDD) identification strategy and show that raising the CSL age from 16 to 18 decreased the probability of teenage motherhood among Roma women by 6.8 percentage points and delayed early motherhood by two years. We contribute to the literature 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 to separate the incapacitation and human capital effects of education on fertility. We find that higher CSL age 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.

JEL Codes: J13, C21, I26

Keywords: education; compulsory school leaving age; teenage fertility; disadvantaged ethnic minorities; regression discontinuity design

Contact Details: Anna Adamecz-Völgyi (a.adamecz-volgyi@ucl.ac.uk), Department of Social Science, UCL Institute of Education, University College London

Acknowledgements: The authors thank Gábor Kézdi, Hessel Oosterbeek, Róbert Lieli, Gábor Kertesi, Dorottya Szikra, Balázs Váradi, Balázs Kapitány, Ildikó Husz, and Béla Janky for their useful comments and suggestions. We thank Flóra Samu her great work as research assistant. 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.

---

[1] Department of Social Science, UCL Institute of Education, University College London and Budapest Institute for Policy Analysis
[2] Budapest Institute for Policy Analysis

1 Introduction

Teenage fertility 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, higher rates of benefit receipt, and infant mortality (Chevalier and Viitanen, 2003; Fletcher and Wolfe, 2009; Wilson, 2012). 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 the high opportunity costs, the prevalence of teenage motherhood has been declining in most developed countries. 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. 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. Kearney and Levine (2012) propose 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".

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); length of school day: Berthelon and Kruger (2011); length of schooling: Black, Devereux, and Salvanes (2008); Silles (2011); Cygan-Rehm and Maeder (2013); Geruso and Royer (2018). Most papers find that having more education does reduce the probability of early childbearing. 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.

The impact of education may work via two main channels: incapacitation or human capital investment. Incapacitation[1] refers to the possibility that teenagers may not have the desire, time, or opportunity to have a child while they are in school. It is important to note that this does not necessarily imply that teenagers are less sexually active while in school. It might also happen that they pay more attention to contraception and/or they are more likely to use morning-after pills or choose abortion in case of emergency.

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. Black, Devereux, and Salvanes (2008) find weak evidence of the incapacitation effect and somewhat larger effects through human capital development, while Wilson (2012) finds strong effects through both channels. 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.

This paper examines whether extended compulsory education may affect teenage childbearing among Roma women, despite their poor labour market opportunities and the associated low opportunity cost of

[1]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.

motherhood. Roma make up the largest ethnic minority in Hungary, accounting for about 5-7% of the total population, or 500,000-700,000 people. Belonging to the Roma minority is strongly correlated with poverty, social exclusion, long term unemployment, and access to low quality public services, including health care and education (Kemény and Janky, 2003; Ladányi and Szelényi, 2002; Kertesi, 2005; Gábos et al., 2006). The Roma employment gap is especially large for women (FRA, 2014). Teenage fertility of non-Roma women is very low in Hungary, 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 effects of longer schooling on the teenage fertility of Roma women by exploiting the reform of compulsory schooling introduced in Hungary 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 age of elementary school entry rule that allows children to start school at age 6. Assuming full compliance with the age rule, children born on June 1, 1991 started elementary school under the new CSL age scheme, while those born born before this date had started elementary school in the previous year, under the old CSL age. Thus, compliance to the age rule generates a discontinuity in the likelihood of starting elementary school under the new CSL age regime at this date of birth. As compliance is not perfect, we set up a fuzzy regression discontinuity design (RDD) identification strategy, using June 1, 1991 as a cutoff, to estimate the intention-to-treat (ITT) effect of the reform on the probability of both early childbearing and pregnancy. We exploit a unique data-set of all known pregnancies, including live births, abortions, fetal losses, and still births, linked to a large subsample of the 2011 Hungarian Census.

We find that the higher CSL age decreased the probability of teenage motherhood among Roma women by 6.8 percentage points. This effect creates a two-year delay in motherhood. We find no effect among non-Roma women, where teenage fertility is rare. We use our rich data sets to reconstruct the conception time of all pregnancies of Roma women. We show that the reform decreased the probability of getting pregnant during the school year but not during summer and Christmas breaks, suggesting that the effect is generated mainly by the incapacitation channel. The fact that we find no evidence on effects realized through the human capital channel is consistent with the explanation that the poor labour market prospects of ethnic minority women must be an important factor in keeping the opportunity cost of childbearing low. Furthermore, it implies that even if a school age reform has a high contemporaneous effect on teenage fertility among ethnic minorities, its human capital effect might still be weak. The results contribute to the understanding of the fertility of disadvantaged ethnic minorities, and also represent the first attempt to show that being in school lowers the probability of getting pregnant (and not just child-bearing).[2] Our findings on the incapacitation effect of education are analogous to that of 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. Our identification strategy and empirical methods are detailed in Section 4. Section 5 presents the results, Section 6 reports several robustness checks, while Section 8 concludes.

2 Data and the teenage fertility of Roma women

2.1 Data

We begin with an overview of our data sources, the data of which we will refer to in later sections. This paper uses three data sources: the 2001 Hungarian Census, the 2011 Hungarian Census, and the Vital Statistics

[2] Berthelon and Kruger (2011) document that being physically present in school contemporaneously lowers the probability of motherhood

database. As it is detailed in Section 4, we estimate the effects of increased CSL age in a fuzzy regression discontinuity design (RDD) framework, where compliance with the age of elementary school entry rule generates a discontinuity in the likelihood of being exposed to the new CSL age scheme. We use the 2001 Hungarian Census to demonstrate the compliance to the age rule, i.e. the jump in the likelihood of starting elementary school under the new CSL age legislation around the cutoff, which is 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 of the observed individuals, and, for those in school, it registers which grade of school they were attending at that time. Information on their school grade in 2001 allows us to estimate the jump in the probability of starting school under the new CSL age legislation around the cutoff in birth year and month bins.

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 birth date 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 in a family, independent from whether the children live with their parents or not. We use this 2011 Hungarian Census data to estimate whether Roma women became mothers during their teenage years.

The Vital Statistics databases cover all pregnancy-ending medical events in Hungary. They record 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 earlier 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.

Our identification strategy is based on a discontinuity in date of birth at June 1, 1991 (see Section 4 for more details). Considering this fact and the purpose of this analysis, our 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 before or after June 1, 1991.

Tables 14-16 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 CSL age legislation 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 CSL age increase 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 9). Furthermore, controlling for year and month of birth, settlement size, and county fixed effects reduces the systematic relationship between the treatment status of women and the link-ability of their Vital Statistics events substantially (see Table 16 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 12).

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 in 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 September 1 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). 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, 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 (Budapest 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)[3]. 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 in Hungary, while the 2011 Hungarian Census reported 315,525 Roma people[4]. In spite of this improvement, demographers estimate that the 2011 Hungarian Census still identifies only half of the total number of the Roma population (Hablicsek, 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 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 CSL age increase on Roma women if a declaration of being Roma is not related to the legislative change. This issue is discussed in Section 4.

2.3 Teenage fertility in Hungary

Table 1 summarizes the timing pattern of first birth-giving of teenagers born before the CSL age increase. The prevalence 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. The share of teenage mothers among Roma women is similar to rates found in developing countries like Kenya, Eritrea, or the Congo (UNFPA, 2013).

There could be several explanations for the high prevalence of early childbearing among Roma women but little evidence exists on this. Teenage fertility has been historically high among Roma women; a 14-year-old girl or boy is treated as an adult in Roma communities. 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[5].

Although teenage fertility has been decreasing in Hungary on average, even among Roma women, this is not observed within some groups of poor (Szikra, 2010). In particular, in some marginalized Roma communities, fertility is either on the rise or it has stabilized at a high rate (Durst, 2007). It should be noted, however, that Roma communities themselves are heterogeneous with respect to their fertility patterns (Janky, 2005).

[3]The total number of Census tract units is over 47,000 (Data source: Hungarian Statistical Office).

[4]Own estimation from the 2001 and 2011 Hungarian Censuses.

[5]Own calculation from the 2011 Census.

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

Probability of giving birth by reaching age ...	All women	Roma women	Non-Roma women	Ethnicity unknown
16	0.007	0.065	0.004	0.007
	(0.000)	(0.003)	(0.000)	(0.085)
17	0.017	0.150	0.010	0.019
	(0.000)	(0.004)	(0.000)	(0.136)
18	0.032	0.260	0.020	0.038
	(0.000)	(0.005)	(0.001)	(0.192)
19	0.052	0.379	0.034	0.063
	(0.001)	(0.006)	(0.001)	(0.243)
20	0.073	0.472	0.052	0.090
	(0.001)	(0.006)	(0.001)	(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 reaching age 16", for example, means giving birth either before, or exactly on, the mother's 16th birthday.

3 Institutional background and the reform

3.1 Roma students in the Hungarian education system

Most adolescents typically are in secondary school when they reach CSL age, although some might still be in elementary school, especially those from the lowest socio-economic backgrounds. Prior to the CSL age increase in 2001, 9% of 15-year-old non-Roma women were attending primary school and 88% were attending secondary school right before reaching the actual CSL age of 16 that was in place at that time. These rates for Roma women were 46% and 29%, respectively[6].

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, to repeat grades in both elementary and secondary school, and if they are admitted to secondary school they are more likely to go to lower-level secondary schools than their non-Roma fellow students. 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 backgrounds (Kertesi and Kézdi, 2014). The Hungarian education system is rated as one of the worsts among the OECD countries when it comes to dealing with 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 in this country group (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).

[6]Own estimation from the 2001 Census. For the rest of the sample (3% and 15%, the information on the type of school is either missing, they live in institutions, or they are not in school due to living with disabilities.

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). We know very little about homeschooling in general.

3.2 Compulsory education and the reform

Before the reform, students had to stay in the formal education system until the end of the academic year in which they reached age 16. The new law, the Public Education Act (1996), extended compulsory school attendance by two years, from age 16 to age 18. Thus, it resulted in students spending two additional years in public education. The new legislation was introduced gradually (i.e., it was grandfathered in), and it first influenced students who started elementary school in the 1998/99 academic year. The first cohort under the influence of the new legislation was already aware of the increase in compulsory schooling at age 6. Although the Public Education Act (1996) 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 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 influenced cohort entered secondary education in 2006, it was completed for half a decade.

Elementary school in Hungary has 8 grades. According to the general rule of school entrance, 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 this 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 and stay in school at least until age 16, while those born on June 1, 1991, were more likely to start elementary school under the new CSL age legislation of age 18 and stay in school two years longer. We build our identification strategy on this cutoff date of birth created by compliance to the rule of elementary school entry.

Compliance to the age rule is not perfect. On top of the rule, elementary school entry 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[7]. 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".

On average, compliance with the age rule is 78-80%. About 18-20% of a cohort start elementary school a year later then they are supposed to based on their date of birth. Early school start is rare, occuring at less than 2% (see Table 2). The share of late starters is the highest among those born right before the cutoff. Among women born in May 1991, 47% of them started elementary school under the higher CSL age scheme, while 85% of the women born in June 1991 began school under the higher CSL age scheme (see Figure 1).

4 Identification strategy and empirical methods

4.1 Identification strategy[8]

As detailed in Section 3.2, our identification strategy is based on compliance with the age of elementary school entry rule, which creates a discontinuity in the probability of starting elementary school under the increased CSL age regime at the date of birth of June 1, 1991. This discontinuity allows us to identify the ITT effects of the increase using a regression discontinuity design (RDD) strategy, with being born on June 1, 1991, as the cutoff.

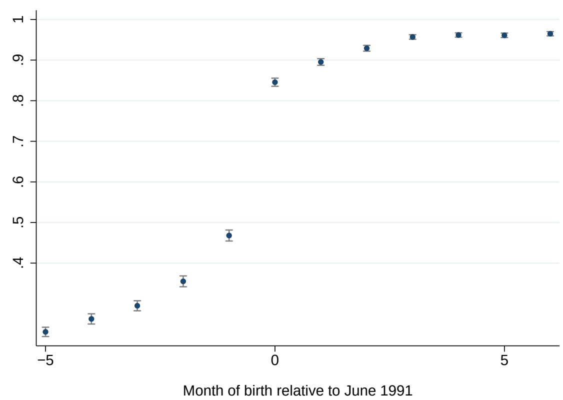
[7]Pedagógiai Szakszolgálat in Hungarian.
[8]Our identification strategy draws heavily on Adamecz-Völgyi (2018).

Table 2: Compliance to the age rule of compulsory schooling

	Share of early starters	Share of comp-liars	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.

Figure 1: The Probability of starting school under CSL age 18, women born in 1991



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.

Table 3: 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 4: 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.

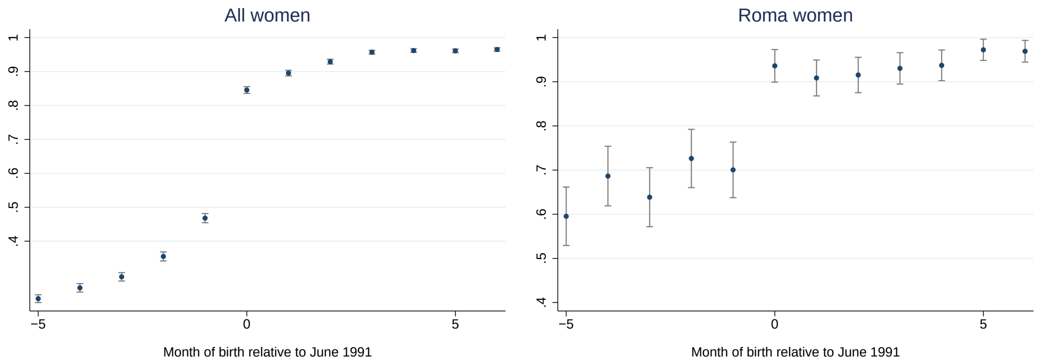
Estimating the size of the jump in the likelihood of starting school under the new CSL age scheme around the cutoff, is straightforward for all women in general but challenging for Roma women in particular. As mentioned in Section 2, we observe compliance with the age rule in birth year and month bins from the 2001 Hungarian Census. The 2001 Hungarian Census registers which grade of school the cohort of students born in 1991 was attending at the time the data was collected. Knowing their grade level in 2001 allows us to estimate the size of the jump, assuming that grade repetition patterns did not change between 1997 and 2001. Those who started elementary school in 1997, under the old CSL age scheme, were in Grade 4, and those who started elementary school in 1998, under the new CSL age scheme, were in Grade 3 in 2001. If we see an individual in a lower grade than where s/he is supposed to be based on the age rule, it means one of two things: s/he started school later, or s/he repeated grades. Although the 2001 Hungarian Census does not register grade repetitions, this information is available on an aggregate level from the Public Education Statistics of the Public Education Information System database. Table 3 summarizes the share of students repeating a grade during Grades 1-3, and shows that the average prevalence of grade repetition is quite low, 1.4-4.0%, and its pattern does not change during this period (National Institute of Public Education, 2006). The prevalence of grade repetition is the highest in Grade 1, and it is lower, at around 1.5%, in Grades 2 and 3. Those who started elementary school in 1997 should have been in a higher grade in 2001 and thus, they are 1.5 percentage points more likely to have repeated a grade by the time of observation. Consequently, the only problem with estimating the size of the jump in general is that this method will result in estimating a 1.5 percentage point lower jump in the probability of starting school under the new scheme because we cannot distinguish between a late school start and a grade repetition.

We face several challenges when looking at the size of the jump in the probability of being treated around the cutoff for Roma women in particular. First, although the 2001 Hungarian Census has ethnic markers, the underreporting problem is much more serious there than it is in the 2011 Hungarian Census (see Section 2). Second, the probability of repeating a grade at least once during Grades 1-4 is five times higher among Roma than among non-Roma students. The only data on grade repetition of Roma students in Hungary come from the Hungarian Life Course Survey. The Hungarian Life Course Survey is a cohort study that follows 10,000 youth, who were in Grade 8 in the spring of 2006 (Hajdú, 2015). As Table 4 demonstrates, only 5% of non-Roma students repeated a grade at least once during Grades 1-4, whereas 20% of Roma students repeated at least once. Thus, our method underestimates the size of the jump around the cutoff in the case of Roma students. Figure 2 compares the jump in the probability of starting school in 1998 between all women and Roma women specifically. The data show three main differences:

- The data of Roma women have a higher variance, and the 95% confidence intervals are much larger. This is partly due to the smaller sample size of Roma women (2,313 vs. 60,302).

- The main difference is on the left hand-side of the cutoff: Roma women born before the cutoff are

Figure 2: The probability of starting school in 1998



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.

more likely to be found in a lower grade than they are supposed to be. This phenomenon is due to the high number of grade repetitions.

- Due to grade repetitions, the estimated jump in the probability of starting school under the new CSL age regime is lower for Roma women than for all women in general (24 vs. 36 percentage points).

Comparing the share of Roma girls who have just started elementary school, and thus, have not been able to repeat a grade, to those who have not started school yet gives a better estimation about the size of the jump. 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[9]. For these cohorts, the probability of not starting school yet is equivalent to the treated status defined above, i.e. starting school the next year. Figure 3 proves that once controlling for grade repetition, the size of the jump is indeed higher, at 28 percentage points in both Censuses.

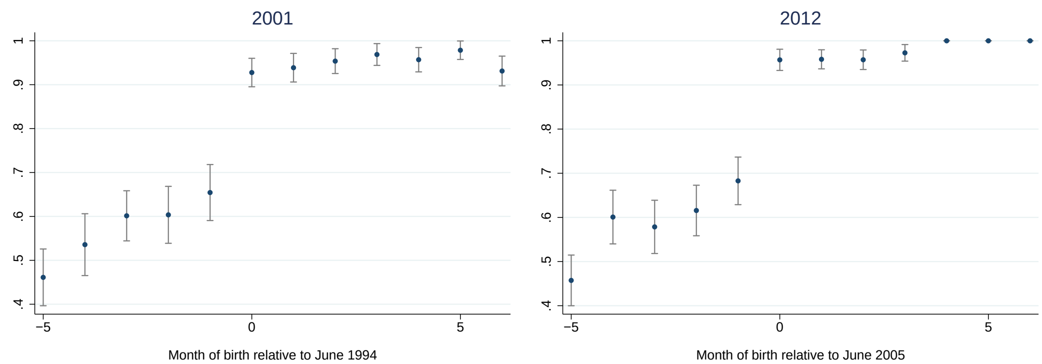
[9]The difference between 2001-1994=7 and 2011-2005=6 comes from the fact that the 2001 Census was taken in the spring while the 2011 Census was taken in the autumn.

Table 5: Compliers, never- and always takers according to the Rubin-framework

	Children born before the cutoff, Z=0	Children born after 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).

Figure 3: The probability of starting school in 2001 and 2012, Roma women



The average probability of starting school in 2001 and 2012, 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: 2,635 and 3,407.

As our data does not allow us to estimate the exact size of the jump for Roma women, this analysis will only estimate the ITT effects of the legislative change. This is a fuzzy RDD setup where compliance with the age of elementary school entry rule is endogenous. Being born right after, rather than before, the cutoff is used as an instrument for starting school under the new CSL age legislation. 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 after or before the cutoff as $Z = 1($*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 $D = 1($*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 5 summarizes the compliers, never- and always-takers according to the Rubin potential outcome framework while their shares in the data is shown in Figure 2. 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 (see Figure 2 and 3).

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 new legislation was introduced five years after the birth of the relevant cohort, the legislative change could not have led to a manipulation of birth. The question whether children born in distinct months of a year might have systematically different 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 legislative change. This assumption is far from being trivial. Independently from the legislative change, those born in May, if compliant with the age rule, are supposed to start school a year earlier than those born in June (Angrist and Krueger, 1991; Hámori and Köllő, 2011). Thus, it is an empirical question whether the fertility decisions of those born in May are different from those born in June in years when no CSL age legislation change occured. On the other hand, 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 legislative change 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 legislation change; 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 legislation change

- did not induce anyone who was born after 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 before the cutoff to purposefully start school a year later in order to be subjected to the new CSL age legislation.

Theoretically, parents could violate the first one if based on their negative preferences for schooling they manipulated school start to save their child born after the cutoff from longer schooling. The second could be violated if parents otherwise compliant to the age rule would withhold their school-ready child born before the cutoff to start school one year later, under the increased CSL age regime. We assume that those born before the cutoff who started school a year later (late starters or always-takers), and those born after the cutoff who started school a year earlier (early starters or never-takers), scheduled their school start because of their general preferences about the best age to start elementary school, and not because they violated the age rule due to the legislative change. According to the data, the share of late starters and early starters was stable in this period. On an aggregate level, the administrative data of the Public Education Statistics of the Public Education Information System[10], which is the Hungarian school census, show the share of early starters (1-2%), compliers (78-80%), and late starters (18-20%) in Grade 1 as stable between 1997 and 1999. This was true both across cohorts measured by date of birth, and measured by academic years (see Table 6).

In the case of Roma women in particular, although we systematically underestimate the size of the jump in the probability of starting school under the CSL age scheme from the 2001 Hungarian Census, and, the data is noisy, the 2001 Hungarian Census still allows us to estimate the school starting patterns of those born between 1990 and 1993, in birth month bins. Figure 4 shows that the share of late starters and early starters do not differ significantly across these four years around the June 1 cutoff.

The fourth identification assumption is that the fact whether one reports herself as Roma or not is independent to the CSL age change. 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. In general, the effects of longer schooling on human capital accumulation could impact self-assessment in two ways. If

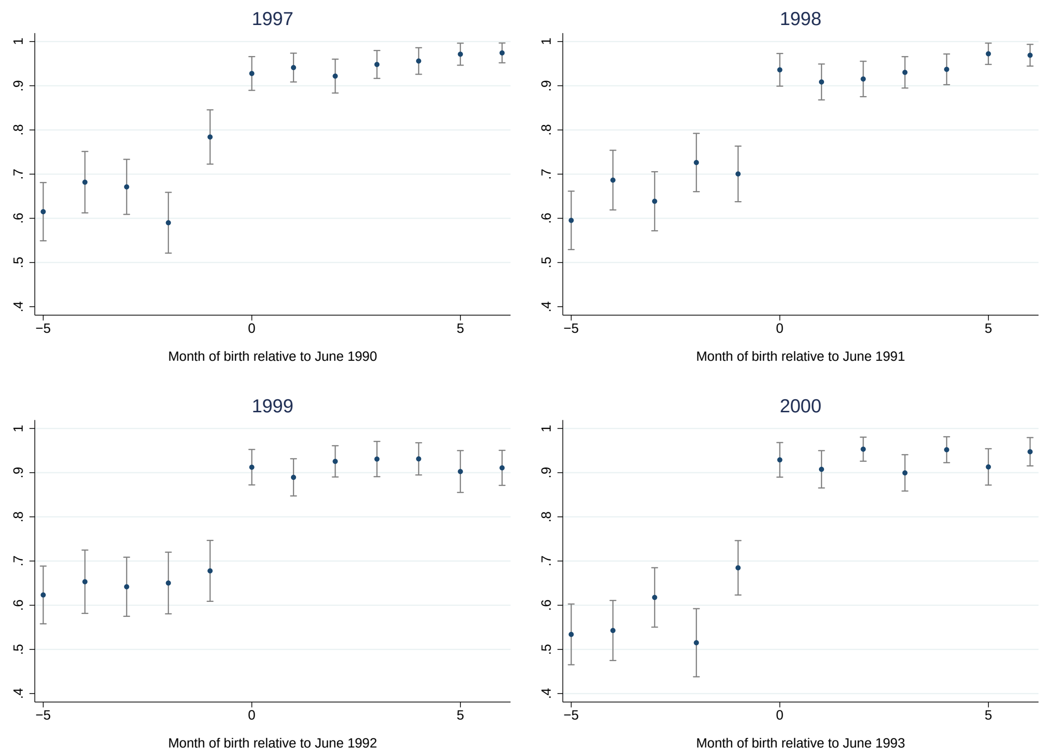
[10] KIR-STAT in Hungarian

Table 6: 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

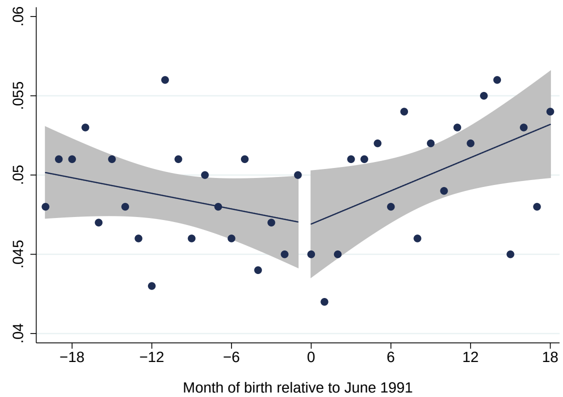
Data Source: Public Education Statistics (PES) of the Public Education Information System (KIR-STAT). "Early starters" refers to those entering elementary school without reaching age 6 by May 31. "Compliers" refers to those entering elementary school according to the age rule of compulsory schooling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

Figure 4: The probability of starting school in 1997-2000, Roma women



The average probability of starting school in 1997-2000, 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: 2,308, 2,313, 2,282, and 2,386.

Figure 5: The share of Roma women by month of birth, 1991



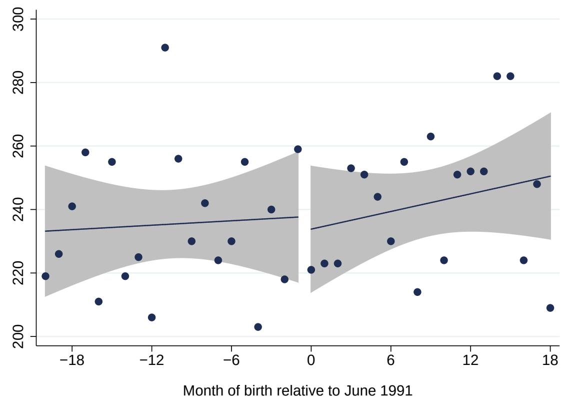
Data source: own estimation from the 2011 Hungarian Census. 0 on the x axis refers to being born in June 1991. Linear regression lines estimated separately below and above the cutoff, plotted with 95% confidence intervals. No. of individual observations: 263,298.

human capital development induces individuals to be more conscious about their identity, those under the new scheme of the higher CSL age could be more likely to reveal their Roma ethnicity, meaning we overestimate the effect of the change. However, this works in an opposite fashion if human capital development induces people to hide their minority status. There is some evidence for such phenomenon 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 to the CSL age change we compare the share of women identifying themselves as Roma below and above the cutoff. Figure 5 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 6).

Fifth, in addition to Roma self-identification being independent from the increased CSL age, we also assume its independence from teenage motherhood. In addition to a legislative change, 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 adolescents of Roma descent are more likely to identify themselves as Roma if their family experienced economic hardship since being Roma is highly 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 legislative change had on Roma women.

Figure 6: 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.

4.2 Empirical methods

Similarly to Adamecz-Völgyi (2018), we estimate the ITT effects of the CSL age legislative change around the June 1, 1991 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
β_{NP} captures the ITT effect of the legislative change 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
β_P captures the ITT effect of the legislative change 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[11]. 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.

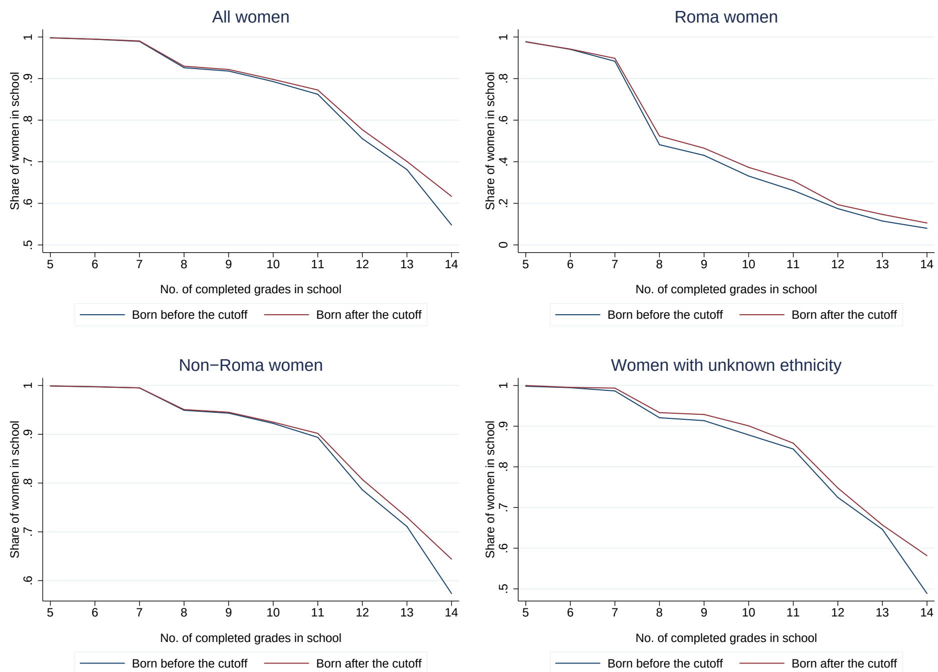
4.3 The number of completed school years around the cutoff

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 7 shows the share of women that completed 1,2,...,13 grades in school. Non-Roma women born right after the cutoff have successfully completed more grades by 2011 than non-Roma women born before the cutoff. This difference is somewhat smaller among Roma women.

We know from the qualitative studies of Mártonfi (2011a,b) that most Roma (and non-Roma) students followed the new CSL age legislation and showed up in school at older ages as required. According to the data, although Roma women born after the cutoff did stay in school longer, they only completed a slightly higher number of grades successfully than before the CSL age increase. This phenomenon is due to the fact that the Hungarian education system has long faced problems with providing quality education for students of various backgrounds (OECD, 2015), and especially for the Roma (Kertesi and Kézdi, 2009). There is a huge variation in the quality of education in Hungary, and Roma students are likely to go to low quality schools and more likely to repeat grades (Kertesi and Kézdi, 2010). This might also be the reason why

[11] Own estimation from the 2011 Census.

Figure 7: The share of women staying in school after completing 1,2,...14 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.

Table 7: Descriptive statistics

Having the first child by age	All women		Roma women		Non-Roma women		Ethnicity unknown	
	Control	Treated	Control	Treated	Control	Treated	Control	Treated
16	0.006	0.006	0.062	0.052	0.003	0.003	0.019	0.016
17	0.016	0.015	0.144	0.135	0.009	0.009	0.030	0.024
18	0.031	0.030	0.259	0.230	0.019	0.019	0.048	0.034
19	0.049	0.047	0.375	0.336	0.031	0.032	0.065	0.051
20	0.068	0.066	0.458	0.428	0.047	0.047	0.092	0.074
No. of obs.	29,275	29,353	1,381	1,394	26,770	26,864	1,124	1,095

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.

among non-Roma women the difference in the number of completed grades appears at Grade 11 (age 17) while among Roma women it appears already after Grade 8 and stays constant.

4.4 Teenage fertility around the cutoff

Table 7 presents the probability that a woman has her first child between ages 16 and 20 for both Roma and non-Roma women. According to this raw data, there is no difference in the prevalence of teenage motherhood between non-Roma women born right before or after the cutoff. However, among Roma women, the probability of beginning motherhood by age 18 and 19 is lower in the group for those women born right after the cutoff.

Figure 8 shows a similar picture by comparing the share of childless women by age among those born right before and after 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.

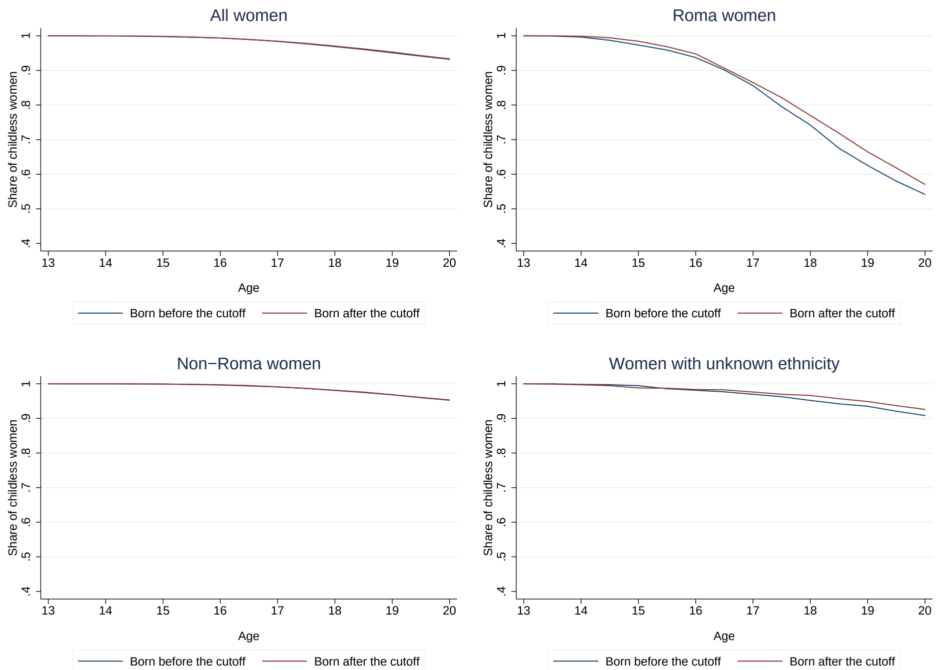
5 Estimation results

5.1 The effects of the CSL age increase on teenage motherhood

Table 8 (and Figure 9) presents the intention-to-treat (ITT) effects of the higher CSL age on the probability of having the first child by ages 16-20. The CSL age increase 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%. However, this effect is temporary and vanishes by age 20.

Our results coincide with the phenomenon also seen in Figure 8. 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 legislative change delayed motherhood by 2 years among Roma adolescents.

Figure 8: The share of childless women by age



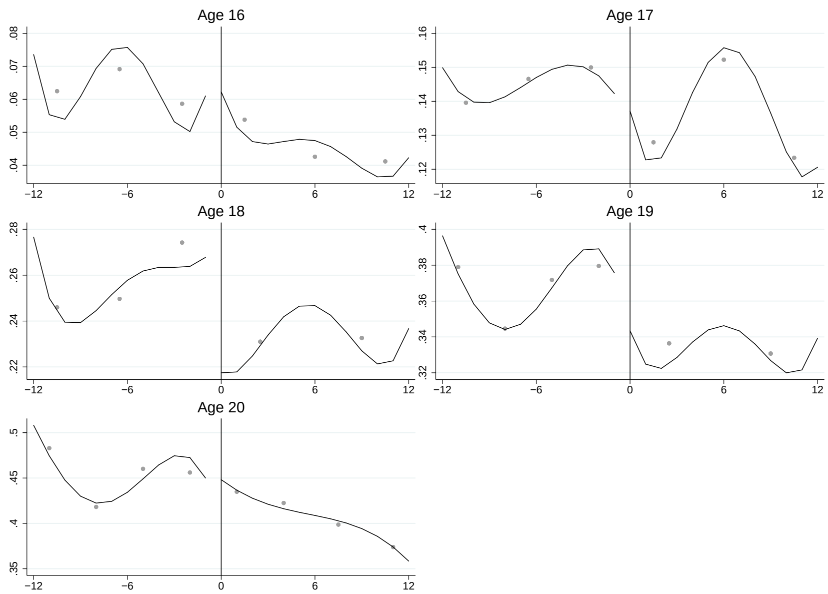
Kaplan-Meier survival functions with respect to giving birth, by age. Age is a continuous variable; giving birth by age 16, for example, means giving birth either before, or exactly on, the mother's 16th birthday. Data source: own estimation from the 2011 Hungarian Census. Born before the cutoff (control 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 8: 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	0.009	0.001	-0.003
	(0.002)	(0.020)	(0.001)	(0.011)
	[50,005]	[2,298]	[51,884]	[2,486]
	{152.2}	{148.2}	{173.6}	{201.3}
17	-0.002	-0.027	0.000	-0.007
	(0.001)	(0.020)	(0.002)	(0.011)
	[60,137]	[3,000]	[53,321]	[2,510]
	{184.3}	{194.8}	{178.4}	{203.0}
18	-0.004	-0.068***	0.000	-0.02**
	(0.002)	(0.022)	(0.003)	(0.008)
	[65,502]	[2,906]	[56,700]	[2,545]
	{201.2}	{188.3}	{190.8}	{206.7}
19	-0.008	-0.067**	-0.001	-0.01
	(0.004)	(0.023)	(0.004)	(0.013)
	[61,070]	[3,015]	[61,000]	[2,100]
	{187.9}	{195.7}	{205.3}	{168.6}
20	-0.003	-0.018	-0.001	-0.044**
	(0.004)	(0.030)	(0.004)	(0.017)
	[51,619]	[2,633]	[44,611]	[2,438]
	{157.2}	{169.8}	{148.0}	{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.

Figure 9: 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 temporary effects on teenage fertility suggest that the impact among Roma women is mainly 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 higher CSL age decreased the probability of getting pregnant during school year only, when adolescents had to be physically in school, and that it had no significant effect on the probability of getting pregnant during summer breaks (Table 9).

Results estimated on the linked data show that the probability of getting pregnant decreased as a result of the legislative change (Table 9, 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 might 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, the probability of giving birth by age 18 decreased by 5.6 percentage points, and the probability of getting pregnant decreased by 6.5 percentage points. Also, the probability of choosing an abortion to end a pregnancy declined by 8.9 percentage points (Block C), although again, this coefficient is not significant either, and it is only estimated on the small sample of Roma women who did get pregnant (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 eariler 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 10 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 legislative change. Table 11 presents our results. The real, 1991 cutoff is the only one producing significant impacts.

Table 9: 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 Time of conceptions defined: Census data: assuming 9-month pregnancies Linked data: based on information on the week of conception	-0.066** (0.019) [2,788]	-0.065** (0.021) [2,327]
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 School years defined: Census data: Sept - May Linked data: 1 Sept - 14 June, excluding 15 Dec - 31 Dec	-0.054** (0.024) [2,788]	-0.054** (0.023) [2,327]
F. Probability of getting pregnant during summer breaks Summer breaks defined: Census data: June - Aug Linked data: 15 June - 31 Aug	-0.011 (0.015) [2,788]	-0.003 (0.009) [2,327]
G. Probability of getting pregnant during Christmas breaks Chritsmas breaks defined: Linked data: 15 Dec - 31 Dec	-	-0.008 (0.007) [2,327]

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), see Table 13.

Table 10: 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	-0.023	-0.051***	-0.001	-0.008	-0.013*
	(0.029)	(0.021)	(0.016)	(0.017)	(0.008)	(0.006)
75%	-0.051*	-0.048***	-0.041*	-0.026	0.006	-0.012
	(0.025)	(0.009)	(0.017)	(0.015)	(0.009)	(0.007)
100%	-0.068***	-0.051***	-0.054**	-0.043**	0.010	-0.007
	(0.022)	(0.012)	(0.018)	(0.017)	(0.009)	(0.006)
	[2,906]	[2,855]	[2,580]	[2,664]	[4,325]	[2,933]
	{188.3}	{221.8}	{191.8}	{205.9}	{329.2}	{226.4}
125%	-0.048**	-0.041**	-0.062***	-0.051**	0.009	-0.008
	(0.022)	(0.014)	(0.018)	(0.019)	(0.009)	(0.006)
150%	-0.050**	-0.034**	-0.045**	-0.038*	0.008	-0.007
	(0.022)	(0.016)	(0.020)	(0.020)	(0.008)	(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 11: 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	-0.012	0.005	-0.007	0.007	0.005
	(0.029)	(0.013)	(0.018)	(0.022)	(0.011)	(0.003)
1991	-0.066**	-0.056***	-0.065**	-0.054**	-0.003	-0.008
	(0.022)	(0.003)	(0.011)	(0.039)	(0.711)	(0.278)
1992	0.056	0.033	-0.005	0.009	-0.020	0.005
	(0.040)	(0.033)	(0.030)	(0.023)	(0.018)	(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.[12]

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 12 shows that the effect of the legislative change in these models is significant and very similar in magnitude to the ones estimated by non-parametric regressions.

6.4 Effects on the probability of pregnancy in two-month intervals

One potential reason for not finding significant effects in school breaks could be that it is "harder" to find effects in shorter periods, where the prevalence of pregnancy is lower, than in longer periods. Thus, in this subsection we look at the effects of the reform on the probability of getting pregnant in two-month intervals (i.e, in Jan-Feb, March-Apr, etc.) on both the Census and the linked data.[13] As Table 13 shows, we see negative effects in all periods, except in July-August. None of these effects are significant though; we are using the same local linear estimation as before and the prevalence of pregnancy is very low (3-4%) in all of

[12]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.

[13]Note that by construction, the probability of getting pregnant in Jan-Feb + the probability of getting pregnant in March-April + ... + the probability of getting pregnant in Nov-Dec=probability of getting pregnant, and this is true for the estimated ITT effects as well.

Table 12: 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***	-0.035**	-0.045**	-0.047***	0.007	-0.004
	(0.019)	(0.015)	(0.018)	(0.015)	(0.011)	(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% .

Table 13: Effects on the probability of pregnancy by age 18 in 2-month intervals (Roma women)

	Census data	Linked data
Jan - Feb	-0.004 (0.010)	-0.015 (0.014)
March - Apr	-0.009 (0.017)	-0.019 (0.009)
May - June	-0.018 (0.009)	-0.020 (0.007)
July - Aug	0.003 (0.006)	0.006 (0.012)
Sept - Oct	-0.009 (0.013)	-0.018 (0.011)
Nov - Dec	-0.028 (0.017)	-0.013 (0.010)
No. of obs	2,788	2,327

Linear probability models estimated by local linear regressions with rectangular kernel using 180-day bandwidths. Outcome variables: getting pregnant in Jan-Feb, March-Apr, etc. 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.

these periods. Still, it is reassuring to see that the estimated coefficients are the smallest in magnitude and non-negative in the summer.

7 Discussion

We find that the increase in compulsory schooling age from 16 to 18 decreased the probability of having the first child by age 18 among Roma women by 6.8 percentage points, or 26%, while did not affect non-Roma women. We find that its effect among Roma women fades away by age 20; thus, we conclude that increasing the CSL age delayed motherhood by about two years. 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. The effect is quite large, which implies that raising the school leaving age can be effective in reducing the incidence of teenage pregnancy even if it does not generate human capital effects. However, as we only observe the relevant cohort until age 20, we can not rule out that there may be further effects later in life. In our interpretation, this paper documents the incapacitation effect of being physically in school that reduces not only the probability of motherhood but also the probability of getting pregnant. We believe that these results have external validity for disadvantaged ethnic minorities living in developed countries.

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

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

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 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., Tárki Social Research Institute, Budapest

Hámori S, Köllő J (2011) Kinek használ az évvesztés? Iskolakezdési kor és tanulói teljesítmények magyarországon. Közgazdasági Szemle (Economic Review - monthly of the Hungarian Academy of Sciences) LVIII(2):133–157

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

Janky B (2005) A cigány nők társadalmi helyzete és termékenysége. 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.

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

Kemény I, Janky B (2003) A cigány nemzetiségi adatokról. Kisebbségkutatás 2003(2):309–315

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 érettségihez. 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

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

Mártonfi G (ed) (2011a) A 18 éves korra emelt tankötelezettség teljesülése és (mellék)hatásai. Oktatáskutató és Fejlesztő Intézet, Budapest

Mártonfi G (ed) (2011b) Hány éves korig tartson a tankötelezettség? Válaszkísérlet egy rossz kérdésre: szakpolitikai javaslat. Oktatáskutató 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. Oktatáskutató 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

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 (2012) Compulsory education and teenage motherhood. Tech. rep., Working paper presented at the Royal Economic Society's 2013 annual conference

A Appendix

Table 14: The balance of the linked data

	Mean, unlinked observations	No. of unlinked observations	Mean, linked observations	No. of linked observations	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 15: The effects of personal characteristics on Vital Statistics database events being linked to the Census

	All events	Live births	Abortions	Miscarriages	Stillbirths
Age	-0.025***	-0.035**	-0.022*	-0.044	0.198
at the event	(0.006)	(0.013)	(0.008)	(0.028)	(0.140)
Educational status at the	0.124***	0.094***	0.188***	0.210*	0.195
event	(0.015)	(0.024)	(0.026)	(0.078)	(0.259)
Settlement	-0.006**	-0.001	-0.008	-0.003	0.011
size	(0.002)	(0.005)	(0.004)	(0.009)	(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 16: 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**	-0.009
	(0.017)	(0.027)
	[3,702]	[18,999]
Live birth only	-0.026	0.032
	(0.060)	(0.048)
	[1,625]	[8,338]
Abortion only	-0.061**	-0.084*
	(0.025)	(0.042)
	[1,759]	[9,020]
Miscarriage only	-0.098	-0.320**
	(0.091)	(0.092)
	[309]	[107]
Stillbirth only	0.113	0.196
	(0.334)	(0.390)
	[21]	[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.