University of California, Berkeley

ECON C142 Final Project

May 2019

Article

Pregnancy and its Effects on Education: An Econometric Insight into the Impact of Teenage Pregnancy on College Attendance

Carlos Gonzalez Perez* †

*** Carlos Gonzalez Perez is an Undergraduate Researcher in Economics at University of California at Berkeley under the supervision of Professor D. Card and PhD Candidates P. Munoz and I. Haegele, Department of Economics**

†Email: carlos.gonzalperez@gmail.com **Contact Number: +34671234034**

Abstract

Pregnancy has been the focus of study of many publications in Labor Economics, Sociology and Medical Anthropology. However, the literature has been unable to measure the impact of (teenage) pregnancy on schooling attendance. Throughout this article we consistently prove the significant but limited effect of pregnancy on education, and more concretely the impact that teenage parenting may have on college attendance. These results are robust to heteroskedasticity and model specifications. The (surprising) low magnitude of pregnancy estimators may suggest the Never Taker status of teenage moms regarding college attendance. We implement a wide variety of checks and approaches using European Survey data for testing our hypothesis. Because this article primarily focuses on the research design and methodology, we have adopted an unconventional narrative style which seeks to lead the reader throughout the heretics of the different econometric tools in this publication.

Keywords

Teenage pregnancy, college attendance, Education Economics.

1. INTRODUCTION

There exists a current debate regarding the socioeconomic impact of pregnancy on women. The literature (Barbezat and Hughes, 2005) has largely analyzed the different incentives and compensation schemes that both, men and women, receive for their job; as well as the different incentives which move an employer to choose a male candidate over a female one or vice versa.

However, the impact that physiological differences may have on earlier stages of development, namely the education period, is still largely unidentified by the academia. Throughout the next few paragraphs, we will present compelling evidence suggesting the negative impact of pregnancy on educational outcomes. More concretely, we will show sound evidence proving that teenage pregnancy is a relevant factor explaining college attendance.

Because this paper is especially focused on its research design and it will be appraised according to its methodology, the following (non-conventional) structure will be adopted. Section 1 introduces and explains the motivation of our work; Section 2 reviews the main arguments in literature; Section 3 presents the data, Section 4 minutely describes the Research Design of this project; Section 5 summarizes the main findings; Section 6 broadly debates the limitations of our study and suggests future lines of research; and Section 7 includes a personal reflection and concludes.

Again, because of the special nature of this article, a series of unusual structural decisions have been taken. At first (i) adverse results will be presented, even if they refute the argument developed by the author; (ii) limitations will be included throughout the different sections, highlighting the sharp difficulties I had confronted when going through this data (in order to show my honesty and my work ethics) and (iii) it is possible that too much space in this presentation has been devoted to methodological concerns (including repetitive or bland results). In that sense, I apologize in advance if the reader considers that I provide with too many of those.

A final concern should be noted. Neither Education Economics nor Medical Anthropology are the matters of study of this course. Moreover, the author has a sound knowledge of Education Economics, Labor Economics and Sociology, but the following work does not lay into his usual research scope.

This publication has been specifically produced for the course ECON 142C at University of California at Berkeley.

2. LITERATURE REVIEW

The Economic literature has largely identified gender differences in terms of labor, education or health outcomes. A non-negligible share of these differences has been explained by physiological characteristics and/or the distinct socioeconomic realities shaping males and females. However, it is still unclear which is the marginal contribution of each of these different features, in both biological and sociological terms. In this article we will explore the impact of pregnancy on education, trying to identify its effects on the probability of dropping school after the delivery takes place.

The arguments explaining the causal link between having a baby and dropping school seem intuitive for the broad public. Among them, the author could highlight the following:

(i) Parents have to satisfy extra income needs for raising their offspring, thus they might be tempted to drop school for joining the labor force; (ii) the incompatibility between school academic requirements and pregnancy, child caring and/or household tasks; and

(iii) the social stigmatization of pregnancy (especially during early education stages) which may push the mother out of the formal education system.

Indeed, according to Espin-Andersen (2009), although the gender gap in terms of hours spent in household work has largely been reduced during the last decades, it remains significant. Moreover, even if this gap has been widely eliminated in under 30-year-old cohorts in Western European countries, differences seem to re-emerge after the delivery of the first kid. If this sociological pattern had an impact on the amount of time devoted to other activities such as work, education or leisure, we would expect a different impact of pregnancy on education depending on gender, i.e. we may expect a significantly larger drop in education enrollment for women than for men.

Nonetheless, it is important to note that up to this point we have simply interpreted all previous mechanisms as mechanic responses to an exogenous shock. That is, women might be dropping school during or immediately after pregnancy just because they need to allocate extra time for their children (because of whichever sociocultural factors), or they need to work in order to satisfy the increasing necessities of their households. However, it should also be acknowledged a potential source of endogeneity in these kinds of decisions. Certain individuals at the micro level and/or certain groups or countries at the macro level might be consciously deciding to have their kids at an early age despite knowing it can potentially affect their academic and labor performance. As a result, it is very likely that these individuals would have dropped school even if they had not had any offspring during their teenage years.

Once more, literature is particularly wanting in this aspect. The most prominent theory is the one presented by Caspi et al. (2001). His *survival theory* exposes that those countries with a lower life expectancy (be it higher probabilities of an early and unpredictable death) are the ones with higher rates of teenage pregnancy. According to Caspi, the main objective of any individual or society would be self-conservation, what in biological terms can be interpreted as the transmission of our genetic load to our offspring. Because, statistically, low-income individuals or low-income countries have a higher chance of early death (without transmitting their genes), these communities might be tempted to have their kids as early as possible, in order to ensure conservation. Under this scenario, education is obviously presented as an inferior preference compared to reproduction, and thereby young mothers and fathers in these environments will be willing, in biological terms, to quit their education for guaranteeing the correct development of their offspring in case of pregnancy.

More recent theories, however, as the one developed by Gonzalez and Videgain (2016) based on the previous work of Caldwell (1967), suggest that it is not the survival ratio but the general economic structure of the country what defines the incentives for teenage pregnancy. Some of the potential factors included in this theory are the expected wealth

flow parent-son, the sectorial distribution of the economy and the diffusion of information between government and civil society.

In this article, we will be testing some of the assumptions and prescriptions enunciated by the literature. More concretely, we will ask ourselves whether having a baby increases the probability of dropping school. Intuitively, we may expect a positive answer to this question, but, because of the aforementioned endogeneity concerns, we will see that this result, in fact, might be tricky and difficult to prove. Note that for Caspi, dropping school is perceived as an outcome of pregnancy, while for Gonzalez and Videgain, it is perceived as a factor explaining teenage pregnancy (poorly educated individuals might be more likely to get pregnant during their teenage years because their lack of awareness regarding long-term income reduction due to teenage pregnancy). This study shows that reality might be somewhere in between.

Our study has been developed using European data because of the following considerations. Europe has experienced the most dramatic decrease in the fertility ratio (Dudley and Bouvier, 2010), dipping way below the replacement ratio and thereby, threatening the demographic sustainability of the continent. Consequently, we may assume that the economic, social and/or cultural cost of having a baby in Europe is higher than in any other region of the world. Moreover, because we are interested in understanding the impact of pregnancy on education outcomes, it is desirable to select a location where compulsory education is pervasive and post-secondary education is widely available to everybody. Furthermore, European Demographic Crisis and its potentially devastating effects on future public pension systems prove the relevance of the question at issue. In that sense, to deeply understand the (dis)incentives of (early) pregnancy (as might be the reduction of years of education) is crucial for the future demographic policy of the continent. This article seeks to reduce the existing literature gap by showing the empirical effects of young age pregnancy on education in Europe.

3. PRESENTATION AND ANALYSIS OF THE DATA

In order to develop our study, we will be using biennial data collected by European Social Survey (ESS). ESS is a socioeconomic questionnaire which provides with cross-sectional data for a total of 23 European countries since 2001. The main advantages of this dataset are its free availability and the inclusion of a large variety of variables (around six hundred different items) in a very complete spectrum of topics (ESS, 2019).

Nevertheless, its limitations are more than obvious. First, it should be pointed out that we are dealing with survey data, which, almost by definition, is going to provide less reliable results than administrative data for objective variables such as participants' level of income or birthdate. Second, it should be noted that ESS does not contain panel data but cross-sectional data. This circumstance further complicates inter-temporal comparison and prevents us from using some quasi-experimental approaches like Difference-in- Difference. A last consideration must be highlighted. This Survey does not contain crucial variables such as earnings. It is always tempting to include a proxy for income in our regression analysis, but it will not be possible in this case. The closest proxies for this kind of variables are items related to subjective considerations about income such as *"in which income distribution decile would you place yourself?"* Based on the author's criteria, these indicators are so problematic that they would be of little use for our analyses.

Although some temporal descriptives have been produced, the scope of this study mostly focuses on the most recent wave (ESS 8, 2016). For this exact year, our database includes a complete set 44,387 observations with a total of 534 variables.

3.1 Presentation of Variables

In this section we characterize the distribution of our variables and we present some cross- country and cross-time summaries of the different indicators. Although it is not the intention of this article to analyze the regional differences across Europe, three evident results pop up. At first, it should be noted that the average parenting age is substantially high. The weighted average age for this sample (using countries and gender as weights) is 28.58 years. Secondly, we can observe pregnancy age differences between countries. Statistically, whereas countries in Northern Europe present the oldest ages for parenting, Central and Eastern European countries present the youngest. Southern European individuals place themselves somewhere between this two. Finally, it should be underlined that there exists a significant parenting age gap between men and women. While for men the cross-country weighted average is 32.08 years, for European women it is only 27.81 years. A raw total difference of 4.29 years.

	Country	Average age when 1st kid born
1	CZ	25.9091
2	HU	26.2184
3	LT	26.9130
4	RU	27.0152
5	EE	27.2695
6	GB	27.2739
7	PL	27.7576
8	AT	28.0638
9	IE	28.1474
10	SI	28.5556
11	BE	29.2500
12	ES	29.2870
13	IL	29.3893
14	CH	29.7385
15	FR	29.7389
16	PT	29.7500
17	IT	29.9630
18	NL	29.9756
19	NO	30.1711
20	SE	30.3108
21	IS	30.5870
22	FI	30.6753
23	DE	30.9143

Table 1. Mean age for parenting across country

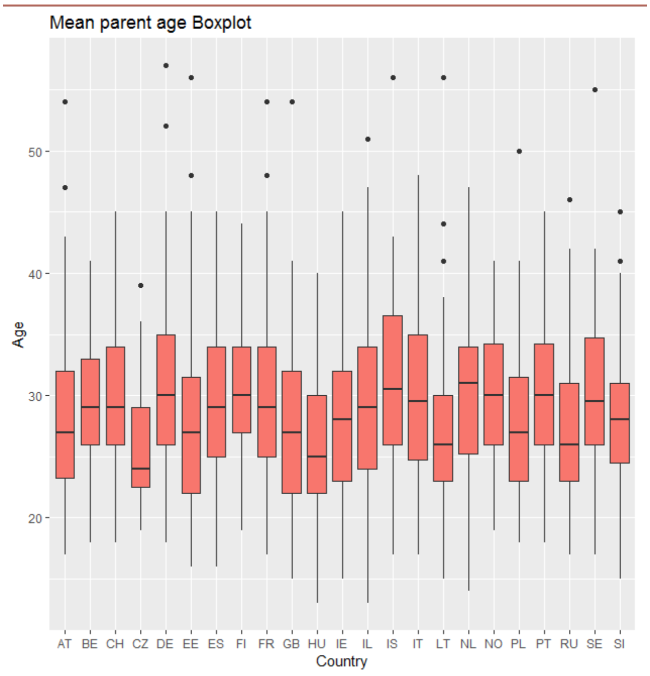


Figure 1. Parenting age distributions across country

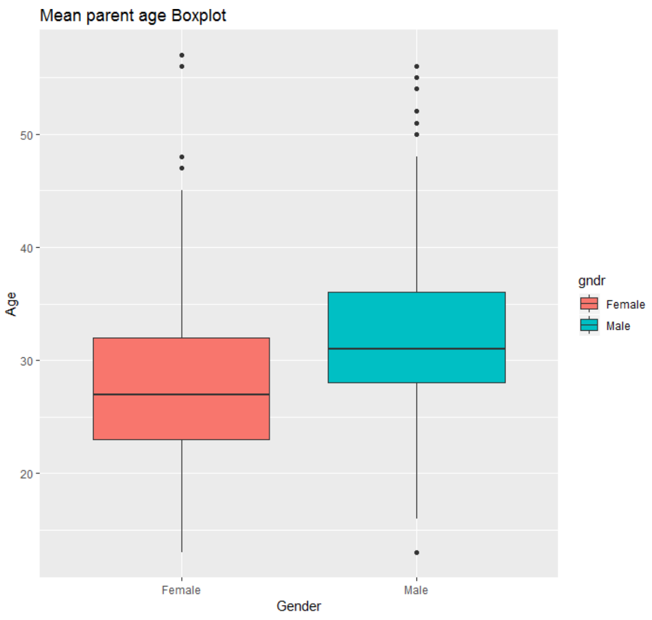


Figure 2. Cross-gender parenting age distributions

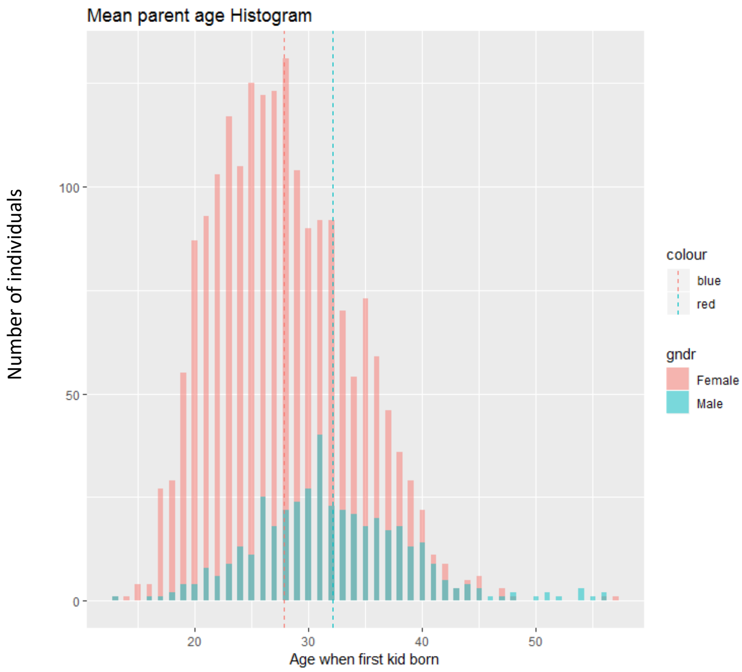


Figure 3. Cross-gender parenting age complete distributions

Eventually, to conclude this data characterization section we show the time evolution of the average age when parenting for first time for waves 2002, 2008 and 2016.

	Male	Female
2002	32.2143	27.3736
2008	29.0566	27.3574
2016	32.0881	27.8185

Table 2. Time evolution of average age when parenting for first time (2002-2016)

As we see, and even if only three waves are being considered, there is no clear time trend on data. This result can be explained by two reasons. First, indeed a stationary ergodic equilibrium has been reached or two (most likely) the time selection is so narrow (15 years) that it is in fact impossible to identify variations in time trends. It should be noted though the sharp decrease in parenting age (especially for men) around 2008 (year of maximum economic expansion before the arrival of the Global Financial Crisis). We suggest as a possible future research line to analyze the (teenage) pregnancy dynamics across time and its implications on education, as well as the possible correlation between parenting age and economic performance. Although some time series intuitions are presented, time evolution is not discussed in depth in this article.

4. RESEARCH DESIGN: Identifying the impact of pregnancy on years of education

Throughout the next few paragraphs, we will try to isolate the effect of pregnancy on education, that is, we will try to avoid the identification problem described in Section 1. To do so, we will use a wide range of econometric approaches. Although it is not standard practice in Labor Economics, we have adopted a semi-narrative structure for this section which will allow the reader to use this paper in a heretic way guiding his/her future research.

One of the results that inspired this article was the positive correlation between education years and the age when parenting for first time.

	Dependent variable:	
	Years of education (women)	Years of education (men)
	(1)	(2)
Age when parents	0.043*	0.067**
	(0.026)	(0.030)
Constant	12.287***	11.673***
	(0.752)	(0.846)
Observations	2,354	1,934
R^2	0.001	0.003
Adjusted R^2	0.001	0.002
Residual Std. Error	7.973 (df = 2352)	7.898 (df = 1932)
F Statistic	2.873* (df = 1; 2352)	5.088** (df = 1; 1932)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 3. Regression of years of education on Age when parenting.

Obviously huge limitations arise from this benchmark model, but possibly the biggest concerns have to do with the well-known "true model" vs "regression model" considerations. Although this result is of little transcendence, we consider it to be foundational to our study, as it pushed us for further analyzing this phenomenon. Note that, at this stage, reverse causality (among other endogeneity concerns) constitutes abig barrier to solve. That is, it might be the case that parenting is not affecting years of education, but years of education is explaining parenting age. In other words, as suggested by Gonzalez and Videgain (2016), it is possible that highly educated people consciously postpone their parenting age because they know in advance the socioeconomic implications of getting pregnant at a younger age. Further considerations regarding reverse causality will be discussed in Section 4.2.

4.1 Regression Discontinuity (RD)

After this heuristic introduction, we proceeded to rigorously confront the research question. Possibly, the most standard approach in Labor Economics would have been to use a Regression Discontinuity (RD) methodology. The underlying economic intuition suggests that when normalizing the age of the respondents (note that only those who have kids will be incorporated into this subsample) to the age when they had their first kid, we expect a sharp decrease in the probability of being enrolled in the education system. As a result, the data should present a discontinuity around the normalized age 0 (=age when the infant was born).

More formally, let us assume a cumulative expectation function (CEF) such that $E[Y_i|x]=E[Y_{0i}|x]$

when x<c (c ∈ R) and $E[Y_i|x]=E[Y_{1i}|x]$ when x≥c. Using this notation, we can define a β_1 such that $\beta_1= E[Y_{1i}|x]- E[Y_{0i}|x]$, and a simple linear model such that

$$Y_i=\alpha_0 + \alpha_1 x_i + \beta_1 D_i + \beta_2 X_i + \beta_{ki} W_{ki} + u_i \quad \text{(i)}$$

Where Di is a dummy variable which equals 1 if the normalized age ≥ 0 (=c), X_i is equal to the normalized age, and W_{ki} a k-row vector controlling for different individual factors. The expected effect of Di is referred in the literature as local average treatment effect (LATE), given that only under very restrictive assumptions we could interpret it as an average treatment effect (ATE). Note that this model only holds under the strong continuity assumption for both CEFs,

$$\lim_{\varepsilon\to 0} E(Yi\ |Di = j, Xi = c - \varepsilon) = \lim_{\varepsilon\to 0} E(Yi|Di = j, Xi = c + \varepsilon)\ for\ j = 0,1$$

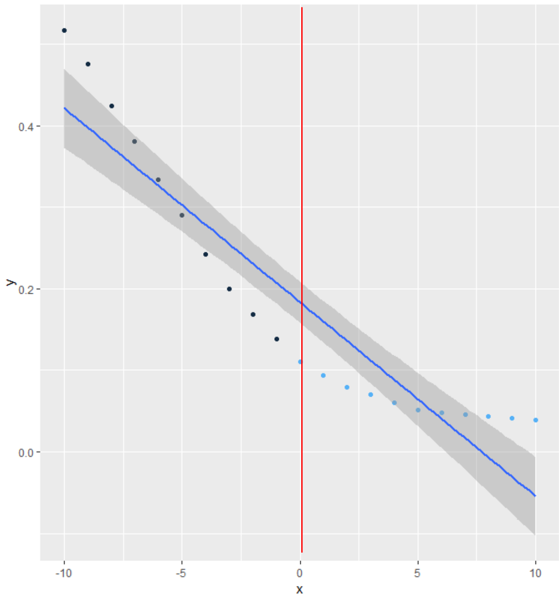
Because CEFs are usually unobserved, the literature usually accepts this design (Angrist and Pischke, 2009) as long as,

(i) Pre-determined characteristics (W_{ki}) evolve smoothly around the benchmark c and,

(ii) The density of the assignment variable Xi does not change discontinuously around the cutoff.

That is to say that selection of parameter c (in this case, having a baby) is considered to be as if random. This assumption is particularly problematic in this context as it does not necessarily hold even under exogenous interpretations of pregnancy (given that families would still have around nine months to accommodate for the upcoming event. Moreover, in an endogenous framework, where parents consciously decide to have a baby, we would most likely expect a non-random variation of W_{ki} before and after year c.

The following scatter plots are presented as a source of evidence. Note that these figures are computed using the mothers' sample (gender=female, having a child = TRUE) for year 2016. They both represent the normalized age (being c=0, the age at which the baby was born) in the x axis and the probability and log probability respectively of "currently at school" for each Xi year. The main (obvious) result is a logarithmic decrease in the probability of attending to school. This pattern can be entirely explained by a time effect. The second (disappointing) result shows no evidence of an extra decrease in the probability of studying when your kid is born. At best, it could be "optimistically" argued that we see the largest marginal decrease in the environment [-1,1], but this result is far from being statistically different from the rest of marginal decays.



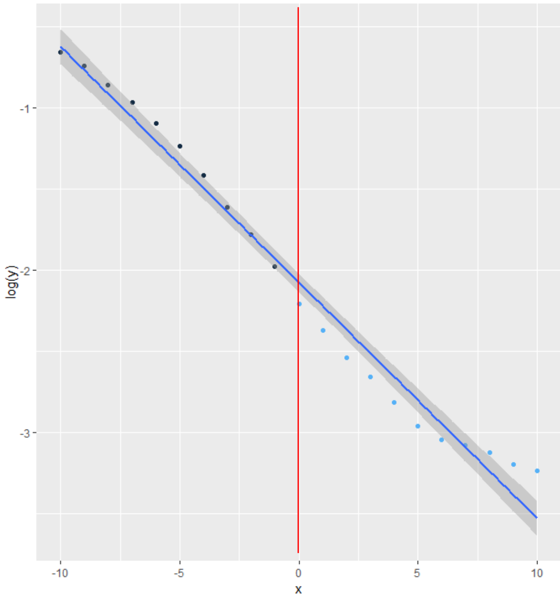


Figure 4. Linear scatter plot: Probability of "currently enrolled in education" across time (left)

Figure 5. Logarithmic scatter plot (right)

In fact, if we run a regression of equation (i) (even with $\beta k=0$) we get that the Treatment effect (having a baby) is far from significant. At best, it could be argued that it is negative rather than positive but still far from conclusive. Note that in this model variable Xi acts as a time control variable.

	Dependent variable:
	Probability of currently studying
Time control	-0.022^{***}
	(0.004)
Dummie	-0.020
	(0.052)
Constant	0.194^{***}
	(0.030)
Log Likelihood	29.944
Akaike Inf. Crit.	-53.889
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4. Model (i) regression table

Severe limitations arise from this methodology. The first one, is that having a baby (as derived from the literature) is far from being random; indeed, in tertiary post-industrial societies there exists a quasi-perfect control of natality as suggested by Jain and Ross, 2012. As a result, we should not expect a sharp decrease in the probability of studying when the kid is born, but a progressive anticipated leave in order to accommodate for the couple's intention to get pregnant.

The other main limitation is not methodological but rather related to the nature of the observed population. As presented in Section 3, European women (and men) tend to have their offspring relatively late (around their late twenties or early thirties). At this point in time, pregnancy does not interfere anymore with education as the schooling period is mostly done for the average European mother. Consequently, we also expect a small marginal contribution of pregnancy to the education outcomes of the overall population given the invariability of education (which equals zero for the vast majority of individuals) at that age.

Given our failure in obtaining any valuable insight from our fuzzy RD research design, and for (partially) overcoming both limitations, we decided to restrict the focus of our study to teenage mothers. We define teenage mothers as those moms who were 20 years old or less when having their first child. The analysis of teenage moms presents a key advantage compared to the general population scenario, as teenagers can indeed be interpreted in terms of compliance as "potential compliers." That is, teenagers are the ones who can indeed change their potential educational outcome.

$$\Pr(C|\text{mother at} >30) \approx 0 \text{ while } \Pr(C|\text{mother at} <20) = k \quad (\text{with } k \neq 0)$$

As a contraposition, note that 40-year-old mothers could be considered in terms of compliance as always takers (AT) as regardless their treatment status (having a baby or not) they do always show an invariable potential outcome $Y_{0i}=Y_{1i}=1$ (not studying). Note that this intuition does NOT assume by any means that all teenagers are compliers (in fact, identifying the characteristics of the complier population is the aim of our study). As a result, *a priori* we do NOT assume (neither expect) that all teenagers who have a baby to drop school.

Based on these considerations, we decided to shift our Research Question (limiting its scope) to: Does teenage pregnancy reduce the probability of attending to college? Before starting with the analysis, we provide some stylized facts about cross-sectional teenage pregnancy in Europe.

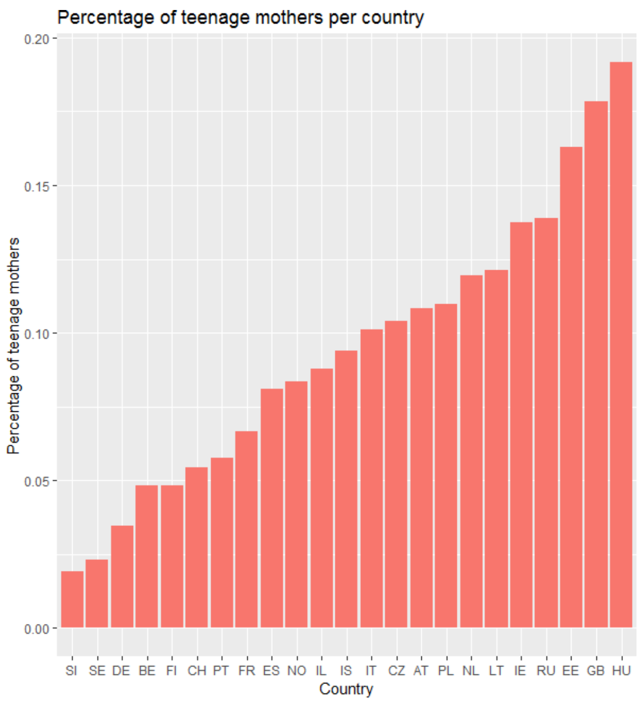


Figure 6. Share of teenage mothers across Europe

Before analyzing Figure 7, we should acknowledge the limitations of survey data. Because we do not know how these surveys take place, it might be true that for unknown reasons, teenage mothers are overrepresented in some countries. However, we will trust the within country representativeness of the sample. As we see, there exist substantial differences across Europe. From a residual 2% in Slovenia and Sweden to almost a 20% in Hungary. Nevertheless, this data should be interpreted carefully as only 1,923 mothers are being observed. As a result, the aggregate country panel can be tricky and unrepresentative. Moreover, unlike in Figure 1, it is difficult to identify groups of countries or subregions from the data.

A first step would be to repeat the RD methodology including only the teenage mothers' sample. However, we unfortunately obtain a very similar result to the one obtained when accounting for the whole set of mothers. Similar limitations hold, suggesting that pregnancy is not a random factor and consequently women leave school anticipating pregnancy.

In fact, this behavior is standard practice among certain collectives. Consider for instance the gypsy communities of Spain or France, which are characterized by very young age pregnancies. In these groups, couples usually marry before intercourse. When marriage takes place, both, wife and husband, usually leave school for taking care of the housework and joining the labor force, respectively. However, it will not be until (at least) nine months after marrying that pregnancy takes place. As we see, even for these communities where teenage pregnancy is quite common, birth is not random at all, as they consciously decide to drop school before the event takes place. It could be argued then, that teenage pregnancy could be quasi-random for other ethnic or income groups such as traditionally privileged middle- and high-income whites. Nonetheless, this dataset does not include proxies for income nor race, thus this argumentation will remain purely speculative.

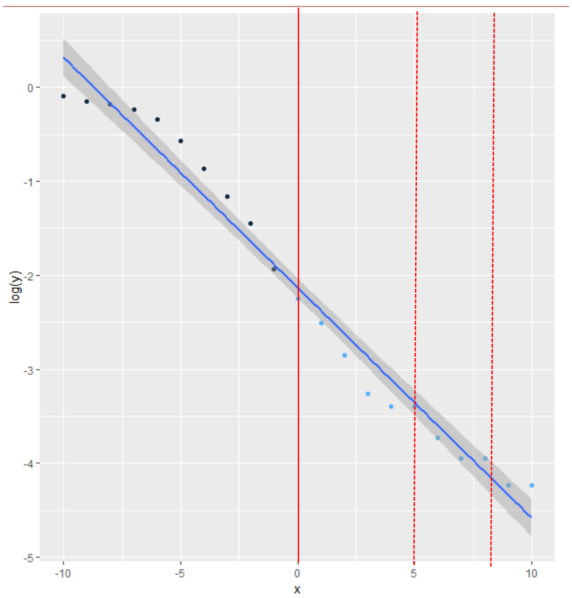


Figure 7. Scatter plot. Logarithmic probability of "currently studying" across time. Teenagers subset.

As matter of fact, it is interesting to note that a small discontinuity takes place around five and eight years after pregnancy. Given that the average parenting age for teenage mothers is equal to 18.83 years, this decay could be explained by the end of college education and the age of post-college (masters) education. Even if at this point, data has not supported any of our intuitions, this is good indirect proof of its representativeness and its robustness.

4.2 Regression Analysis

Given the current limitations of our Research Design, we decided to explore another appealing line of research. It might be that pregnancy does not immediately drop you out of school but has an impact in the decision of enrolling into subsequent education levels. The narrative goes as follows: Let us consider certain teenage mom m_i whogives birth at the age of seventeen. It would not be too surprising that the family and relatives of the respondent take care of the baby supporting the academic career of the young mom during her final year in high school. However, after finishing compulsory education, the decision of attending to college gets dismissed because of time allocation and/or income concerns. According to this narrative, we should expect a sharp decrease on the probability of going to college when having a baby before twenty.

We formally proceed then, to test the following Research Question: Does having a baby before twenty (teenage pregnancy) reduces the probability of attending to college? A series of models and controls will be presented throughout the upcoming paragraphs. The previously identified advantages associated with limiting the sample to teenagers notwithstanding, there exists one more advantage. Because up to 15-16 year-old education is compulsory in most European countries (thus they cannot drop school before this age) and the decision of pursuing college education (in most of the cases) takes place during the pregnancy or after giving birth, we highly limit the possibilities of reverse causality or simultaneity, as temporary causality will prevent this effect from happening.

Regarding the explanatory factors, and considering the ESS database limitations described in Section 3, the following variables and controls have been introduced:

1. Mother's and father's education (defined by the dummy variable college_mom, college_dad=1 if TRUE). According to literature (Christensen et al., 1975), parent's education acts as a good proxy for the general socioeconomic situation, as well as for learning environment and learning expectations.
2. A full set of country dummies (cntry_i =1 if TRUE). Note that cntry_24 (Belgium) has been removed from the analysis to avoid multicollinearity. This set of variables seeks to control for country fixed effects as well as to fit the mean of the regression for each of the means of the countries.
3. Dummie variable domicil =1 if the respondent lives in an urban area.
4. A series of labor situation controls. They include a full set of dummies forretired status,

disabled, full-time housework, unemployment status and ever in paid job.

5. A series of dummies controlling for the type of work. These two sets account for the relationship with employers (self-employed, employee, employer) and type of job (dummies for sectorial distribution).

As we see, a series of controls in very different fields have been introduced. The goal is to isolate the contribution of having a child before 20 (son.bef20). Let us note that the real contribution of this variable will never be properly identified through multivariable analysis, but given the large set of controls, it is not unreasonable to believe that regression estimators could be close to their real value. (Frisch-Waugh interpretation.)

A good alternative to the aforementioned approach is the use of Instrumental Variables. There exists a prominent literature concerning the use of IVs for predicting teenage pregnancy (Klepinger et al., 1999) but none of them are available in this dataset.

4.2.1 Analysis of results

In order to gain robustness, we have tested our data using a large variety of models: OLS (ii), Logit (iii), Lasso-Gaussian (iv), Lasso-Logit (iv), Ridge (v) and a set of nine intermediate estimators with alpha = (0.1, 0.2, …, 0.9; vi, vii, […], xv). **The variable son.bef20 is significant at 0.01 for the whole set models**. Despite its reduced marginal contribution to overall college attendance prediction (previous methods already proved that), it is significant and robust to heteroskedasticity. In fact, we can proudly state that the sign of these estimators is negative, meaning that having a kid before 20 statistically reduces the probability of attending to college. In spite of being an exciting statement, it should be interpreted cautiously as, so far, it can only be shown that on average, within our sample, people getting pregnant before 20 show a lower probability of attending to college, but this probability may or may not be explained by pregnancy itself, this is the true model versus regression model interpretations of the regression analysis. Because of the large size of our tables, these have been placed in Annex 1.

Before implementing a wider variety of robustness checks, we have interpreted the key results derived from our set of models. Note that the aim of this article is not to predict overall college attendance, but to identify the contribution of early pregnancy in college attendance decisions, thus, traditionally valuable indicators as MSRE or R^2 have little interest for this study. Consequently, standard econometric procedures as Cross- Validation have not been implemented either. Some interesting results from models (ii) to (xv) are

(i) The model which best predicts **in sample** college attendance is the Lasso- Gaussian model,

(ii) Both Lasso-Gaussian and Lasso-Logit exclude a large number of variables (11 variables, a 23.4 % of the variables get excluded). In any model, son.bef20 is removed,

(iii) Ridge and hybrid models (which automatically vary the contribution of the indicators) tend to increase the contribution of son.bef20, what implies that teenage pregnancy is a robust predictor of college attendance,

(iv) As presented before, the son.bef20 estimator is negative and statistically different from zero in all fourteen models.

4.2.2 Testing the robustness of our indicator

In this section we present extra robustness checks for our variable of interest. Unsurprisingly, we firstly implement a heteroskedasticity test. We have selected OLS for the subsequent robustness checks for a matter of simplicity given that all previous models yielded similar results. **Our estimators for son.bef20 are robust under heteroskedastic errors** (using R sandwich package). Although only three variables are presented in the summary Table 5, results included the whole set of variables described in Section 4.2.

	Dependent variable:
	college
son.bef20	−0.016***
	(0.0003)
college.mom	0.244***
	(0.035)
college.dad	0.233***
	(0.030)
Observations	1,934
R^2	0.275
Adjusted R^2	0.257
Residual Std. Error	0.342 (df = 1888)
F Statistic	15.895*** (df = 45; 1888)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5. Regression table under heteroskedasticity assumptions

Other alternative method for testing the robustness of our results is to conduct a simple difference in means test, which compares average college attendance between mothers who gave birth before twenty and those who did not. This approach is particularly suitable for our analysis because of its clarity and its fit to in-sample moments.

$$t = \frac{\mu 1 - \mu 0}{\sqrt{\frac{s1}{n1} - \frac{s0}{n0}}} = \frac{0.207416 - 0.1009615}{\sqrt{\frac{0.4055735}{1726} - \frac{0.3020046}{208}}} = 2.59188 > 1.96$$

Where t is the parameter of interest, $\mu_1$is the probability of attending to college when son.bef20=0 and $\mu_0$is the probability of attending to college when son.bef20=1. Note that according to this raw subtraction, pregnancy would be reducing college attendance by 48.67 points (!!) (from 0.2 to 0.1.) Because of our previous multivariable analysis, we know that, in fact, the contribution of pregnancy is much more limited, but still it is good to show that our statistic is significant (at 5 percent confidence level) under different model specifications and approaches.

Because of well-known identification problems it will be interesting to disentangle the effects of the different variables explaining absolute differences in means across the two groups (Treatment=son.bef20 and Control=son.aft20). It could be for instance, that those women who got pregnant before twenty sharply differ in the learning environment from the ones who did not. If that was the case, differences might have not been driven by pregnancy itself but by unobserved factors (omitted variable bias). Just as an illustration, the difference in means across parents' education has been computed for teenage and non- teenage progenitors.

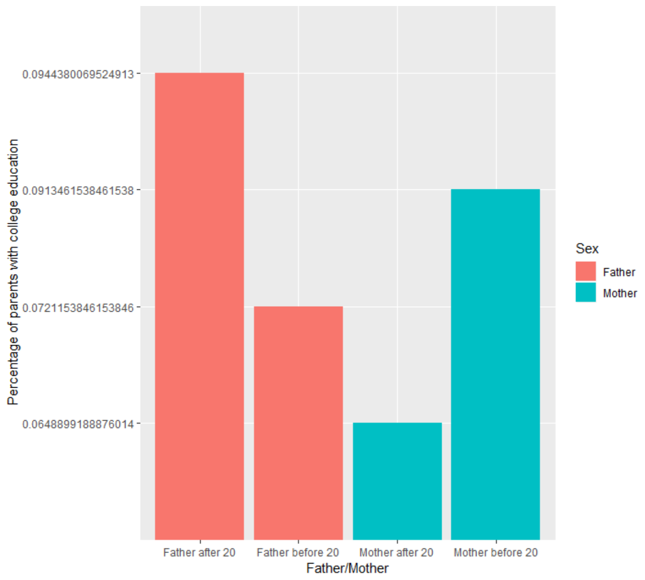


Figure 8. Difference in means of parents' college education across son.bef20

As we see, our intuitions were correct and there exist substantial differences in the characteristics of both groups. Note that surprisingly (as it will be expected that women getting pregnant before 20 years old come from poorer educative environments), mother education of those getting pregnant before 20 is substantially higher than the one from mothers whose daughters have a kid after 20. For solving this composition problem, we have developed a more complex difference in means equation, where differences can be decomposed into a set of observed differences across a set of characteristics j and a not- explained component $(\hat{\beta}_{k+1})$, which, by definition, is identical to the one obtained through the regression model (ii).

$$\overline{y}^b - \overline{y}^a = \sum_{j=2}^{K} (\overline{x}_j^b - \overline{x}_j^a)\widehat{\beta}_j + \widehat{\beta}_{K+1} \qquad \text{(xv)}$$

Even if the results yielded by this approach are quite dull and repetitive, this specification posts a crucial question: What if the returns (coefficients) of our X_{ki} variables differed across groups too? That is, it is not only that teenage progenitors have a different set of endowments but that the contribution of these endowments in terms of college attendance differs across groups too. This question is far from obvious and it underlines the possibility of different marginal contributions of X_{ki}. For tackling this problem, we have adopted an *Oaxaca decomposition* approach. We have conducted three different Oaxaca specifications. The first one assumes that teenage mothers have the same distribution across countries as old mothers. The second one assumes that the mothers' education of both groups is the same, and a third one, supposes that their father's education is the same across teenage and not-teenage moms.

$$Y^{g}_{i} = \beta_0 + \sum_j W_{ji}\psi_j^{g} + u_i \qquad \text{(xvi)}$$

Where g identifies the group having kids before 20 and W_j is a j-row vector including a set of dummies for country.

$$Y^{g}_{i} = \beta_0 + M_{1i}\psi^{g} + u_i \qquad \text{(xvii)}$$

Where M_1 accounts for the dummie mother attending to college

$$Y^{g}_{i} = \beta_0 + F_{1i}\psi_1^{g} + u_i \qquad \text{(xviii)}$$

Where F_1 accounts for the dummie father attending to college

Firstly, we have derived linear models for each of the groups according to the previous specifications.

	Dependent variable:			
	college			
	(Before 20)	(After 20)	(Before 20)	(After 20)
college.mom	0.485***	0.410***		
	(0.038)	(0.067)		
college.dad			0.428***	0.250***
			(0.032)	(0.079)
Constant	0.176***	0.063***	0.167***	0.083***
	(0.010)	(0.020)	(0.010)	(0.021)

Table 6. Independent regressions by parent education

	Dependent variable:	
	college	
	(1)	(2)
cntry2	−0.000	−0.105*
	(0.175)	(0.059)
cntry3	0.000	−0.019
	(0.223)	(0.066)
cntry4	0.000	−0.103
	(0.196)	(0.071)
cntry5	−0.000	−0.030
	(0.253)	(0.064)
cntry6	0.174	0.132**
	(0.158)	(0.052)
cntry7	−0.000	0.033
	(0.188)	(0.060)
cntry15	0.000	−0.135**
	(0.175)	(0.058)
cntry16	0.083	0.115**
	(0.168)	(0.056)
cntry17	0.000	−0.054
	(0.179)	(0.063)
cntry18	0.250	0.055
	(0.206)	(0.070)
cntry19	0.000	0.005
	(0.175)	(0.059)
cntry20	−0.000	0.093
	(0.223)	(0.067)
cntry21	0.280*	0.221***
	(0.157)	(0.049)
cntry22	−0.000	−0.054
	(0.326)	(0.071)
Constant	0.000	0.173***
	(0.146)	(0.037)

Table 7. Independent regressions by country

Although on average differences across coefficients are not striking there exist a few exceptions (see college.dad estimators). Consequently, our approach is relevant, and it deserves some attention. The intuition behind the Oaxaca decomposition is to interpolate the expected college attendance if women in son.bef20 were distributed in the same way as women in son.aft20 (across our variables of interest, in this case, country/mom into college/dad into college).

We have used a DFL method (Card, 2019) for reweighting the distribution because its simplicity and intuitiveness. We just need to regress our variable son.bef20 on the set of dummies presented above (in this occasion the logit model has been used for convenience), then predict college attendance of son.bef20 using that same model and eventually computing our weights ω_i in the following way,

$$w_{g(i)} = \frac{\widehat{m}_i}{1-\widehat{m}_i} \; if \; g \neq 1$$

$$w_{g(i)} = 1 \qquad if \; g = 1 \tag{xix}$$

Where $\widehat{m}_i$ is the fitted value of son.bef20 using logit regression. Finally, we can simply compute our counterfactual mean in the following way:

$$\bar{y}^{b}_{counterf} = \sum_{i \in b} w_g y_i \left[\sum_{i \in b} w_g\right]^{-1} \tag{xx}$$

The following chart summarizes the main findings of this process, where the first-row accounts for the non-weighted mean.

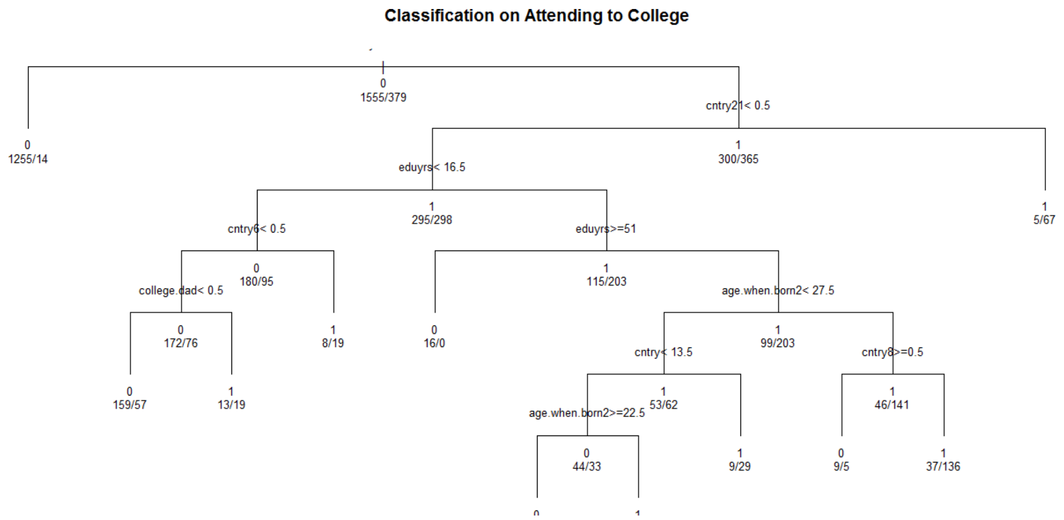
	Weighted mean	Difference in means	% gap explained when weighting
Son.bef20	0.1010	0.1065	0.0000
Counterfactual for country	0.1593	0.0481	45.2301
Counterfactual for mother educ	0.1606	0.0468	43.9764
Counterfactual for father educ	0.1494	0.0580	54.4885

Table 8. Oaxaca decomposition results

As we see, when introducing weighting, a big share (around 50% of the gap) can in fact be explained by the different distributions of son.bef20 across the selected variables. My coding skills have not allowed me to implement an Oaxaca decomposition whose counterfactual corrects for the three sets of dummies at the same time (possible future line of investigation). Although a big gap is explained by control variables, optimistically, it could be argued that still a 50 percent of the difference remains unexplained and may potentially be attributed to pregnancy itself. Because of previous methods, we know indeed that the contribution of teenage pregnancy to college attendance when correcting for a larger set of variables is much smaller; but still, we consider it to be a nice result for presenting to the reader.

Finally, we have implemented an unusual robustness check, i.e. a tree analysis. Trees and other bootstrap classification algorithms, such as random forest, are usually used in classification processes (not that much in causality identification). Although it is not the intention of this work to predict overall college attendance, trees schemes could be used to determine the explanatory power of son.bef20. If this variable was in fact a strong predictor of college attendance, as previous analyses suggest, it might be possible that the algorithm would place it as a main branch (a main classifier). Unfortunately, my guesses were mistaken.

Figure 10. Tree classification model



Son.bef20 is not even selected as a classifier within the first six levels of the classification tree. This result underlines the fact that, despite being a significant factor explaining college attendance as tested above, its overall impact might be very small compared to other variables.

4.3 Differences across genders

The last methodology part of this paper will be focused in analyzing the different impact of parenting depending on gender. Intuitively, because women are the ones who get physically pregnant and because traditional socialization models impose a higher home labor tasks burden on women (especially after pregnancy, Espin-Andersen 2009) we would expect a larger magnitude of female estimators. Although many different approaches could be used to test the different impact of pregnancy across genders, we have selected the following two:

1. Method 1 (M1): Regressing son.bef20 on probability of college using the father sample and check for the significance of the estimator. Ideally, we would observe a non-significant statistic (indicating no contribution at all) or a smaller- magnitude statistic (indicating a smaller contribution) when using the male sample compared to the female sample.

2. Method 2 (M2): Producing a similar fuzzy RD design to the one developed for teenage mothers. Ideally, we would have observed a decay in the probability of attending to college in the women subsamples (but we did not) and we could have compared this decay with the one observed in men (which potentially may act as counterfactual distribution). This design would have been a very solid evidence of cross-gender differences, but because no discontinuity had been observed in women's data, we expect no discontinuity in male sample either. Moreover, because of the small sample of teenage fathers I have been forced to use the whole subset of fathers and mothers for this RD, what brings back most of the problems identified in Section 4.1.

The results of the first approach are summarized in Table 9.

	Dependent variable:
	college.masc
son.bef20.masc	−0.107
	(0.07)
Constant	0.183***
	(0.018)
Observations	469
R^2	0.008
Adjusted R^2	0.006
Residual Std. Error	0.375 (df = 467)
F Statistic	3.835* (df = 1; 467)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 9. College attendance regression on son.bef20 (male subsample)

As presented in Section 4.2, this raw estimator may suffer from all kinds of biases but in broad terms it is very similar to the one obtained in the women subsample. It should be noted though that unlike the one for women, this coefficient is far from significant what may suggest no significant variation in the probability of attending to college for men when having a child before 20. Nevertheless, there exists another potential mechanism explaining the no-significance of this estimator, namely the reduced number of male observations in ESS. If we further take into consideration that (as proved in Figure 4) men statistically become parents at an older age than women, we may understand why such a negligible share of the respondents were in fact teenage fathers (n<100). In that sense, the limited number of observations may be responsible for the lack of overall significance.

Surprisingly, M2 provides with two very interesting results,

a) Men tend to have their offspring at an older age. Although we were already aware of this trivial result, it is interesting to confirm it through this circuitous approach. This intuition can be derived from the fact that there is a lower number of men in school before parenting (despite the fact that we know that for the cohorts included in the Survey men were more educated than women). This fact can indeed be explained in terms of aging, when men become parents, they are significantly older thus their probability of being in school is lower than the one for women (note that ages are normalized to zero=age at birth).

b) More surprisingly, a fact that was completely undetected when including only the women sample, is that men in fact do not decrease their schooling probability after giving birth while women sharply do. In previous sections we interpreted women's probability decay as a simple timing effect, but in fact, their schooling decline could be explained by parenting (for instance because of increasing time devoted to housework). Obviously, the same aging problems detected in a) could still be relevant. It could simply be that aging is the driving factor of this effect. When parenting, men are "so old" that their schooling probability had already reached a floor and simply could not decrease any further. Analogously, women reach a floor at a lower level and at later relative age (five years after pregnancy, which precisely is the expected average difference in age when parenting across genders).

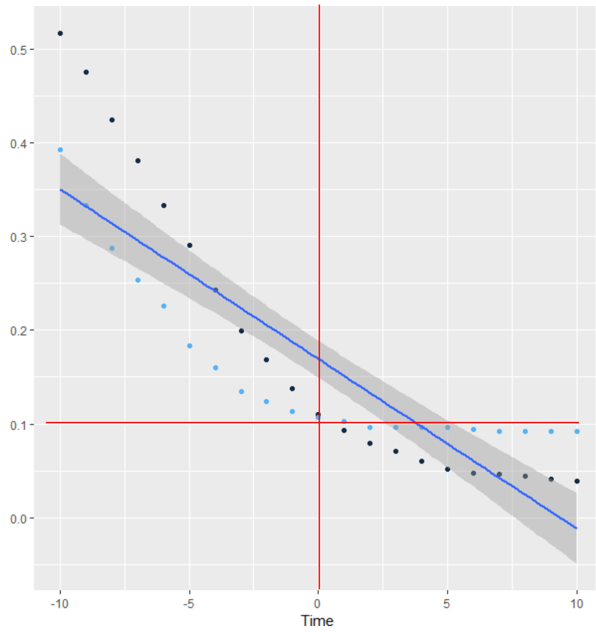


Figure 10. Cross-gender scatter plot. Probability of currently studying across time (light-blue = men, black=women)

Further investigation concerning the nature of those floors should be conducted. A possible design for that experiment would be to use an Oaxaca approach. That is, if men were parents at the same age as women, which would be their college density and distribution function.

Note too, that although we were looking for a RD approach a quasi Diff-in-Diff design has emerged. Obviously, this approach would require indefensible assumptions such as parallel trends before treatment and equivalent schooling distributions across cohorts (given that we are using cross-sectional data). However, it might be the case that when correcting for age, these assumptions are not “that unrealistic” anymore, and Diff-in-Diff interpretations could hold.

5. SUMMARY OF RESULTS

Despite the severe limitations of data, **it has been proved that pregnancy has a small magnitude impact on school attendance**. This result is robust to different samples, approaches and specifications. This result holds even for teenage parents, as this cohort does not present a relevant decay in the probability of attending to college after parenting.

Nevertheless, the lack of relevance does not affect the lack of significance, as it has been consistently proved too that **teenage pregnancy is in fact a significant factor explaining the probability of**

attending to college. It has also been proved through Oaxaca decomposition that teenage mothers are substantially different from older mothers in terms of country and socioeconomic background. These two results could be used to reasonably argue that teenage mothers are in fact Never Takers (NT), that is, that they would have dropped school (or they would have not enrolled into subsequent education levels) regardless pregnancy.

To summarize, we can conclude that teenage pregnancy on average, has a significant impact on education, but this impact might be much more limited in terms of magnitude than originally believed. Formally,

$$E[Y_{1i}|son.bef20=1, mother=TRUE] \neq 1 \qquad \text{(xxi)}$$

The different educational effects of (teenage) parenting between men and women have remained largely unexplained. Preliminary results suggest potential differences across groups, but these are far from conclusive.

6. FUTURE LINES OF INVESTIGATION AND RECOMMENDATIONS

The main recommendations have to do with the quality of the data. As presented in the introduction and data analysis sections, unreliable survey data usually leads to an attenuation bias. Using administrative data could potentially unveil new results.

Despite this big limitation, appealing lines of research have been settled in this article. Among them, we must underline the different returns to pregnancy for men and women and the development of mixed Oaxaca decomposition methods for simultaneously accounting for different variables.

It could also be interesting to expand the geographical scope of this study to other regions where teenage pregnancy is (i) more frequent and (ii) may have a higher impact than in Europe, where the State provides with abundant resources for childcaring.

Another possible line of research might be related with the time-series development of teenage pregnancy and its potentially different contributions to education across time. In this sense, it will be attractive to develop methods accounting for the variation of teenage pregnancy in relation with overall economic performance at both micro (personal income) and macro level (general economic performance of the country).

It might be interesting too to account for different variations in schooling probability depending on the baby characteristics. For example, we might expect a higher impact on education of (teenage) pregnancy if the baby is born with health issues. More mundane questions such as the heterogenous effects of pregnancy depending on the gender of the infant (might be the case that boys are more time demanding than girls?) could be analyzed too.

Eventually, and taking into consideration the sharp variations in teenage pregnancy across territories, it could be tempting to develop empiric models supporting or disregarding existing untested theories explaining cross-country variation (Gonzalez and Videgain, 2016).

7. PERSONAL REFLECTION AND CONCLUSION

Going through this data analysis has been incredibly challenging. Its quality was really poor, and the absence of certain key variables (such as income or race) further complicated the task. In addition, the permanent failure in finding results has been mentally exhausting and somehow discouraging, especially when you research a topic you know very little about, and you do not know what exactly you are looking for.

Nevertheless, I think I have successfully provided with a decent variety of tests and intuitions (difference in means, Oaxaca decomposition, linear regression, Lasso type models, RD, trees…) which hopefully have revealed some surprising and robust results in the field of teenage pregnancy. I hope you have enjoyed this paper as much as I had.

ANNEX 1:

Complete set of specifications for regression analysis

	OLS	Logit	Lasso Gaussian	Lasso Logit	Ridge
constant	-0.2615	-0.2615	-0.2128	-5.2447	-4.9774
son.bef20	-0.0162	-0.0162	-0.0079	-0.3167	-0.4039
college.mom	0.2437	0.2437	0.2428	1.2460	1.1995
college.dad	0.2332	0.2332	0.2280	1.2003	1.1596
cntry2	-0.0483	-0.0483	-0.0364	-0.6941	-0.6270
cntry3	-0.0487	-0.0487	-0.0262	-0.3665	-0.3624
cntry4	-0.0931	-0.0931	-0.0723	-0.9042	-0.8654
cntry5	-0.0846	-0.0846	-0.0675	-0.6762	-0.6058
cntry6	0.0825	0.0825	0.0612	0.4824	0.5497
cntry7	0.0434	0.0434	0.0089	0.0317	0.1937
cntry8	-0.0724	-0.0724	-0.0597	-0.5566	-0.5147
cntry9	0.0335	0.0335	0.0046	0.0740	0.1682
cntry10	0.0141	0.0141	0.0000	0.0000	0.0184
cntry11	-0.0471	-0.0471	-0.0394	-0.7063	-0.7345
cntry12	0.0765	0.0765	0.0355	0.3938	0.4311
cntry13	-0.0025	-0.0025	-0.0085	-0.0748	-0.0908
cntry14	0.3752	0.3752	0.3329	1.6848	1.6373
cntry15	-0.0590	-0.0590	-0.0719	-1.1999	-1.0529
cntry16	0.0791	0.0791	0.0358	0.2970	0.3330
cntry17	-0.0605	-0.0605	-0.0628	-0.6549	-0.6789
cntry18	0.0117	0.0117	0.0000	0.0000	-0.0276
cntry19	0.0499	0.0499	0.0000	0.0228	0.1007
cntry20	0.1154	0.1154	0.0582	0.4117	0.4357
cntry21	0.1641	0.1641	0.1158	0.8400	0.7730
cntry22	-0.0913	-0.0913	-0.0906	-0.8766	-0.9077
cntry23	0.0339	0.0339	0.0000	0.0000	-0.1244
cntry24			0.0000	0.0000	0.0875
domicil	-0.0272	-0.0272	-0.0246	-0.2270	-0.2135
uempli	-0.0473	-0.0473	-0.0257	-0.3187	-0.5032
dsbld	0.0048	0.0048	0.0000	0.0000	-0.0067
rtrd	-0.0415	-0.0415	-0.0278	-0.5065	-0.5203
hswrk	-0.0024	-0.0024	0.0000	0.0000	-0.0046
pdjobev	0.0145	0.0145	0.0142	0.1413	0.1256
emplrel_1	0.0876	0.0876	0.0000	0.4298	0.3606
emplrel_2	0.0351	0.0351	0.0609	0.9795	0.7889
emplrel_3	0.1694	0.1694	0.0273	0.9074	0.9118
wrkctra_1	-0.1440	-0.1440	0.0000	0.0000	0.0315
wrkctra_2	-0.1293	-0.1293	0.0029	0.0398	0.1109
wrkctra_3	-0.1735	-0.1735	-0.0176	-0.2308	-0.2831
tporgwk_1	0.1598	0.1598	0.0517	0.4754	0.4911
tporgwk_2	0.2266	0.2266	0.1248	0.9018	0.8689
tporgwk_3	0.1077	0.1077	0.0010	0.0000	0.0583
tporgwk_4	0.0913	0.0913	0.0000	0.0000	0.0199
tporgwk_5	0.0764	0.0764	0.0000	0.0000	0.1676
tporgwk_6	0.0927	0.0927	0.0000	0.0000	-0.0537
cntry			0.0015	0.0131	0.0197
age.when.born2	0.0088	0.0088	0.0082	0.0692	0.0613
eduyrs	0.0075	0.0075	0.0074	0.0473	0.0442
Selected predictors	48.0000	48.0000	37.0000	37.0000	48.0000
RMSE	0.3380	0.3380	0.1254	0.7979	0.8890

Table 10. Complete set of specifications for OLS, Logit, Lasso-Gaussian, Lasso-logit and Ridge models. Regression of college attendance on son.bef20 and controls

	Btw1	Btw2	Btw3	Btw4	Btw5	Btw6	Btw7	Btw8	Btw9
constant	-4.5267	-4.9609	-4.9326	-5.0502	-4.9452	-5.1005	-5.1480	-4.9802	-5.3174
son.bef20	-0.3412	-0.3726	-0.3323	-0.3309	-0.3126	-0.3202	-0.3191	-0.2987	-0.3257
college.mom	1.1687	1.2118	1.2196	1.2299	1.2261	1.2361	1.2396	1.2328	1.2485
college.dad	1.1283	1.1737	1.1774	1.1871	1.1822	1.1919	1.1949	1.1878	1.2031
cntry2	-0.5515	-0.6473	-0.6297	-0.6543	-0.6318	-0.6645	-0.6744	-0.6432	-0.7069
cntry3	-0.2883	-0.3626	-0.3291	-0.3447	-0.3219	-0.3473	-0.3537	-0.3244	-0.3786
cntry4	-0.7249	-0.8688	-0.8294	-0.8610	-0.8226	-0.8690	-0.8810	-0.8291	-0.9252
cntry5	-0.5196	-0.6251	-0.6102	-0.6360	-0.6130	-0.6470	-0.6569	-0.6240	-0.6885
cntry6	0.4576	0.5118	0.4824	0.4872	0.4712	0.4818	0.4823	0.4645	0.4907
cntry7	0.0652	0.1204	0.0523	0.0521	0.0199	0.0350	0.0341	0.0000	0.0483
cntry8	-0.4275	-0.5204	-0.5025	-0.5241	-0.5029	-0.5325	-0.5407	-0.5101	-0.5675
cntry9	0.0791	0.1252	0.0847	0.0862	0.0650	0.0756	0.0753	0.0529	0.0852
cntry10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
cntry11	-0.6052	-0.7115	-0.6664	-0.6857	-0.6526	-0.6861	-0.6931	-0.6492	-0.7232
cntry12	0.3272	0.4064	0.3724	0.3851	0.3592	0.3817	0.3860	0.3549	0.4087
cntry13	-0.0456	-0.0766	-0.0591	-0.0654	-0.0565	-0.0675	-0.0700	-0.0566	-0.0779
cntry14	1.5655	1.6500	1.6516	1.6678	1.6532	1.6716	1.6763	1.6596	1.6938
cntry15	-0.9166	-1.0766	-1.0846	-1.1285	-1.1024	-1.1532	-1.1693	-1.1258	-1.2143
cntry16	0.2658	0.3210	0.2866	0.2945	0.2729	0.2892	0.2919	0.2682	0.3101
cntry17	-0.5505	-0.6489	-0.6155	-0.6338	-0.6062	-0.6365	-0.6429	-0.6038	-0.6680
cntry18	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
cntry19	0.0080	0.0728	0.0165	0.0242	0.0000	0.0139	0.0169	0.0000	0.0413
cntry20	0.3699	0.4295	0.3996	0.4081	0.3879	0.4034	0.4062	0.3858	0.4241
cntry21	0.7322	0.8026	0.7983	0.8155	0.8018	0.8213	0.8277	0.8117	0.8504
cntry22	-0.7334	-0.8651	-0.8279	-0.8516	-0.8161	-0.8546	-0.8624	-0.8121	-0.8932
cntry23	0.0000	-0.0148	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
cntry24	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
domicil	-0.1981	-0.2167	-0.2168	-0.2211	-0.2182	-0.2230	-0.2244	-0.2203	-0.2285
uempli	-0.3762	-0.4358	-0.3549	-0.3507	-0.3159	-0.3285	-0.3255	-0.2898	-0.3350
dsbld	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
rtrd	-0.4425	-0.5084	-0.4821	-0.4942	-0.4722	-0.4932	-0.4977	-0.4685	-0.5197
hswrk	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
pdjobev	0.1232	0.1308	0.1354	0.1376	0.1387	0.1397	0.1403	0.1415	0.1406
emplrel_1	0.1708	0.3537	0.2709	0.3277	0.2383	0.3390	0.3673	0.2290	0.4932
emplrel_2	0.5530	0.8061	0.7761	0.8517	0.7490	0.8707	0.9051	0.7462	1.0505
emplrel_3	0.5176	0.8574	0.7013	0.7843	0.6345	0.7881	0.8261	0.6098	1.0031
wrkctra_1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
wrkctra_2	0.0400	0.0640	0.0416	0.0432	0.0332	0.0394	0.0397	0.0287	0.0442
wrkctra_3	-0.2362	-0.2745	-0.2317	-0.2346	-0.2131	-0.2269	-0.2282	-0.2017	-0.2423
tporgwk_1	0.4067	0.4679	0.4521	0.4633	0.4497	0.4653	0.4688	0.4516	0.4826
tporgwk_2	0.7995	0.8655	0.8670	0.8815	0.8735	0.8888	0.8935	0.8813	0.9061
tporgwk_3	0.0000	0.0113	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
tporgwk_4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
tporgwk_5	0.0857	0.0821	0.0018	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
tporgwk_6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
cntry	0.0148	0.0155	0.0143	0.0141	0.0136	0.0136	0.0134	0.0127	0.0132
age.when.born2	0.0575	0.0634	0.0650	0.0666	0.0662	0.0677	0.0682	0.0674	0.0695
eduyrs	0.0423	0.0450	0.0455	0.0462	0.0458	0.0466	0.0468	0.0462	0.0476
Selected predictors									
RMSE	0.8122	0.8046	0.7966	0.7989	0.8013	0.7981	0.7986	0.7983	0.8028

Table 11.In between estimators. Regression of college attendance on son.bef20 and controls

BIBLIOGRAPHY

1. Angrist, J. D., and Pischke, J.S. (2009). *Mostly harmless econometrics: An empiricist's companion*.
2. Barbezat, A., and Hughes, J. (2005) *Salary Structure Effects and the Gender Pay Gap in Academia*. Research in Higher Education, vol. 46, no. 6,, pp. 621–640. *JSTOR*, www.jstor.org/stable/40197439.
3. Caldwell, J. C. 1967. *"Fertility attitudes in three economically contrasting rural regions of Ghana,"* Economic Development and Cultural Change 15: 217-238.
4. Caspi A., Jaffee S., Moffitt T.E., Belsky J, Silva P. (2001) *Why are children born to teen mothers at risk for adverse outcomes in young adulthood? Results from a 20-year longitudinal study. Development and Psychopathology*, vol. 13, pp. 377–397.
5. Card D. (2019) *Lecture on Oaxaca Decomposition*. Course ECON C142. University of California at Berkeley
6. Christensen, Sandra, et al. (1975). *Factors Affecting College Attendance. The Journal of Human Resources*, vol. 10, no. 2, pp. 174–188. JSTOR, www.jstor.org/stable/144825
7. Dudley P. and Bouvier L., (2010). *An Introduction to Demography*, Cambridge Press, pp. 265-285
8. Esping-Andersen G. (2010). *The incomplete Revolution*, Polity Press pp.17- 55
9. European Social Survey, (2018). *European Social Survey: Phase 8.* https://www.europeansocialsurvey.org/
10. Gonzalez, C. and Videgain, M. (2016). *Teenage Pregnancy, a Sociological Approach*. Univerisidad Carlos III de Madrid
11. Jain, Anrudh K., and John A. Ross. (2012) *Fertility Differences Among Developing Countries: Are They Still Related to Family Planning Program Efforts and Social Settings*? International Perspectives on Sexual and Reproductive Health, vol. 38, no. 1, pp. 15–22. JSTOR, www.jstor.org/stable/41472761.
12. Klepinger, D., Lundberg, S., & Plotnick, R. (1999). *How does adolescent fertility affect the human capital and wages of young women*? Journal of Human Resources, 421-448.