Anna Adamecz-Volgyi – Flora Samu - Agota Scharle[1]

The Effect of a School-leaving Age Rise on the Prevalence of Teenage Pregnancies

November 2015

We examine the intention to treat effect of an increase in the compulsory school-leaving (CSL) age in Hungary on the prevalence of teenage pregnancy in a regression discontinuity design. We find that the rise from age 16 to 18 decreased the probability of first birth-giving by 7.3 percentage points among Roma women. This is a relatively large effect, however, it vanishes by age 20. We support our findings by several robustness checks. It is puzzling that we find no effect on the education outcomes of these women; thus, we conclude that the impact of the legislative change might have gone through the time-use (incarceration) channel. We discuss the potential problems with our empirical design and point out the further directions of this research.

1. Introduction

We examine the effect of raising compulsory school-leaving (CSL) age on the prevalence of teenage pregnancy in Hungary. Through its effect on dropping out early from school, teenage childbearing is one of the most important channels of the intergenerational transmission of poverty (e.g. Bonell, 2004). The literature presents clear evidence on the negative health, social and economic consequences of teenage pregnancy, such as lower academic attainment and labour market attachment rates, and higher rates of benefit receipt and infant mortality (Chevalier and Viitanen 2003, Fletcher and Wolfe 2008, Wilson 2012). Negative impacts have been found to affect the second generation as well: Paniagua and Walker (2012) shows that having a teenage mother decreases the probability of post compulsory education, and the daughters of adolescent mothers are more likely to become teenage mothers themselves, because their social conditions had a higher chance to remain unchanged.

Although international evidence suggests a link between education and teenage pregnancy, the existing empirical results are mixed. The increase of the average education status of the youth and the decrease of early pregnancy happened parallel to each other in all countries, but it is not straightforward to tell which way causality ran: teenagers with more years of schooling are less likely to get pregnant, or, because the prevalence of teenage pregnancy decreased, more teenagers had the possibility to stay in school longer and finish secondary education with maturity exam. Moreover, it complicates the analysis that there may be several other things affecting educational attainment and teenage pregnancy at the same time: for example, disadvantaged adolescents are more likely to drop out of school and have early sexual life at the same time (Blum et al. 2000). The usual solution to identify the causal effects of education is to use policy

[1] The authors thank Dániel Horn, Zoltán Hermann, Gábor Kézdi, Dorottya Szikra, Balázs Váradi, and participants of the MTA-KRTK workshop in Szirák and seminars at the Budapest Institute and CEU for their useful comments and suggestions. The research leading to this paper was supported by Grant No. 112792 of the National Scientific Research Programme (OTKA).

changes (school entry rules: McCrary and Roger, 2006; length of school day: Berthelon and Kruger, 2011; length of schooling: Black et al., 2008; Silles, 2011; Wilson, 2012, Cygan-Rehm and Maeder 2013, Clark, Geruso and Royer 2014). These empirical papers typically use either labour force or household surveys, and they compare pregnancy prevalence in several cohorts from before and after the legislation change.

The conclusions of these papers do not fully converge. While McCrary and Roger (2006) find no significant effect of education on teenage pregnancy, the other papers come to the conclusion that there is a significant effect, however, they do not agree on the mechanisms generating the effect. These authors consider two possible impact mechanisms: an "incarceration" effect, meaning that these teenagers are "incarcerated" in the school for longer so they have no possibility to get engaged in activities leading to early pregnancy; and a human capital effect, i.e. that the human capital of these teenagers develops further in school, and hence they make better decisions, or their preferences change. Black et al. (2008) find very weak evidence of the incarceration effect and somewhat larger effects through human capital development, while Wilson (2012) finds strong effects through both channels. These analyses cover the US, Norway, Mexico, and the UK; to our best knowledge, no evidence has been published regarding the CEE-countries.

There is a wide range of studies regarding early pregnancy in Hungary. One strand of the literature examines the sharp rise of teenage pregnancy in the 1970s. A study was conducted between 1981 and 1993 to explain this unexpected trend. The first representative survey of the study in 1983 found that both the number of adolescent births and abortions skyrocketed at that time (Pongrácz, 1983). The follow-up survey of the young mothers in 1993 concluded that they lagged behind significantly in their educational status with respect to their age group as they dropped out of school. Although during the 1983 survey they had had education plans, with few exceptions, most of them did not finish secondary school and did not gain professional degrees (Pongrácz- S. Molnár, 1994).

The National Longitudinal Child Growth Study supplies evidence on the effect of the mother's age on child development (Szikra, 2010). The study followed more than 6000 mothers and their children starting from 1979. Using these data, Gárdos-Joubert (1990) shows that younger mothers are more prone to early delivery, and infants of mothers under 19 weigh significantly less. Szikra (2010) compares several features of participant mothers below and above 18. She finds that young mothers came from a significantly worse background, were more likely to be Roma, had their first period earlier, and their children did not just weigh less at delivery, but the weight difference was still significant at the age of 4.

Although teenage fertility (and fertility in general) has been decreasing in Hungary on average, this does not apply to some groups of the poor (Szikra, 2010). In particular, in some marginalized Roma communities fertility is either on the rise or it has stabilized at a high level (Durst 2007). Janky (2007) cites data from the 2003 Roma survey of Kemény suggesting that one-third of Roma women give birth to their first child before turning 18. It should be noted however that Roma communities themselves are heterogenous with respect to their fertility patterns (Janky 2005).

Kertesi-Kézdi (2009) report that one-quarter of Roma and 10% of non-Roma early dropouts in the Életpálya surveys mentioned having a child and/or starting a family among their reasons for leaving school. Kertesi and Kézdi (2010) have also documented using the Életpálya data that almost 40% of Roma and about 5% of non-Roma 8-graders drop out of school before finishing secondary school, and these ratios reach 50 and 20 per cents among the children of non-educated parents, respectively. This is important since earning a maturity exam (érettségi) implies a large employment and wage advantage in the Hungarian labour market (Kertesi-Kézdi, 2010).

To measure the causal effect of longer schooling on early pregnancies we exploit a legislative change in Hungary in 1996, which increased school leaving age from 16 to 18 years for those staring elementary school in 1998. Thus, the new regulation intended to affect those born in June 1991 or later, while those born in May 1991 or earlier were left unaffected. Using this definite point in time regarding the introduction of the higher school-leaving age measure, we aim to estimate the intention to treat effect of this policy change on the probability of early pregnancy.

The rest of the paper is structured as follows. Section 2 summarizes the theoretical considerations behind the relationship between schooling and teenage fertility. Section 3 introduces the empirical methodology used by this paper, Section 4 describes the data. Empirical results and robustness checks are presented in Section 5. Section 6 provides additional evidence on the potential channel of the effect, Section 7 concludes and points out the potential problems of this paper and sets out directions for further research.

2. The mechanism – why CSL age matters with respect to teenage pregnancy?

The existing literature explicitly or implicitly relies on human capital theory in which the individual decides about investments into human capital considering life-time returns and preferences for consumption and leisure. Extensions of this basic model may include other sources of utility (the technical term for happiness or satisfaction), such as preferences for having a partner or a child. Black et al (2008) outline a static model of schooling and fertility decisions in which women decide about the desired level of education and fertility at the beginning of their teenage years. If there is no regulation on CSL age, their decisions only depend on their utility function (preferences) and their ability. If there is a CSL age rule in place, this constrains the educational choice for some of the women and makes them to choose stay in school longer. This altered choice of schooling time may be associated with a postponement of fertility. This is referred to as the "incarceration effect": while in school, women do not have the desire, time, or opportunity to have a child.

Within the same model, there is an alternative mechanism that may come into play: an increase in education increases human capital, and thus the expected wage rate, which in turn increases the opportunity cost of having a teen birth.[2] The so called "current human capital" comes from the fact that

[2] There is both income and substitution effect at play, but, as Black at al (2008) note, "the consensus in the female labour supply literature is that substitution effects are more important so more education should reduce teen pregnancy".

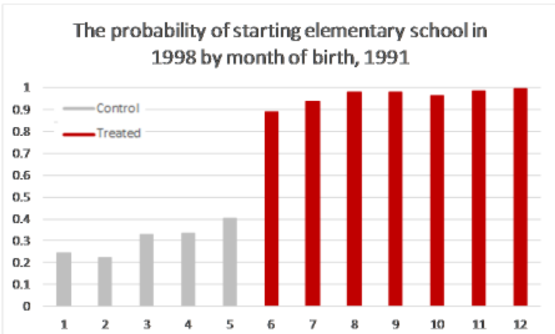
the additional schooling may makes girls smarter and hence decide to postpone childbearing, while the "future human capital" effect works via expectations about the future acquisition of human capital, which are changed by compulsory schooling laws.

3. The methodology and the 1996 Law of Education

We examine the intention to treat effect of a CSL age rise in Hungary on teenage fertility using a regression discontinuity approach. The 1996 Law of Education increased compulsory schooling age from 16 to 18 years. In particular, the new legislation was introduces in a grandfathered manner and came into force in the case of those who started elementary school in Sept 1998. In Hungary, elementary school starting year depends on three factors: date of birth, the results of a school readiness test and parental preferences. Compulsory schooling in this time started in the September of the calendar year in which one reached age 6 by 31 May. Thus, those born before 1 June 1991, referred as cutoff in the rest of the paper, were supposed to start elementary school in 1997 under the old legislation, while those born after 1 June 1991 were supposed to start in 1998 under the new law of 18-year CSL age. We make the usual identification assumption of an RD setup that whether one was born right before or right after 1 June is exogenous (random).

On average, 80 percent of each cohort start elementary school in time, while the rest, those who failed on the readiness test and whose parents decide so, enter only a year later (Hámori and Köllő, 2011). The share of "late starters" is the higher the closer one was born to the cutoff (see Chart 1).

Chart 1: The probability of starting elementary school in 1998



Data source: own estimation from the 10-grade waves of the Hungarian Assessment of Basic Competencies database

Among those born right before the cutoff, this ratio is around 40%. As the probability of treatment, i.e. the probability of starting school in 1998, jumps from about 40% to 90% at the cutoff, and not from 0 to 1, our setup is a fuzzy RD design.

This fact constraints our analysis the following way. As we do not have data about who started elementary school in 1998, we can only estimate the intention to treat (reduced form) effect of the legislation change. We can test the hypothesis of whether the CSL age change had any effects on teenage fertility, but we cannot test the magnitude of this effect. As the probability jump at the cutoff is roughly 50 percentage points, our estimated intention to treat effects would be equal to about a double-size two-stage least square (2SLS) local average treatment effects. In other words, our estimated effects are underestimated.

We estimate the intention to treat effect of the CSL age change by a nonparametric approach using first-order local linear regressions weighted by a rectangular kernel of the following form (Imbens and Lemieaux, 2010):

$$Y_i = a + b_1 * treated_i + b_2 * running_i + b_3 * treated_i * running_i + u_i, \text{ where}$$

Y_i is a binary variable showing whether individual I gave birth by a certain age,

$treated\_i$ is a binary variable indicating whether individual I was born after 1 June 1991,

$running\_i$ is the running (assignment) variable, capturing the time dependent nature of the data and showing the number of days before (-) or after (+) 1 June 1991 individual I was born[3], and

$treated_i * running_i$ is the interaction term of the two previous variables allowing for time dependence to be different on the two sides of the cutoff.

In this setup, the coefficient on $treated_i$, b_1 is going to capture the impact of the legislation change (after controlling for the time dependent nature of the outcome variables). In other words, b_1 is the difference of constant terms estimated by two local linear regressions on the two sides of the cutoff. This local kernel regression method is the standard way in the literature to estimate RD effects because it has good properties in handling the arising boundary problem (Cheng, Fan and Marron, 1997). The method involves the use of a weighting function around the cutoff; however, the shape of this function in practice does not make a difference. (Lee and Lemieaux, 2010) We use a rectangular (uniform) kernel which leads us to the simple functional form explained above.

The choice of bandwidth, i.e. how close we need to go to the cutoff to be "close enough" to believe that those on its two sides really do not differ from each other in a systematic way, is not a trivial question. Relying on the latest RD literature, in this paper we use the optimal bandwidth choice routine of Calonico, Cattaneo and Titiunik (2014), abbreviated as CCT in the rest of the paper. This is the most conservative procedure in the sense that it avoids to get "too large" bandwidths as the earlier routines in the literature and it also applies a novel standard error calculation method that accounts for the remaining potential bias of the estimated coefficient. As a robustness check, we are estimating the effect of the change on several versions of the optimal bandwidth as well.

4. The Data and Descriptive Statistics

We are using the individual-level database of the 2011 Hungarian Census, which includes information on the year, month and day of birth of individuals, their education history and the date of birth of their children. We construct the following outcome variables from the data.

Number of completed school-years: unfortunately, we do not see explicitly exactly when one left school; the Census data only contains the number of successfully completed school-years. For example, if one repeated grades two time in the 8-grade elementary school, thus spent 10 years there, we see only 8 completed school-years in the data.

[3]By time dependence we mean the fact the older one is the more likely that she had already given birth at any point in time.

The age of giving birth for the first time: this variable is constructed based on the date of birth of the mother and the year and month of her first birth.

Probability of giving birth by age 16-20: these variables are constructed based on the date of birth of the mother and the year and month of her first birth.

Table 1: The share of teenage birth-giving in the previous cohorts born in 1988-1990

Probability of Giving Birth by the Age of	Total	Non-Roma Women	Roma Women
16	0.007 (0.000)	0.004 (0.000)	0.066 (0.003)
17	0.017 (0.000)	0.009 (0.000)	0.152 (0.004)
18	0.033 (0.000)	0.020 (0.001)	0.265 (0.005)
19	0.053 (0.001)	0.034 (0.001)	0.385(0.006)
20	0.075 (0.001)	0.051 (0.001)	0.481 (0.006)
No. of obs.	132,227	124,862	7,284

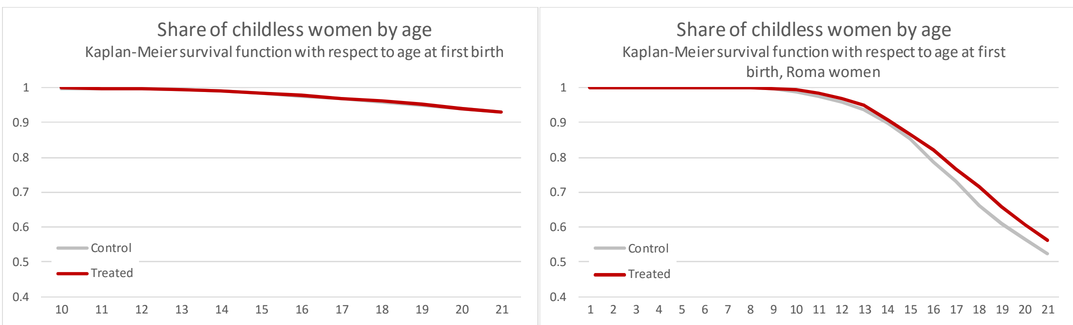
Data source: estimation from the 2011 Hungarian Census data. Standard errors of the group means are in parenthesis.

The Census contains information on the nationality of individuals. The 2011 Census relies on the self-assessment of the respondent to identify Roma ethnicity. The questionnaire asks about primary and other national identity, allowing respondents to identify themselves as Roma *and* Hungarian, and there are two questions on language use as well. Compared to the previous census, this has improved the identification of the Roma population (Kertesi- Kézdi 1998). However, underreporting is still substantial, as the 2011 Census reported only 315,000 self-assessed Roma, which is only about half of the Roma population estimated by demographers (Hablicsek 2007).

4.1 Teenage fertility in the treated and control groups

Chart 1 compares the stock of childless women at every age. On the total sample there is no difference between the survival functions of the control and the treated group while this difference is significant on the subsample of Roma women.

Chart 1: Share of childless women



Data source: own estimation from the 2011 Hungarian Census data. Estimation is done using a 180-day bandwidth. Control: born at most 180 days before 1 June 1991. Treated: born at most 180 after 1 June 1991. Number of observations: 223,341 and 12,479. The difference between the two Kaplan-Meier survival functions estimated on the subsample of Roma women is significant (p=0.0367).

5. Estimation results

5.1 The effect of the CSL age rise on the number of completed years in school

In this section we estimate the effect of the CSL age rise on the number of completed school-years in a survival analysis framework. By survival we mean how long individual i stayed ("survived") in the school system before exit. Unfortunately, in our data we do not see the point in time when individuals left the schooling system; we only see the number of school-years they have successfully finished. Assuming that there should have been many cases in which students dropped out of school without completing the last academic year in which they were enrolled (i.e. they failed), the actual effect of the legislation change on the time spent in school should have been higher than the one we can estimate here. Still, we see a highly significant negative effect of the higher CSL age on the probability of exiting school (see Table 2.) on the sample of all women, however, we do not see any significant effect in the case of Roma girls. Considering that we see the number of completed years but not the real dropping-out time, we assume that the legislation change must have induced even this groups to stay in school longer, however, it probably has not increases their human capital. We are going to show some more results supporting this hypothesis in Section 6.

Table 2: The effect of the legislation change on the probability of finishing school in a survival model

	All Women	Roma Women	Non-Roma Women
Effect of the CSL raise	**-0.165*** (0.000)	0.021 (0.664)	-**0.176*** (0.000)
Control for the number of days before and after 1 June 1991	yes	yes	yes
Control for the interaction of treatement status and the number of days	yes	yes	yes
No. Of obs.	46,088	43,574	2,495

Exponential survival model. Treated: born at most 180 days after 1 June 1991. Control: born at most 180 days before 1 June 1991. Robust p-values, clustering by settlement. Clustering by month of birth gives the same results. The estimated effect of -0.165 means that those under the higher CSL age had 16.5 percent lower probability to exit education, thus attended school longer. Significance: *10%, **5%, ***1%

5.2 The effect of the CSL age rise on teenage fertility

The CSL age rise seems to have a significant negative effect on the probability of giving birth by the age of 18, and this effect is realized on Roma women only. (see Table 3). Considering that the share of those giving birth by the age of 18 is 26.5% in the control group (see Table 1), the estimated 7.3 percentage point effect is quite large.[4] However, the effect vanishes completely by age 20..

Table 3: The effect of the legislation change on the probability of giving birth

Probability of Giving Birth by the Age of	Total		Roma Women		Non-Roma Women	
16	0.000		0.004		0.000	
	(0.976)		*(0.649)*		*(0.866)*	
	Bandwidth	172	Bandwidth	150	Bandwidth	224
	No. of obs	44,154	No. of obs	2,083	No. of obs	53,571
17	-0.003		-0.029		-0.001	
	(0.197)		*(0.287)*		*(0.518)*	
	Bandwidth	194	Bandwidth	195	Bandwidth	214
	No. of obs	49,452	No. of obs	2,700	No. of obs	51,292
18	**-0.007***		**-0.073***		-0.000	
	(0.029)		*(0.043)*		*(0.0989)*	
	Bandwidth	187	Bandwidth	189	Bandwidth	199
	No. of obs	47,123	No. of obs	2,615	No. of obs	47,855
19	-0.007		-0.071		-0.003	
	(0.070)		*(0.053)*		*(0.382)*	
	Bandwidth	157	Bandwidth	191	Bandwidth	171
	No. of obs	40,404	No. of obs	2,633	No. of obs	41,453
20	-0.004		-0.001		-0.023	
	(0.384)		*(0.654)*		*(0.608)*	
	Bandwidth	154	Bandwidth	151	Bandwidth	184
	No. of obs	38,879	No. of obs	2,548	No. of obs	37,496
No. of obs.		223,341		12,479		210,745

First order local linear regressions with rectangular kernel. Optimal bandwidth calculated by the CCT (2014) routine. Heteroscedasticity-robust p-values are in parenthesis in the sense of CCT (2014). This is a hypothesis generating, not testing table, thus p-values are not corrected for multiple testing. Significance: *10%, **5%, ***1%

5.3 The heterogeneity of the effect with respect to social background

In this subsection we examine the effect by splitting the whole sample into five subsamples based on two proxy measures. Our first measure is the share of uneducated adults in 1990 in the settlement where the individuals were born.[5] We interpret these ratios as proxies to parental education. As Table 4 shows, we find significant effect in the highest quintile only.

[4] Earlier estimates of the probability to become a teen mother at age 18 range between -0.003/-0.004 (Wilson 2012, Silles 2011) and -0.0047/-0.0058 (Black et al. 2008).

[5] This data is constructed using the 2011 and 1991 Hungarian Censuses.

Table 4: The heterogeneity of the effect with respect to the share of uneducated at the place of birth in 1990

Probability of giving birth by the age of	Lowest quintile	Second quintile	Third quintile	Fourth quintile	Highest quintile
16	0.001	0.000	0.002	0.001	-0.005
	(0.346)	*(0.999)*	*(0.525)*	*(0.561)*	*(0.378)*
17	-0.003	-0.001	0.004	0.000	**-0.021****
	(0.608)	*(0.668)*	*(0.517)*	*(0.760)*	*(0.017)*
18	-0.008	-0.002	-0.001	0.000	**-0.021*****
	(0.149)	*(0.812)*	*(0.754)*	*(0.977)*	*(0.008)*
19	-0.014	-0.002	-0.008	0.002	-0.021
	(0.078)	*(0.864)*	*(0.440)*	*(0.929)*	*(0.132)*
20	-0.011	0.000	-0.014	0.013	-0.019
	(0.185)	*(0.861)*	*(0.216)*	*(0.304*	*(0.265)*
No. of obs.	44,847	44,126	44,656	44,510	45,202

First order local linear regressions with rectangular kernel. Optimal bandwidth calculated by the CCT (2014) routine. Heteroscedasticity-robust p-values are in parenthesis in the sense of CCT (2014). This is a hypothesis generating, not testing table, thus p-values are not corrected for multiple testing. Significance: *10%, **5%, ***1%

Our second measure is the share of teenage birth-giving on the place of birth in 2001. Unfortunately, we cannot go back to 1990 as the time of giving birth is not available for the whole sample of women in the 1990 Census. The picture is quite mixed according to this measure as it is shown in Table 5. Although we can find some positive coefficients, we estimate the same -2 percentage point effect in the fifth quintile, where the share of teenage pregnancies are the highest, just as in the case of the highest quintile with respect to the share of uneducated adults.

Table 5: The heterogeneity of the effect with respect to the share of teenage pregnancies at the place of birth in 2001

Probability of giving birth by the age of	Lowest quintile	Second quintile	Third quintile	Fourth quintile	Highest quintile
16	0.002	-0.001	-0.001	**0.007****	-0.004
	(0.203)	*(0.520)*	*(0.898)*	*(0.036)*	*(0.413)*
17	-0.001	-0.008	0.000	**0.010***	-0.013
	(0.598)	*(0.157)*	*(0.993)*	*(0.066)*	*(0.103)*
18	-0.003	-0.010	-0.003	0.003	**-0.022***
	(0.627)	*(0.126)*	*(0.656)*	*(0.574)*	*(0.072)*
19	**-0.012***	-0.015	-0.003	0.000	**-0.026***
	(0.083)	*(0.205)*	*(0.715)*	*(0.998)*	*(0.073)*
20	-0.007	-0.011	0.003	0.003	-0.009
	(0.318)	*(0.325)*	*(0.810)*	*(0.609)*	*(0.435)*
No. of obs.					

First order local linear regressions with rectangular kernel. Optimal bandwidth calculated by the CCT (2014) routine. Measured in days as being born before and after 1 June 1991. Heteroscedasticity-robust p-values are in parenthesis in the sense of CCT (2014). This is a hypothesis generating, not testing table, thus p-values are not corrected for multiple testing. Significance: *5%, **1%, ***0.1%

5.4 Robustness checks of the effect on Roma women

In this section we conduct several robustness checks for the effect we estimated on the subsample of Roma women.

5.4.1 Different bandwidth choice

In this subsection we investigate the effect of the legislation change on the sample of Roma women. We do same estimation as before, but now we check whether the choice of the bandwidth alters the earlier results. In particular, we repeat the estimations using the 75 and 1250 per cent of the optimal bandwidth by the CCT routine. As Table 6 shows, our results with respect to the probability of giving birth by the age of 18 do not depend on the bandwidth choice in magnitude, however, by using the 75% bandwidth we lose significance. The reason of this phenomenon can be that, as we explained before, the jump in the probability of the treatment is not 0 to 1, and the closer we are to the cutoff the smaller the jump is. Thus, using larger bandwidths is more realistic in this case as it increases the difference of treatment probabilities.

Table 6: The effect of the legislation change on the probability of giving birth using different bandwidths, Roma women

Probability of Giving Birth by the Age of	Optimal bandwith, 75%		Optimal bandwith, 100%		Optimal bandwith, 125%	
16	0.010		0.004		0.001	
	(0.696)		*(0.827)*		*(0.937)*	
	Bandwidth	112	Bandwidth	150	Bandwidth	187
	No. of obs	1,529	No. of obs	2,083	No. of obs	2,588
17	-0.014		-0.029		-0.031	
	(0.652)		*(0.282)*		*(0.743)*	
	Bandwidth	146	Bandwidth	195	Bandwidth	244
	No. of obs	2,036	No. of obs	2,700	No. of obs	3,397
18	-0.051		**-0.073****		**-0.061****	
	(0.188)		*(0.034)*		*(0.046)*	
	Bandwidth	142	Bandwidth	189	Bandwidth	236
	No. of obs	1,975	No. of obs	2,615	No. of obs	3,284
19	-0.046		**-0.071***		**-0.074****	
	(0.292)		*(0.062)*		*(0.028)*	
	Bandwidth	143	Bandwidth	191	Bandwidth	
	No. of obs	1,988	No. of obs	2,633	No. of obs	
20	0.002		-0.001		-0.034	
	(0.969)		*(0.562)*		*(0.323)*	
	Bandwidth	138	Bandwidth	151	Bandwidth	230
	No. of obs	1,925	No. of obs	2,548	No. of obs	3,198
No. of obs.		12,479		12,479		12,479

First order local linear regressions with rectangular kernel. Optimal bandwidth calculated by the CCT (2014) routine. Measured in days as being born before and after 1 June 1991. Heteroscedasticity-robust conventional p-values are in parenthesis in the sense of CCT (2014). Significance: *10%, **5%, ***1%

5.4.2 Effect of placebo cutoffs in 1984-1992

In this subsection we compare the effects estimated at the legislation change to placebo cutoffs: to cutoffs at 1 June in years 1984-1992 when, to the best of our knowledge, no similar policy change was introduced. Table 8 shows our results, the cutoff in 1991 is the real one caused by the increase of CSL age. Interestingly, we find one more significant cutoff that influenced the prevalence of teenage pregnancies both in the group of all women and in particular in the group of Roma women: between those born right before and right after 1 June 1988. Although the effect of this on Roma women is quite high, 12.6 percentage point, this effect is not robust as it is much smaller and becomes non-significant in case of a different bandwidth choice.

Table 8: The effect of placebo cutoffs in 1984-1992

Cutoff year, 1 June	All Women CCT bandwidth	All Women 189-day bandwidth	Roma Women CCT bandwidth	Roma Women 189-day bandwidth
1984	-0.007 (0.107)	-0.005 (0.192)	-0.049 (0.262)	-0.057 (0.166)
1985	0.000 (0.852)	0.001 (0.702)	-0.013 (0.741)	-0.014 (0.705)
1986	0.003 (0.480)	0.001 (0.756)	-0.018 (0.741)	-0.036 (0.328)
1987	**0.008**** (0.034)	**0.008**** (0.031)	0.062 (0.124)	0.049 (0.198)
1988	**0.008*** (0.055)	0.004 (0.254)	**0.126**** (0.013)	0.023 (0.541)
1989	-0.002 (0.657)	-0.003 (0.294)	-0.008 (0.973)	-0.010 (0.785)
1990	0.003 (0.328)	0.003 (0.334)	0.058 (0.199)	0.043 (0.197)
1991	**-0.007**** (0.029)	**-0.006*** (0.061)	**-0.073**** (0.043)	**-0.073**** (0.029)
1992	0.000 (0.728)	0.012 (0.715)	**0.102*** (0.096)	0.049 (0.135)

First order local linear regressions with rectangular kernel. Number of observations is around 30,000 in ó the case of all women and around 4,000 in the case of Roma women. Heteroscedasticity-robust conventional p-values in the sense of CCT (2014). Significance: *10%, **5%, ***1%

5.4.3 Effect of placebo cutoffs at each month in 1991

Similarly to the previous sections, we also find an unexpectedly significant placebo cutoff between those born right before and right after 1 Feb, 1991. (see Table 7.) This effect, unlike the one we saw in 1989, is robust across different bandwidth choices as well. At this stage of the research, we cannot supply an explanation why the prevalence of teenage pregnancies is higher in the case of those born after 1 Feb 1991. Instead, in the next section we are estimating an extended model where we control for potential yearly and monthly seasonality.

Table 7: The effect of placebo cutoffs in each month of 1991

Cutoff day, born in 1991	All Women		Roma Women	
	CCT bandwidth	189-day bandwidth	CCT bandwidth	189-day bandwidth
January 1	0.000 (0.962)	0.000 (0.904)	0.038 (0.247)	0.031 (0.372)
February 1	**0.007**** (0.026)	**0.008**** (0.016)	**0.059*** (0.079)	**0.072**** (0.032)
March 1	0.002 (0.437)	0.003 (0.388)	-0.003 (0.996)	-0.003 (0.923)
April 1	0.003 (0.309)	0.001 (0.691)	-0.055 (0.117)	-0.051 (0.147)
May 1	-0.004 (0.233)	-0.005 (0.108)	-0.008 (0.921)	-0.020 (0.579)
June 1	**-0.007**** (0.029)	**-0.006*** (0.061)	**0.073**** (0.043)	**-0.073**** (0.029)
July 1	-0.004 (0.424)	-0.004 (0.206)	-0.037 (0.525)	-0.038 (0.256)
Aug 1	-0.004 (0.315)	-0.002 (0.599)	-0.038 (0.438)	-0.002 (0.945)
Sept 1	0.004 (0.315)	0.005 (0.126)	0.049 (0.161)	0.062 (0.054)
Oct 1	0.004 (0.428)	0.005 (0.114)	-0.001 (0.741)	0.030 (0.366)
Nov 1	0.003 (0.608)	0.003 (0.426)	0.042 (0.375)	0.045 (0.170)
Dec 1	-0.004 (0.235)	-0.001 (0.819)	0.016 (0.854)	0.011 (0.747)

First order local linear regressions with rectangular kernel. Number of observations is around 30,000 in the case of all women and around 4,000 in the case of Roma women. Heteroscedasticity-robust conventional p-values in the sense of CCT (2014). Significance: *10%, **5%, ***1%

5.4.4 Controlling for monthly and yearly seasonality

Using the total sample of women born in 1980-1993, we estimate the following 4th-order global polynomial model:

$$Y_i = a + b_1 * treated_i + B_2 * f(treated_i, running_i) + B_3 * controls_i + u_i, \text{ where}$$

Y_i is a binary variable showing whether individual I gave birth by a certain age,

$treated\_i$ is a binary variable indicating whether individual I was born after 1 June 1991,

$running\_i$ is the running (assignment) variable, capturing the time dependent nature of the data and showing the number of days before (-) or after (+) 1 June 1991 individual I was born[6],

$f(treated_i, running_i)$ indicates a 4-th order function of the running variable and its interaction term with the treatment variable, and we also control for the followings:

- year of birth fixed effect
- month of birth fixed effect
- day of the week at birth fixed effect
- settlement type fixed effect
- county fixed effect.

If there were any change in the prevalence of teenage pregnancies in any particular year or month of birth, with this specification we are directly controlling for it, and we estimate the effect that is still remaining. As Table 9 shows, the effect of the legislation change on Roma women is still there, still significant, and very similar in magnitude to the one estimated by non-parametric methods. (-7.3 vs. -6.4 percentage points).

Table 9: The effect of the legislation change in a global polynomial model

Probability of giving birth by the age of	All Women	Roma Women
16	-0.000	-0.008
	(0.726)	*(0.675)*
17	-0.002	-0.028
	(0.287)	*(0.306)*
18	-0.003	**-0.064****
	(0.286)	*(0.037)*
19	-0.001	-0.037
	(0.713)	*(0.281)*
20	0.000	-0.006
	(0.958)	*(0.864)*
No. of obs.	620,560	31,723

Sample: women born in 1980-1993. 4th order global polynomial of the running and running*treatment variable. Additional controls: year of birth fixed effect, month of birth fixed effect, day of the week fixed effect, settlement type, county. Significance: *10%, **5%, ***1%

6 The potential channels of the effect

The impact of education on teenage motherhood may be explained by two potential mechanisms, namely the human capital and the time use channel. According to the latter, the time use or so called "incarceration effect", compulsory schooling reduces the time available to engage in risky behavior and reduces the

[6]By time dependence we mean the fact the older one is the more likely that she had already given birth at any point in time.

likelihood of adolescence pregnancy. The human capital theory states that education affects the fertility decision via better chances in both current and the expected future (Black et al. 2008).

In Section 5.1 we have shown that we the legislation change did not affect the number of successfully completed school-years of Roma women. To investigate this question further, we examine whether it affected any other education outcomes in this group: the probability of gaining a secondary degree. Table 10 shows that the CSL age rise did not increase the probability of gaining a secondary degree among Roma women, nor the level of this degree.

Table 10: The effect of the CSL age rise on the education outcome of Roma women

Education outcomes	Effect on Roma women
Gaining any secondary degree	-0.033 (0.427)
Gaining a lower-tier secondary degree	-0.018 (0.416)
Gaining a higher-tier secondary degree	-0.007 (0.696)

First order local linear regressions with a CCT bandwith and rectangular kernel. Robust p-values in the sense of CCT (2014). Number of observations are around 2,000 in each estimations. Significance: *10%, **5%, ***1%

As we see no effect on the education outcomes, thus the human capital stock of Roma women, at this stage of the research we support the hypothesis that the effect of the legislation change should have gone through the time channel. However, further research is needed in this direction.

7 Discussion , potential problems and the direction of further research

Our analysis finds that the increase of compulsory schooling age from 16 to 18 years by the 1996 Law of Education decreased the prevalence of teenage pregnancies among Roma women by 6-7 percentage points. We make use of the fact that those born right before 1 June 1991 were supposed to start elementary school in the old regime while those born right after started it under the new regime of elevated CSL age; thus, we estimate the intention-to-treat effect of the policy change. We do several robustness checks to support this result. The probability of first birth-giving by the age of 18 is 26.5% in the control group and the CSL age rise decreased this ratio by about a quarter, which is a considerable effect. However, this estimated effect seems to be completely vanished by the age of 20. Thus, the higher CSL age delayed first birth-giving by about 1-2 years.

There are several potential problems to be tackled by further research in order to refine these results. The most important potential shortfall of our empirical strategy could be that the fact whether one declares herself as Roma may depend on the legislative change itself; if so, our results may be biased. As the examination of this question remained out of the scope of this piece of research, we can only hypothesize regarding the direction of any potential bias. If more schooling induce individuals to be more conscious about their identity, those under the new scheme of higher CSL age might have been more likely to reveal their Roma nationality in the Census survey and we might have overestimated the beneficial effect of the change. However, this phenomenon works the other way round if more schooling induces people to hide their minority status. There is some evidence for this in relation to the Afro-American population in the US, called as “acting white” in the literature (Fryer and Torelli (2010). If this is the case, our results are downward biased and the actual beneficial effect of the change could have been even higher. One of the next steps of this research should work on this question.

It is also questionable whether the subsample of Roma girls is the best target group to find an effect on teenage fertility. A better approach may be to find the subsample on which the effect of the CSL age rise is significant on education outcomes in the first place, and then investigate this group further with respect to their fertility decisions. We have tried to do this in Section 5 of this paper already when we split the sample into quintiles by two social background proxy measures. Indeed, we have found significant effects in the most disadvantaged quintiles of the sample on the prevalence of teenage fertility. One direction of further research could be the identification of the appropriate measure for such analysis.

References

Alam, Andaleeb, Javier Baez and Ximena V. Del Carpio (2011). Does cash for school influence Young Women’s behavior in the Longer Term? Evidence from Pakistan, IZA Working Paper 5703.

Berthelon, M. E., and Kruger, D. I. (2011). Risky behavior among youth: Incapacitation effects of school on adolescent motherhood and crime in Chile. *Journal of Public Economics*, 95(1), 41-53.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2008). Staying in the Classroom and out of the maternity ward? The effect of compulsory schooling laws on teenage births. *The Economic Journal*, 118(530), 1025-1054.

Blum RW, Beuhring T, Shew ML, Bearinger LH, Sieving RE, Resnick MD (2000). The effects of race/ethnicity, income, and family structure on adolescent risk behaviors. *Am J Public Health*. 2000 Dec; 90(12):1879-84.

Bonell, C. (2004). Why is teenage pregnancy conceptualized as a social problem? A review of quantitative research from the USA and UK. Culture, health and sexuality, 6(3), 255-272.

Breierova, Lucia and Esther Duflo (2004). The Impact of education on fertility and child mortality: Do fathers matter less than mothers? NBER Working Paper 10503.

Chevalier, A. and Viitanen, T. (2003). The long-run labour market consequences of teenage motherhood in Britain. *Journal of Population Economics*. EPDP, 516. Centre for Economic Performance, London School of Economics and Political Science, London, UK.

Cygan-Rehm, K. and Maeder, M. (2013). The effect of education on fertility: Evidence from a compulsory schooling reform, *Labour Economics*, Volume 25, December 2013, Pages 35–48

David S. Lee & Thomas Lemieux (2010). Regression Discontinuity Designs in Economics, *Journal of Economic Literature*, American Economic Association, vol. 48(2), pages 281-355, June

Fletcher and Wolfe (2008). Education and labor market consequences of teenage childbearing: evidence using the timing of pregnancy outcomes and community fixed effects. NBER working paper number 13847

Fryer, Roland G. and Paul Torelli (2010). An empirical analysis of ‘acting white’, *Journal of Public Economics*, 94(5–6), pp. 380–396

Gárdos, É. and Joubert, K. (1990). Terhesek és csecsemők egészségügyi és demográfiai vizsgálata. A terhes nők összefoglaló adatai. – KSH Népesedésstatisztikai főosztály Termékenység-statisztikai osztály kiadványa. Bp.

Hablicsek, L. (2007): Kísérleti számítások a roma lakosság területi jellemzőinek alakulására és 2021-ig történő előrebecslésére. Demográfia, L. évf. 2007/1: 5–54.

Janky B. (2005): A cigány nők helyzete és termékenysége. In: Nagy Ildikó – Pongrácz Tiborné – Tóth István György (szerk.): *Szerepváltozások. Jelentés a nők és a férfiak helyzetéről*. SZCSEM-Tárki, Budapest, 136-148.

Janky, B. (2007). A korai gyermekvállalást meghatározó tényezők a cigány nők körében. *Demográfia*, 50. évf. 1. szám, 55-73

Jones, E. F., Forrest, J. D., Goldman, N., Henshaw, S. K., Lincoln, R., Rosoff, J. I., Westoff, C. F. and Wulf, D. (1985). Teenage pregnancy in developed countries: determinants and policy implications. Family Planning Perspectives, 17(2), 53-63.

Joubert, K., Darvay, S., Gyenis, Gy., Éltető, Ö., Mag, K., Hof, M. van't, Ágfalvi, R. (2006). Az Országos Longitudinális Gyermeknövekedés-vizsgálat eredményei születéstől 18 éves korig I. (Szerk.: Joubert, K.) — KSH Népességtudományi Kutató Intézetének Kutatási jelentések 83.

Kamarás, F. (2009). Tizenéves terhességek demográfiai jellemzői. KSH, 2009.

Kertesi G. and Kézdi G. (2006). A hátrányos helyzetű és roma fiatalok eljuttatása az érettségihez. (Egy különösen nagy hosszú távú költségvetési nyereséget biztosító befektetés). Megjelent: Budapest Working Papers on the Labour Market, 2006.

Kertesi G. and Kézdi G. (2009). Roma és nem roma fiatalok középiskola továbbtanulása. Megjelent: Fazekas K. (szerk.): Oktatás és foglalkoztatás. MTA Közgazdaságtudományi Intézet, Budapest, 2009. 122-136. old.

Kertesi G. and Kézdi G. (2010). Iskolázatlan szülők gyermekei és roma fiatalok a közép-iskolában. Beszámoló az Educatio Életpálya-felvételének 2006 és 2009 közötti hullámaiból. Társadalmi Riport, 2010. Tárki, Budapest, 2010

McCrary, J., and Royer, H. (2006). The effect of female education on fertility and infant health: Evidence from school entry policies using exact date of birth (No. w12329). National Bureau of Economic Research.

McCrary, Justin and Heather Royer (2011). The Effect of Female Education on Fertility and Infant Health: Evidence from School Entrey Policies Using Exact Date of Birth, *American Economic Review*, 101(1): 158-95

Navarro, M., and Walker, I. (2012). The impact of teenage motherhood on the education and fertility of their children: Evidence for Europe. IZA DP No. 6995 November 2012

Poh Lin Tan (2014). Re-examining the Impact of Education on Teenage Motherhood: Evidence from North Carolina. Working Paper presented at the Population Association of America Annual Meeting 2014, Session 146. Adolescent Sexual and Reproductive Health.

Pongrácz, T. (1983). Fiatalkori terhességek társadalmi, demográfiai vizsgálata. KSH Népességtudományi Kutatóintézetének Kutatási Jelentései 13., Sorozatszerkesztő: Hablicsek László. Budapest, 1983/6.

Pongrácz, T. and S. Molnár, E. (1986). A terhességmegszakításon átesett 18 éven aluli hajadonok társadalmidemográfiai jellemzői. KSH Népességtudományi Kutatóintézetének Közleményei 62., Budapest, 1986/2

Pongrácz, T., and S. Molnár, E. (1994). Serdülőkorban szült anyák társadalmi, demográfiai jellemzőinek longitudinális vizsgálata. Népességtudományi Kutatóintézet Kutatási Jelentései. 53. Budapest, 1994/4

Silles, M. A. (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.

Wilson, T.J. (2012). Compulsory Education and Teenage Motherhood. July 30, 2012. Working paper presented at the Royal Economic Society's 2013 annual conference