LUND
UNIVERSITY

Female Labour Supply and Fertility at the Extensive Margin

Master Thesis Year 1, Department of Economics

Author: Anna Linder
Supervisor: Jan Bietenbeck

Seminar: 2016-08-22

Abstract

This paper studies the female labour supply response to fertility at the extensive margin. The strategy pursued in this analysis is to identify two groups of childless women who utilized family planning services for the first time, one group for help with pregnancy and one group for help with contraception. The identification strategy is meant to resemble a natural experiment where actual treatment of a child within 21 months from the family planning service is as good as random among women in the two samples. The results of this paper indicate strong and negative fertility effects, with dissipating importance as age and years of education increases. It is argued that the assumption on randomness holds, and that the observed differences in employment status between the treatment groups can be interpreted as the causal impact of having a first child (below age one) on female labour supply. Further, the stability of the results between the two different samples suggests strong external validity.

Keywords

Fertility, extensive fertility margin, female labour supply, female labour force participation

Table of Contents

1. Introduction

During the 20th century a historical decline in fertility took place in many developed countries. For example, in the United States the total number of births per woman decreased from 3,7 in 1960 to 1,9 in 2016 (World Bank, 2016). During the same period of time female labour force participation increased dramatically, in the United States the female labour force participation rate steadily increased from practically non-existing in the beginning of the century to around 52 % in 1990, making female participation add to around 40 % of the total work force (World Bank, 2016)). Since then, the trend subsided and even today female labour force participation in the United States is just over 50 % (World Bank, 2016). An important question this raises is whether the decline in fertility caused the increase in female labour force participation. Understanding the causal channels between fertility and female labour supply is important for both theoretical and practical reasons, but especially from policy perspective. If post war increases in female labour force participation can be explained by the contemporaneous decline in fertility, it would have important policy implications in several areas, for example regarding child welfare, social security and sustainability through equality across gender and socioeconomic status. It is plausible that extended access to childcare, or more stable working conditions regarding terms of employment, could encourage earlier re-entrance to the labour market for women after childbearing. It is also possible that legislation regarding parental benefits could be used to achieve a more equal distribution of labour force participation between parents.

Estimating the causal effect of fertility on female labour force participation is complicated by several issues; firstly, it could be the case that women with different preferences for children also differ in unobserved ways. For example, if women who choose to have children also have weaker attachment to the labour force or possess lower potential labour market outcomes than women who do not choose to have children, it would constitute an adverse selection problem (Lundborg, Plug and Rasmussen, 2014). Secondly, if decisions regarding fertility and labour supply either are made simultaneously or if the fertility decision depends on potential labour market outcomes, it is possible that joint determination or reversed causality could cause bias in the observed relationship between fertility and female labour supply (Miller, 2011). Observed differences in female labour market outcomes between women with and without children are therefore difficult to interpret, as the observed differences not only reflects the average causal effect of fertility on those actually treated with children, but also the differences in potential labour market outcomes as untreated, irrespective of actual treatment (Angrist and Pischke, 2009).

A large existing literature, which is discussed in detail in the following section of this paper, has convincingly addressed these issues using instrumental variables. However, these approaches have mainly allowed for the study of female labour responses to intensive margin fertility changes, i.e. comparing females with one versus several children. Since it is plausible that female labour responses at the intensive margin of fertility are stunted by those responses at the extensive margin, it is argued that extensive margin fertility changes are more important to study, evidence regarding such responses is however still very rare.

To overcome some of the mentioned selection and endogeneity issues, this paper uses utilization of family planning services in combination with birth data as a source of random variation in fertility. Data from the United States National Survey of Family Growth is used to study the female labour force participation response to fertility at the extensive margin using two different empirical strategies. The first strategy follows a recent study by Cristia (2008), and compares the labour force participation of a sample of childless women who sought medical help to get pregnant. 21 months after seeking help with pregnancy, i.e. 12 months after the earliest possible date of birth, some of these women had indeed conceived a child while others had not. Absence of a child 21 months after seeking help with pregnancy is an event, which I refer to as "conception failure". Under the assumption that having a conception failure is as good as random among these women, comparing employment status between women in the sample gives the causal impact of having a first child (below age one) on female labour supply. This strategy is complemented with a second approach using a sample of childless women who sought help with contraception. Similarly to the first empirical strategy, 21 months after seeking help with contraception some of these women had indeed conceived a child while others had not. Presence of a child 21 months after seeking help with contraception is an event, which I refer to as "contraception failure". Under the assumption that having a contraception failure is as good as random among these women, comparing employment status between women in the sample gives a second estimate of the causal impact of having a first child (below age one) on female labour supply.

This paper contributes to the existing literature by extending the analysis of Cristia (2008) to include a complementary empirical approach on a different sample of women, which allows for assessment of the external validity of the results. The results of this paper show that the treatment effect of having a first child reduces female labour supply by; 26 percentage points when experiencing a contraception failure, and by 21 percentage points when not experiencing a

conception failure. The stability of the results between the two identification strategies supports the external validity of the results, the fact that the treatment effect between the different samples are of comparable magnitude, is evidence on the generalizability of the results. The results of this paper further imply that level of attachment to the labour force may be of great importance regarding the female labour response to fertility. It is plausible that older and more educated individuals have a relatively stronger attachment to the labour force through more work experience or higher eligibility for employment. If true, this could be an important implication for policy as it would be possible to reduce the negative fertility effects on female labour supply through for example human capital investments or more stable working for young or socioeconomically vulnerable individuals.

The empirical strategy pursued in this paper rests on assumptions of randomness, the validity of this assumption is however not evident. On one hand, robustness tests indicate that some pre-treatment characteristics are not completely independent of treatment status, which is a possible caveat when interpreting the results. On the other hand, the coefficients are very stable across specifications when adding control variables, and placebo regressions show that baseline labour supply is independent of treatment status. Therefore it is argued that the estimates of this paper can be given a causal interpretation among these women.

The outline of this paper proceeds as follows; section 2 discusses the existing literature on the area fertility and female labour supply, section 3 presents the data, sample selection and descriptive statistics of this paper, section 4 describes the empirical strategy and section 5 presents the results. Section 6 presents the robustness checks on the main results, while section 7 discusses and section 8 makes some concluding remarks on this paper.

2. Literature Review

This section discusses the large existing literature related to fertility and female labour supply.

Earlier studies have estimated robust negative correlations between fertility and female labour market outcomes and consensus is that women with children both work less and have lower income than women without children. However, since the relationship between fertility and labour market outcomes is likely to be affected by endogeneity issues, it has been difficult to draw credible inferences regarding causality (Browning, 1992). To overcome this issue, later studies have addressed the relationship between fertility and female labour force participation using instrumental variables. Two strategies have been of particular interest among researchers when looking at fertility on the intensive margin; (1) using twin births, and (2) using parental preferences for mixed gender composition of children as instruments. Lately, the ambition to study fertility at the extensive margin has compelled innovative empirical strategies using for example family planning failures and in vitro fertilization success as instruments.

Twin births have been used to provide exogenous variation in family size, and thereby measure the effect of a higher order child on labour market outcomes. This approach rests on the assumption that a twin birth randomly distributes an unanticipated child to some families, and therefore a causal effect of this variation can be estimated. Rosenzweig and Wolpin (1980) use the occurrence of a twin birth to estimate the effect of a second child on female life cycle labour force participation. Rosenzweig and Wolpin (1980) find that the impact of a second child through a twin birth reduces female labour force participation in the age group 15-24 years with 37,1 %. This effect dissipates rapidly with age and in the oldest age group examined, 35-44 years, the sign of the effect is reversed and indicates a 14,2 % higher probability of labour force participation. An important conclusion from Rosenzweig and Wolpin (1980) is that although the initial negative effect on labour force participation is more common for mothers of twins compared to mothers of singletons, this effect is compensated by an earlier return to the labour force by mothers of twins later in the life cycle.

Similarly, Bronars and Grogger (1994) and Gangadharan, Rosenbloom, Jacobson and Wishart Pearre (1996) uses twin births to instrument fertility change and estimate its economic consequences. In line with earlier results using twin birth strategies, these articles find negative short-term effects with dissipating importance in the long run, however some additional important conclusions are drawn in these articles. Bronars and Grogger (1994) find racial disparities regarding both the magnitude and the persistence of the effects, while the negative

effect of a twin birth for white mothers is smaller and evens out in the long run; the effect for black mothers is stronger and more persistent. Gangadharan et al. (1996) interestingly find that as the female labour force participation rate grew stronger, from around 0,52 in 1980 to 0,63 in 1990, the negative effect of fertility on female labour force participation also grew considerably, from 7,5 % to 13,1 % during the same time period.

Angrist and Evans (1998) estimate the effect of fertility on parents' labour supply using parental preferences for mixed gender composition of children as instrumental variable. This empirical strategy assumes that the mix of gender between children is as good as random, and uses the fact that having same-sex children equals a higher probability of having another higher-order child because of parental preferences for mixed gender composition of children. Angrist and Evans (1998) shows that the effect of fertility on female labour supply is negative and significant, however not very strong and dissipating in the long run. Their results also indicate that the effect of fertility on female labour supply among females with college education and for females with high-income husbands is almost insignificant (Angrist and Evans, 1998).

Using twin births and parental preferences for mixed gender composition of children have attracted many researchers as they provide natural experiments with exogenous variation in family size, and thereby circumvent some of the complicating issues of unobserved heterogeneity and endogeneity when estimating the relationship between fertility and labour supply. However, on a whole these techniques have observed relatively small effects on labour supply that dissipate in the long run as they only provide estimates of responses to fertility on the intensive margin.

Intuitively it is plausible that the effect of a first child matters most for labour supply responses, this has also been confirmed by some more recent articles using family planning failures and in vitro fertilization (IVF) success as natural experiments to induce exogenous variation in having a first child (Cristia, 2008; Lavado, 2014; and Lundborg, Plug and Rasmussen, 2014). Cristia (2008) estimates the effect of a first child on female labour supply using conception failures as a source of exogenous variation in fertility. He uses a sample of women who sought help to get pregnant and argues that the variation in fertility (having conceived a first child or not) 21 months later is exogenous. Cristia (2008) is aware of the possible problem that early success in fertility treatment is not completely random but provides evidence that prior employment and subsequent pregnancy is uncorrelated. Cristia (2008) finds that having a child younger than one year reduces female employment with around 26 percentage points. These results, and the identification

strategy used, are especially interesting for two reasons; firstly, the effect of a first child is estimated, and secondly, the results indicate that female labour supply responses to fertility are stronger than earlier suggested in related literature. Lavado (2014) and Lundborg et al. (2014) study female labour outcome responses to fertility on the extensive margin using two different instrumental variables approaches. Lavado (2014) uses the imperfect control of fertility as a natural experiment by combining data of monthly contraception use with birth data to detect contraception- and conception failures, which he exploits as instruments for exogenous variation in family size. The strongest negative effect of a first child on female labour supply is found for women with low education and strong preferences for children and leisure, for women with high education and low preferences for children and leisure the effects are substantially lower (Lavado, 2014). Further, Lavado (2014) argues that the effect of a first child is more persistent than the intensive margin fertility effects suggested in earlier literature. Lundborg et al. (2014) study the effect of having children on female labour market outcomes using a sample of women participating in IVF, where success at first IVF-attempt is expected to be as good as random. Lundborg et al. (2014) study labour supply responses to fertility at both the extensive and intensive margins, and find that responses to fertility at the extensive margin are much stronger than those responses at the intensive margin. Lundborg et al. (2014) also show that there are strong and negative fertility effects on female labour market outcomes regarding both earnings and labour force participation during the first two years, however, while the fertility effects on earnings are long lasting, the responses in labour force participation dissipate in the long run.

Results throughout this literature review agree on the direction of results, however they are conflicting regarding the intensity of the effect. It is plausible that these differences are related to sample selection. Consequently, in spite of relatively high internal validity, some of these findings only relate to specific samples and are therefore not generalizable to other settings. Some articles have shown that the effect of a first child is of greater importance than higher-order children, however evidence regarding this relationship is still relatively rare. This motivates the present analysis to study female labour supply responses to fertility on the extensive margin. The use of two different empirical strategies for the purpose of this paper allows for comparison of the results between the two different samples, which enables assessment of the external validity of the results. Earlier studies also indicate that age, race and level of education are significant variables when estimating the female labour response to fertility, therefore a heterogeneity analysis and several robustness test will be made, to make sure that the results of this paper are not driven by differences in age, level of education, race or ethnicity.

3. Data, Sample Selection and Descriptive Statistics

This section presents the data, sample selection process and descriptive statistics used in this analysis.

3.1 Data

The empirical analysis of this paper uses data from the National Survey of Family Growth (NSFG), which is conducted by the National Centre of Health Statistics in the United States. The NSFG collects information on family life, marriage and divorce, pregnancy, contraceptive use, and general and reproductive health. The NSFG has been conducted in six cycles between 1973 and 2002, as well as three surveys between 2006-2010, 2011-2013 and 2013-2015. In general these surveys are comparable, however some of the cycles contain specific features that can be used for different purposes. The NSFG cycle 5 from 1995 is the only cycle that provides retrospective employment history for up to 10 periods of employment for each of the 10847 females included in the survey. The NSFG cycle 5 also provide information on specific months when family planning services were used, as well as birth data. Since this paper aims to study the effect of fertility on female labour supply using utilization of family planning services in combination with birth data and employment history, the NSFG cycle 5 is the most suitable cycle for this analysis.

3.2 Sample Selection

This section presents the selection process of the two samples used in this analysis, PREG-HELP consisting of those who sought help with pregnancy and BC-HELP consisting of those who sought help with contraception. Some of these steps are similar to the sample selection process of Cristia (2008), especially regarding which months are used to compare employment. Further in this paper the month of seeking help, the 21st month after seeking help and 12th month prior to seeking help will be referred to as month 0, month 21 and month -12, respectively.

The sample of women included in this analysis is selected as follows: (1) start with the full sample of 10847 females included in the NSFG cycle 5. (2) Include only women who could specify the month when they first sought help with pregnancy or contraception. (3) Exclude women who sought help for the first time before age 19, as well as women who sought help for the first time <21 months prior to the interview. The reason is that the NSFG only register employment history for women above age 18, and that employment status in month 0, month 21 and month -12 are used in this analysis. (4) Exclude women who already had children when they sought help with pregnancy or contraception for the first time, and finally, (5) exclude women who had non-biological children living with them in month 21, since taking care of non-biological children

could confound the estimates (up to five non-biological children are taken into account). This process generated two samples, PREG-HELP consisting of 530 women and BC-HELP consisting of 163 women.

3.3 Descriptive Statistics

The identification strategy is based on the assumption that conception and contraception failure provides a distribution of individuals' characteristics between the different treatment groups that is as good as random. To assess the validity of this assumption a good strategy is to compare pre-treatment outcomes or other covariates between the different treatment groups (Angrist and Pischke, 2009). Sample means and standard deviations of the two samples, PREG-HELP and BC-HELP, as well as the difference in means between the treatment groups are presented in Tables 1 and 2. Treatment refers to having given birth to a first child between month 0 and month 21 after seeking help with pregnancy or contraception.

Table 1: Descriptive Statistics PREG-HELP sample

	PREG-HELP Treatment=0			PREG-HELP Treatment=1			
VARIABLES	N	mean	sd	N	mean	sd	Difference in means
Years of education	341	13.96	2.563	189	14.06	2.602	0,100
Employed month 0	341	0.839	0.368	189	0.847	0.361	0,008
Employed month 21	341	0.860	0.348	189	0.630	0.484	0,230***
Employed month -12	341	0.842	0.365	189	0.831	0.376	0,011
Age month 0	341	26.26	4.423	189	25.42	4.481	0,840**
Hispanic	341	0.0502	0.219	189	0.130	0.337	0,080***
Black	341	0.0804	0.272	189	0.0851	0.280	0,005

Note: Descriptive statistics is presented by treatment group, treatment refers to giving birth to a first child between month 0 and month 21 after seeking help with pregnancy. Data is collected from the NSFG *** p<0.01, ** p<0.05, * p<0.1

Table 2: Descriptive Statistics BC-HELP sample

	BC-HELP Treatment=0			BC-HELP Treatment=1			
VARIABLES	N	mean	sd	N	mean	sd	Difference in means
Years of education	143	14.46	1.883	20	12.70	2.452	1,760***
Employed month 0	143	0.671	0.471	20	0.600	0.503	0,071
Employed month 21	143	0.755	0.431	20	0.550	0.510	0,205*
Employed month -12	143	0.566	0.497	20	0.750	0.444	0,184*
Age month 0	143	19.78	0.840	20	19.75	0.910	0,030
Hispanic	143	0.154	0.362	20	0.300	0.470	0,146
Black	143	0.182	0.387	20	0.350	0.489	0,168

Note: Descriptive statistics is presented by treatment group, treatment refers to giving birth to a first child between month 0 and month 21 after seeking help with contraception. Data is collected from the NSFG. *** p<0.01, ** p<0.05, * p<0.1

There are statistically significant differences between the treatment groups in both samples, for the PREG-HELP sample regarding age and being of Hispanic origin, and for the BC-HELP sample regarding years of education and employment status 12 months prior to seeking help. The difference in means of employment in month 21 is statistically significant in both samples, which is suggestive of the main results of this paper, however this estimate should not be interpreted as the causal effect, as it does not control for potentially confounding variables. The differences in means between the treatment groups regarding the rest of the variables are statistically insignificant, which indicates that distribution of these characteristics across the different treatment groups appear to be as good as random

Comparison of the statistics between the two samples tell us that the BC-HELP sample is considerably younger, this observation is in line with expectations since intuitively it makes sense that those seeking help with contraception are younger than those seeking help to get pregnant. This is not a problem for the internal validity of the results but needs to be taken into consideration when generalizing the results to other settings, or when comparing the results between the two samples.

4. Empirical Strategy

This section describes the empirical strategy of this paper presented in method and econometric model.

4.1 Method

The objective of this paper is to study the female labour force participation response to fertility at the extensive margin using utilization of family planning services in combination with birth data as a source of exogenous variation in fertility. The empirical strategy is based on two different approaches where the treatment effect of having a first child below age one on female labour supply is estimated. The first strategy follows a recent study by Cristia (2008) where a sample of women who sought help to get pregnant is identified. The specific month of when each of these women sought help for the first time is used to compare employment status 21 months later between those treated and those not treated with a child, i.e. not having or having a conception failure. These women are observed in month 21, i.e. 12 months after the earliest possible date of birth, if a gestation period of 9 months is assumed. The reason for using this particular time span is twofold; firstly, it enables a clearer definition of the treatment effect since the conceived child cannot be above the age of one, and secondly, it delimits the possibility of additional children being born, which could have confounding effects on labour supply. This strategy is complemented with a second approach where a sample of women who sought help with contraception is identified. The specific month of when each of these women sought help for the first time is used to compare employment status 21 months later between those treated and those not treated with a child, i.e. having or not having a contraception failure. The proposed empirical strategies rests on the assumption that among the sample of women who sought help to get pregnant, and among the sample of women who sought help with contraception, actual treatment of having a child 21 months later is as good as random, and therefore unrelated to their baseline labour supply. Under this assumption, comparing labour force participation between the treatment groups gives the causal impact of having a first child below age one on female labour supply.

4.2 The Econometric Model

Regression analysis is performed using Ordinary Least Squares (OLS) to estimate the effect of a first child (below the age of one) on female labour supply. Consider the following model,

$$Y_{it} = \alpha + \rho D_i + X_i'\gamma + \delta YEAR_t + \eta_{it}$$

where Y_{it} is the binary outcome variable of female labour supply representing employment status in month 21. The treatment variable D_i, also binary, represents treatment of a first child between month 0 and month 21, and ρ is the OLS-estimator of the average treatment effect, our

coefficient of interest in this case. X'_i represents a vector of control variables represented by age at month 0 and years of education at interview date, as well as dummies for being of Hispanic origin and being black. Finally, year fixed effects are added to the model, as it is possible that time specific differences might influence the results.

5. Results

This section discusses the main results of the paper and presents a heterogeneity analysis regarding age, years of education, race and ethnicity.

5.1 Main Results

The main results are presented in Table 3 for the PREG-HELP sample, and Table 4 for the BC-HELP sample. The results indicate that the treatment effect of having a first child (below age one) by not experiencing a conception failure reduces female labour supply with around 21 percentage points. Further, the results indicate that the treatment effect of having a first child (below age one) by experiencing a contraception failure reduces female labour supply with around 26 percentage points. These results are statistically significant on 1 % and 5 % level, respectively.

When control variables are added to the regression the standard errors decrease. This could be explained by the fact that if treatment is randomly assigned, including control variables provides a more efficient estimate of the treatment effect, since including additional determinants of the outcome variable reduces the residual variance. The coefficients of interest are quite stable across the different specifications, which implies that selection regarding the observed control variables is only a minor issue. Consequently, if selection on unobservables is similar to that on observables, the assumptions on randomness, which the empirical strategy builds on, are likely to hold.

The validity of the results can further be assessed through comparison with earlier evidence. Cristia (2008) suggests that treatment of a first child by not experiencing a conception failure reduces female labour supply with around 26 percentage points, which is in line with present paper's estimations, however slightly stronger. This disparity is unexpected since the empirical strategies between these papers are similar; a possible explanation could be that Cristia (2008) uses some different control variables and does not include year fixed effects. Lavado (2014) shows that females with strong preferences for children have a negative short-run effect on female labour supply of between 10-22 percentage points, and that females with weak preferences for children have a negative short-run effect of between 2-6 percentage points. If we assume that women who sought help with pregnancy have strong preferences for children, and women who sought help with contraception have weak preferences for children, the results of this paper go in the same direction as Lavado (2014), although the actual magnitude of the effect differs regarding women with weak preferences for children. These disparities are likely to be affected by the different assumptions made by Lavado (2014) in his structural econometric

model. Lundborg et al. (2014) find that female labour force participation reduces with 6 to 7 percentage point during the first two years after childbearing. A pre-treatment labour force participation rate of around 90 % indicates that the effect estimated by Lundborg et al. (2014) is weaker than the estimations of this paper. This could be explained by the fact that Lundborg et al. (2014) estimate the effect of any child on female labour force participation, i.e. include both extensive and intensive fertility changes. Lundborg et al. (2014) also estimate the female labour response to a first child; in this case the authors use earnings as the outcome measure, which is less comparable to the results of this paper. Overall the results of this paper agree with earlier evidence, there are some disparities regarding magnitude of the effect, which are likely to be caused by differences in the empirical approaches.

Table 3: OLS estimates of the effect of a first child on female labour supply PREG-HELP sample

	(1)	(2)	(3)
	PREG-HELP	PREG-HELP	PREG-HELP
VARIABLES	Employed month 21	Employed month 21	Employed month 21
Treatment	-0.230***	-0.214***	-0.208***
	(0.0429)	(0.0419)	(0.0413)
Age month 0		0.00898*	0.000456
		(0.00483)	(0.00520)
Years of education		0.0150**	0.0154**
		(0.00647)	(0.00665)
Hispanic		-0.130*	-0.0977
		(0.0675)	(0.0687)
Black		0.0421	0.0417
		(0.0528)	(0.0512)
Constant	0.860***	0.418***	0.629***
	(0.0202)	(0.137)	(0.149)
Observations	530	530	530
R-squared	0.069	0.104	0.194
Year FE			YES

Note: Table 3 presents regression output on the PREG-HELP sample. Treatment refers to giving birth of a first child between month 0 and month 21 after seeking help with pregnancy. Data is collected from the NSFG. Sample weights provided by the NSFG are used to get a nationally representative sample. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 4: OLS estimates of the effect of a first child on female labour supply BC-HELP sample

	(1)	(2)	(3)
	BC-HELP	BC-HELP	BC-HELP
VARIABLES	Employed month 21	Employed month 21	Employed month 21
Treatment	-0.225*	-0.259**	-0.255**
	(0.136)	(0.131)	(0.124)
Age month 0		-0.00332	-0.00674
		(0.0513)	(0.0600)
Years of education		-0.0238	-0.0217
		(0.0195)	(0.0196)
Hispanic		-0.0775	-0.0636
		(0.0990)	(0.106)
Black		0.0444	0.0635
		(0.0753)	(0.0774)
Constant	0.785***	1.203	1.236
	(0.0354)	(0.959)	(1.124)
Observations	163	163	163
R-squared	0.023	0.037	0.063
Year FE			YES

Note: Table 4 presents regression output on the BC-HELP sample. Treatment refers to giving birth of a first child between month 0 and month 21 after seeking help with contraception. Data is collected from the NSFG. Sample weights provided by the NSFG are used to get a nationally representative sample. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

5.2 Heterogeneity Analysis

It is implied in previous research that age, as well as racial and educational differences could yield differences in the labour supply response to fertility. This section provides a heterogeneity analysis that interacts the treatment indicator with variables for age, level of education, race and ethnicity, to test for differences in treatment effect regarding these variables.

The results for the PREG-HELP sample are presented in Table 5 and show that the coefficients for interaction between treatment status and two of the variables, age at month 0 and years of education are statistically significant. This shows that there seem to be differences in the treatment effect between women of different age and level of education. Further, the coefficients of the interactions with age at month 0 and years of education are positive, indicating that employment in month 21 increases with both age and years of education, among those treated with children. The coefficients for interaction between treatment status and the two dummies for being black and being of Hispanic origin are statistically insignificant, which indicates that the treatment effects between black and non-black women, and between women of Hispanic and non-Hispanic origin, seem to be similar.

The results for the BC-HELP sample are presented in Table 6 and show that the coefficients for interaction between treatment status and variables for age at month 0, years of education, being black and being of Hispanic origin, are all statistically insignificant. However, since several of these coefficients actually are quite large, this is more likely the result of poor statistical power due to few observations in this sample, rather than evidence on homogeneity of treatment effects. Thus, it is possible that there are differences in the treatment effect regarding age, level of education, race and ethnicity also in the BC-HELP sample, and that the main results are driven by these differences. This is also indicated by the fact that the coefficients for the main treatment indicator loose their statistical significance when the interactions are included in the specifications.

Table 5: Heterogeneity Analysis PREG-HELP sample

VARIABLES	(1) PREG-HELP Employed month 21	(2) PREG-HELP Employed month 21	(3) PREG-HELP Employed month 21	(4) PREG-HELP Employed month 21
Age month 0	-0.000255	0.000538	0.000527	-0.00628
	(0.00526)	(0.00522)	(0.00521)	(0.00564)
Hispanic	-0.0868	-0.0988	-0.0279	-0.0968
	(0.0679)	(0.0685)	(0.0837)	(0.0690)
Years of education	0.00476	0.0154**	0.0152**	0.0143**
	(0.00721)	(0.00663)	(0.00660)	(0.00661)
Treatment	-0.664***	-0.213***	-0.196***	-0.762***
	(0.222)	(0.0440)	(0.0435)	(0.260)
Treatment#Black		0.0600		
		(0.110)		
Black	0.0386	0.0202	0.0438	0.0392
	(0.0512)	(0.0583)	(0.0515)	(0.0503)
Treatment#Years of education	0.0325**			
	(0.0151)			
Treatment#Hispanic			-0.129	
			(0.129)	
Treatment#Age at month 0				0.0215**
				(0.00996)
Constant	0.796***	0.629***	0.627***	0.823***
	(0.164)	(0.149)	(0.149)	(0.164)
Observations	530	530	530	530
R-squared	0.203	0.195	0.196	0.206
year FE	YES	YES	YES	YES

Note: Table 5 presents regression output on the PREG-HELP sample. Treatment refers to giving birth of a first child between month 0 and month 21 after seeking help with pregnancy. Data is collected from the NSFG. Sample weights provided by the NSFG are used to get a nationally representative sample. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 6: Heterogeneity Analysis BC-HELP sample

VARIABLES	(1) BC-HELP Employed month 21	(1) BC-HELP Employed month 21	(1) BC-HELP Employed month 21	(1) BC-HELP Employed month 21
Age month 0	0.00441	-0.00598	-0.00840	0.00477
	(0.0602)	(0.0601)	(0.0593)	(0.0634)
Hispanic	-0.0614	-0.0634	-0.00731	-0.0593
	(0.106)	(0.106)	(0.112)	(0.106)
Years of education	-0.0124	-0.0212	-0.0207	-0.0183
	(0.0203)	(0.0198)	(0.0195)	(0.0196)
Treatment	0.475	-0.228	-0.167	1.842
	(0.765)	(0.143)	(0.147)	(2.439)
Treatment#Black		-0.149		
		(0.256)		
Black	0.0723	0.0837	0.0637	0.0680
	(0.0757)	(0.0791)	(0.0745)	(0.0762)
Treatment#Years of education	-0.0554			
	(0.0551)			
Treatment#Hispanic			-0.380	
			(0.243)	
Treatment#Age at month 0				-0.105
				(0.123)
Constant	0.879	1.211	1.246	0.958
	(1.140)	(1.128)	(1.109)	(1.201)
Observations	163	163	163	163
R-squared	0.072	0.065	0.075	0.067
year FE	YES	YES	YES	YES

Note: Table 6 presents regression output on the BC-HELP sample. Treatment refers to giving birth of a first child between month 0 and month 21 after seeking help with contraception. Data is collected from the NSFG. Sample weights provided by the NSFG are used to get a nationally representative sample. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

6. Robustness

As mentioned earlier in this paper, selection may cause potential outcomes to not be independent of treatment status, which would cause assumptions on randomness to be wrongful. To assess the severity of this potential failure, this section provides robustness test on the main results through an F-test on the joint significance of the selected covariates on treatment status, as well as placebo regressions on baseline and lagged employment status. The final part of this section discusses the sensitivity of the main results to subgroups of different age, level of education, race and ethnicity.

6.1 F-test

If individuals of different age, education level, race or ethnicity respond differently to fertility regarding labour supply, these variables could be systematically related to treatment status, which could bias the results. The statistically significant difference in means between the different treatment groups regarding some of the control variables suggest that this might be the case. To further test for random variation between the treatment groups, treatment status is regressed on all the selected covariates in a joint significance F-test. The results are presented in Table 7 and indicate that there are statistically significant predictors of treatment success in both samples, for the PREG-HELP sample regarding age and being of Hispanic origin, and for the BC-HELP sample regarding years of education. The P-values show that the selected covariates are jointly significant, in the PREG-HELP sample on 1 % level and in the BC-HELP sample on 10 % level. These results imply that treatment success to some extent is predicted by these variables, and therefore might not be independent of potential outcomes.

Table 7: F-test on joint significance

	(1)	(2)
	PREG-HELP	BC-HELP
VARIABLES	Treatment	Treatment
Age month 0	-0.0117**	0.0188
	(0.00504)	(0.0292)
Years of education	0.0137	-0.0341*
	(0.00857)	(0.0174)
Hispanic	0.256***	0.0192
	(0.0715)	(0.0747)
Black	0.0297	0.0435
	(0.0655)	(0.0651)
Constant	0.428***	0.203
	(0.154)	(0.701)
Observations	530	163
R-squared	0.032	0.064
P-value F-test joint significance	0.0004	0.0667

Note: Table 7 presents regression output on PREG-HELP and BC-HELP samples. Treatment refers to giving birth of a first child between month 0 and month 21 after seeking help with pregnancy or contraception. Data is collected from NSFG. Sample weights provided by the NSFG are used to get a nationally representative sample. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

6.2 Placebo Regressions on Baseline and Lagged Employment

To be able to interpret the observed difference in employment rates in month 21 between the treatment groups as the causal impact of having a first child, it is important to make sure that baseline employment rates are similar between the two treatment groups. Therefore, placebo regressions on employment status in month 0 and month -12 are performed. The results of these regressions are presented in Table 8 and show that lagged employment statuses are independent of treatment status in the PREG-HELP sample. In the BC-HELP sample the same applies for employment in month 0, however employment status in month -12 appear to be correlated with treatment status. It is possible that this result is influenced by the fact that the BC-HELP sample is relatively young with a mean age just above 20 years, in combination with the fact that employment in young ages is more volatile.

Table 8: Placebo test on baseline and lagged employment status

VARIABLES	PREG-HELP Employed month 0	PREG-HELP Employed month -12	BC-HELP Employed month 0	BC-HELP Employed month -12
Treatment	0.0418	0.0192	0.0764	0.158**
	(0.0327)	(0.0331)	(0.101)	(0.0726)
Age month 0	0.00511	0.00388	0.106	0.00113
	(0.00489)	(0.00493)	(0.0712)	(0.0398)
Years of education	0.00565	0.0195***	-0.0214	0.000316
	(0.00595)	(0.00642)	(0.0226)	(0.0148)
Hispanic	-0.118*	-0.0898	-0.0482	0.0569
	(0.0700)	(0.0685)	(0.107)	(0.0929)
Black	0.00949	-0.0120	-0.170	-0.141
	(0.0518)	(0.0545)	(0.113)	(0.115)
Employed at age 18	0.138***	0.131***	0.206**	0.659***
	(0.0392)	(0.0389)	(0.0842)	(0.0635)
Constant	0.534***	0.381**	-1.233	0.176
	(0.157)	(0.159)	(1.308)	(0.749)
Observations	530	530	163	163
R-squared	0.115	0.123	0.103	0.470
year FE	YES	YES	YES	YES

Note: Table 8 presents regression output on PREG-HELP and BC-HELP samples. Treatment refers to giving birth of a first child between month 0 and month 21 after seeking help with pregnancy or contraception. Data is collected from NSFG. Sample weights provided by the NSFG are used to get a nationally representative sample. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

6.3 Sensitivity to Subgroups

As mentioned before, it is implied in previous research that individuals of different age, race and educational level might differ regarding their labour supply response to fertility. If this statement is true, there is a possibility that the results of this paper are biased, because the treatment groups are not comparable. For example, Hispanic origin individuals are substantially overrepresented in the treated group of the PREG-HELP sample, at the same time it is empirically implied that this group is underrepresented in the labour force (see for example Mora, 2015). This is a potential source of selection bias, which could cause a simple control strategy like OLS to overestimate the effect of fertility on labour supply. Further, there are statistically significant differences regarding age between the treatment groups in the PREG-HELP sample and regarding years of education

in the BC-HELP sample. If women of different age or education level also differ regarding for example labour force attachment, it could cause selection bias. Since the treated group in the PREG-HELP sample have a lower mean age, treatment status is negatively correlated with age. If younger women at the same time work less than their older counterparts, OLS estimations could overestimate the effect of a first child on female labour supply. Similar argument can be made regarding differences in education. The treated group in the BC-HELP sample is less educated, if those with lower education at the same time have lower attachment to the labour force, OLS-estimations could overestimate the effect of a first child on female labour supply.

The heterogeneity analysis presented in section 5.2 suggests that there are differences in the treatment effect regarding some of these variables, the question that remains is to what extent the results of this paper are driven by such selection. Including control variables for these potential confounders in the last specifications of the main results should correct for some of this potential bias, and the fact that the results are relatively stable across the specifications suggests that selection due to these variables is only a minor issue (see Tables 3 and 4). Further, the fact that the main treatment coefficients remain statistically significant when the interaction terms are added in the heterogeneity analysis suggests that the main results are not fully driven by selection, at least regarding the PREG-HELP sample. The results in Table 5, specification 2 and 3, indicate that the treatment effects for non-black women, and for non-Hispanic women, are similar to the main results of this paper. This provides further evidence that the results of this paper not are driven by selection due to race and ethnicity. The results in Table 5, specifications 1 and 4, indicate that although the coefficients of the main treatment variable remain statistically significant when the interaction terms are added, the magnitude of these coefficients is substantially higher than in the main results. This indicates that the treatment effect is much stronger for those with low levels of education and for those with young ages, which also suggests that the results of this paper to some extent are likely to be driven by selection due to age and educational level.

7. Discussion

This paper aims to estimate the effect of extensive margin fertility changes on female labour supply using OLS regression analysis, however several difficulties arise with this task. The relationship between fertility and labour supply is likely to be affected by unobserved heterogeneity and endogeneity issues, which complicates the analysis as a simple control strategy like OLS is unable to draw credible inferences with observational data. As mentioned in the introduction, the observed differences in employment between women with and without children does not only reflect the average causal effect of fertility on those actually treated with children, but also the differences in potential labour market outcomes as untreated, irrespective of actual treatment. The strategy pursued in this analysis is to identify two groups of women who utilized family planning services for the first time, one group for help with pregnancy and one group for help with contraception. The identification strategy is meant to resemble a natural experiment where actual treatment of a child within 21 months from the family planning service is as good as random. If the assumption of randomness holds, this strategy is a solution to the selection problem, and the observed differences between these counterfactuals can be interpreted as the causal effect of having a first child on female labour supply. Whether the assumptions on randomness holds can be discussed since this paper presents evidence of both success and failure. On one hand some observed control variables, for example age and years of education, seem to be correlated with treatment status. On the other hand, the stability of the coefficient of interest across the different specifications implies that selection is only a minor issue. Placebo regressions show that baseline employment at month 0 is uncorrelated with subsequent fertility, which further supports the assumptions made on randomness.

Experiments are considered the gold standard when causality is to be estimated (see for example Angrist and Pischke, 2009). Striving to find natural experiments to exploit, researchers often construct samples with specific characteristics. However, selecting these specific samples for the benefit of high internal validity often come at a cost of low external validity, as the estimated effect only holds for the selected sample. This issue is a potential caveat of this paper, as it is possible that the PREG-HELP and BC-HELP samples are different from the general population. On one hand, this implies that generalization of the results to other populations of interest should be made with caution. On the other hand, the OLS estimates between the two different samples are stable and of comparable magnitude, which instead supports the external validity of this paper. It would perhaps be desirable to compare baseline variables between the PREG-HELP and BC-HELP samples and the general population to assess whether they are representative of other populations of interest, unfortunately the identification strategy pursued

in this paper is based on observation of employment at specific times from utilization of family planning services, which makes a credible comparison to another sample difficult.

Another potential limitation of this paper is the fact that the empirical results are based on relatively old data from 1995, generalization of the effects to present times could be questioned. Previous research have shown that changes over time regarding female labour force participation has had substantial impact on the magnitude of the fertility effects, for example Gangadharan et al. (1996) show that as the female labour force participation rate grew from 0,52 in 1980 to 0,63 in 1990, the negative fertility effect grew from 7,5 % to 13,1 % during the same time period. This could imply that the female labour response to fertility is fluctuant, and that generalization of these effects over time should be done with caution. However, changes in female labour participation have since 1990 been insignificant (World Bank 2016), therefore it is possible that the estimated effect can be generalized over time to current settings.

8. Concluding Remarks

This paper is able to establish a statistically significant relationship between female labour supply and fertility on the extensive margin and shows evidence of a strong and negative treatment effect of a first child (below age one) on female labour force participation. It is argued that the assumption imposed in the empirical strategy holds, and that the estimated effects can be given a causal interpretation. The introduction of this paper questions whether the post war decline in fertility caused the contemporaneous increase in female labour force participation; the results of this paper imply that this is likely to be the case.

In general, the results of this paper are in line with previous research. Firstly, regarding the strong and negative extensive margin fertility effects also found by Cristia (2008), Lundborg et al. (2014) and Lavado (2014). Secondly, regarding the differences in treatment effect between women of different ages. Strong and negative fertility effects were found for individuals of young ages with dissipating importance as age increased, which assents with the results of Rosenzweig and Wolpin (1980). And thirdly, regarding differences in the treatment effect between individuals of different educational level. Strong and negative fertility effects were found among low educated individuals with dissipating importance with years of education, which agrees with the results of Angrist and Evans (1998).

Differences in the level of attachment to the labour force are plausible causes of the disparities between groups of different age and educational level. Intuitively one would think that with age and level of education comes a more profound attachment to the labour force through more work experience or greater eligibility for employment. Further, it is possible that among groups with lower socioeconomic status, less qualified jobs and worse employment conditions leads to a weaker attachment to the labour force. If we assume what is stated above is true, that attachment to the labour force has a large impact on the female labour supply response to fertility, then the conclusions of this paper are of great importance, especially from public policy perspective, as they imply that it is possible to influence women's labour supply response to fertility. For example, policies to induce human capital investments, or policies that allocate young or socioeconomically vulnerable individuals with more stable employment conditions, could reduce the negative fertility effects on female labour supply. This finding is definitely the most interesting point of this paper, and one that would benefit from further investigation.

References

Angrist, Joshua D and William N Evans, "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size", *American Economic Review*, June 1998, 88(3), 450_77.

Angrist, Joshua D and Jörn-Steffen Pischke, Mostly Harmless Econometrics, An Empiricist's Companion, Princeton University Press, 2009.

Bronars, Stephen and Jeff Grogger, "The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment", *American Economic Review*, December 1994, 84 (5), 1141–56

Browning, Martin, "Children and Household Economic Behavior", *Journal of Economic Literature*, 1992, pp. 1434–1475.

Cristia, Julian P, "The Effect of a First Child on Female Labor Supply: Evidence from Women Seeking Fertility Services", *Journal of Human Resources*, 2008, 43 (3), 487_510.

Gangadharan, Jaisri, Joshua Rosenbloom, Joyce Jacobson, and James Wishart Pearre, "The Effects of Child-Bearing on Married Women's Labor Supply and Earnings: Using Twin Births as a Natural Experiment", NBER Working Papers 5647, National Bureau of Economic Research, Inc July 1996.

Lavado, Pablo, "The Effect of a Child on Female Work When Family Planning May Fail", CEMFI Working Paper No. 1405, 2014.

Lundborg, Petter, Erik Plug and Astrid Rasmussen, "Fertility Effects on Female Labor Supply: IV Evidence from IVF Treatments", Working Paper, 2014.

Maurin, Eric and Julie Moschion, "The Social Multiplier and Labor Market Participation of Mothers", *American Economic Journal: Applied Economics*, January 2009, 1 (1), 251–72.

Miller, Amalia R, "The effects of motherhood timing on career path", *Journal of Population Economics*, July 2011, 24 (3), 1071_1100.

Mora, Maria T, "The increasing importance of Hispanics to the U.S. workforce", *Monthly Labor Review,* Sep 2015, p1-3. 3p

Rosenzweig, Mark and Kenneth Wolpin, "Life-Cycle Labor Supply and Fertility: Causal Inferences from Household Models", *Journal of Political Economy*, April 1980, 88 (2), 328_48

World Bank website
Data on female labour force participation and total births per woman
http://data.worldbank.org/indicator/SL.TLF.CACT.FE.ZS/countries (accessed 2016-05-11)
http://data.worldbank.org/indicator/SP.DYN.TFRT.IN?locations=US (accessed 2016-08-08)