Erasmus University Rotterdam

Erasmus School of Economics

Bachelor Thesis [Economie en Bedrijfseconomie]

The effect of an increase in fertility on the labour market outcomes of parents in the United Kingdom: an instrumental variable approach

Name student: Rindert Ruit

Student ID number: 507750

Supervisor: dr. A.A.F Gerritsen

Second assessor: dr. A.L. Boring

Date final version: 11-7-2021

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam

In this paper I research what happens to the labour outcomes of existing parents in the United Kingdom if they decide to have additional children or an additional child. I distinguish three separate labour market outcomes; whether or not a parent is employed; for how many hours a week a parent works and the gross labour income a parent earns each month. I use data gathered using surveys in the years 1992, 2001 and 2002 and pool observations from these separate years into one larger sample. To circumvent endogeneity problems I use whether a parent's first two children are of the same sex as an instrument for the amount of children this parent has and as an instrument for whether or not this number is larger than two. I thereby restrict my sample to parents with at least two children. All my estimates of treatments effects on all three of the separate outcome variables are statistically insignificant. The results remain insignificant when I split up my sample in smaller subsamples. The relatively large confidence intervals accompanying my estimates indicate that I don't obtain accurate estimates of null effects either. It is likely that this is caused by the main pitfall of my analysis; the limited size of my sample.

Table of contents

I. Introduction

In this paper I aim to investigate what happens to labour market outcomes of parents in the United Kingdom, if these parents already have two children and decide to have additional children or an additional child. Three labour market outcomes are of interest to me. These are: whether or not a parent is employed; for how many hours a week a parent works and how much a parent earns each month by working, before taxation.

It is important to investigate these effects. If an increase in fertility does influence labour market outcomes, then knowledge about these effect helps us explain why certain individuals end up in certain positions in the labour market. Knowledge about why individuals end up in certain positions is necessary if we want to improve these positions. In other words, the knowledge is necessary to guide economic policy. Assume, for example, that an increase in fertility negatively affects the amount of hours a parent works, because this parent has to take care of the additional child. Assume, furthermore, that the parent earns less because of this. If a government knows why this parent earns less, namely an increase in fertility perhaps accompanied by a lack of childcare options, then this government can implement precise policy to improve the labour market position of this parent. A government could, for example, subsidize childcare for parents, which is in fact a relevant measure in the United Kingdom (BBC News, 2020).

At first I focus on all parents with at least two children, the youngest of whom is at most seventeen years old. Afterwards I divide this sample into smaller subsamples. These are: one subsample containing only men; one containing only women and four separate subsamples with different combinations of sex and educational attainment. I make these divisions in order to obtain separate estimates, enabling me to investigate whether the effects differ between subgroups of the population. This type of knowledge is also desirable. If increased fertility affects different groups of individuals differently, then knowing this may help us explain disparities in labour outcomes between individuals from different groups. An example showing that the latter is important to society is the debate about the gender wage gap found in the United Kingdom (Petter, 2020) and its potential causes. It is also true in this context that knowledge about why disparities exist helps in resolving them, because it enables us to directly address the causes.

To create a dataset I pool observations from three separate waves of a survey. In my data section I will explain why this doesn't mean I use multiple observations for a single individual. I could run a simple OLS regression using this dataset, however, I suspect that this would lead to biased estimates because of reverse causality and unobserved confounders. To solve these problems I use an instrumental variable approach, in which I use whether or not a parent's first two children are of the same sex as an instrument for how many children that parent has. The necessary assumptions are not unlikely in this context. Some parents may want to have children of both sexes, in which case parents with two children of the same sex would be more likely to have additional children than parents of two children with differing sexes. A correlation between the instrument and other determinants of labour market outcomes is also unlikely, since the sex of a child is, arguably, as good as randomly assigned. Finally, it is hard to imagine mechanisms other than fertility through which the instrument might affect labour market outcomes. This makes the exclusion restriction a reasonable assumption. I also add control variables to my regressions to further warrant my estimates against bias.

My analysis can be seen as an attempt to answer the following research question: What is the effect of having additional children on the probability of employment, the amount of hours worked per week and the gross monthly labour income for parents in the United Kingdom? None of my estimates of these effects is statistically significant at any of the conventional levels, whichever treatment variable I use, whichever subgroup I focus on and whichever outcome variable I use as a dependent variable. This shouldn't be interpreted as evidence that there are in fact no economically significant treatment effects. The 95 percent confidence intervals of my estimates are relatively wide and include effects that could be considered economically significant.

II. Relevant literature

II.A. Theoretical literature

In this section I discuss previous research into the effects of fertility on labour market outcomes. This will help explain the academical relevance of my research and help me when interpreting my results. I start by discussing several theoretical papers.

Becker (1981) provides a theoretical explanation for why fertility might affect a parent's labour market outcomes. He argues that parents maximize household utility, which is a function of domestically produced goods and goods bought on the market, by dividing their time between their children and the labour market. Becker also argues that human capital increases with the cumulative time spent working and that wages increase with human capital. This makes specialization within the household optimal, with one parent investing time in the children and the other spending time on the labour market. The parent who (partially) stops working thereby accumulates less human capital, leading to lower wages even after a (partial) return to the labour market. It can be argued that women were traditionally more likely to be the parent that focussed on the children. This theory could then partially explain the wage gap between men and women. Then the wage gap should also be smaller in societies or periods where emancipation enables both parents to spend equal amounts of time with the children. According to this theory the effects of fertility on labour market outcomes should also be smaller if there are better or more affordable options for domestic production outside the family environment. For example better or more affordable childcare provision.

Willis (1973) uses a model similar to Becker's, in which household utility is a function of how many children a couple has, the 'quality' of these children and other sources of satisfaction. These other sources of satisfaction include goods that can be bought with labour income, allowing individuals to indirectly derive utility from working. Willis assumes that some individuals have a higher starting level of human capital. He argues that the opportunity cost of time invested in children is higher for these individuals, since wages increase with human capital. This, ceteris paribus, leads parents with a higher starting level of human capital to invest relatively less time their children. An example of a group with a higher starting level of human capital could be college educated individuals. This theory would then predict that the effects of fertility on labour supply are smaller for these individuals.

II.B. Empirical literature

Angrist and Evans (1998) investigate fertility effects using data from the United States. They do so for two separate years: 1980 and 1990. They use a variable indicating whether or not the first two children a parent has are of the same sex as an instrument for the amount of children this parent has. This 'Same-sex' variable is missing for individuals with less than two children. I will explain why the instrument affects fertility in my methodology section. Angrist and Evans argue that this instrument solves endogeneity problems typically encountered when trying to estimate fertility effects using an OLS approach, since the sex of a child is randomly assigned. They divide their sample into three subsamples the smallest of which contains about 250,000 observations. The subsamples are; all women; only married women and only the husbands of married women. They obtain separate estimates for these separate groups and also check whether or not the effects for married women differ with their educational attainments. The authors find that having additional children reduces the labour supply of women in both years, regardless of their marital status. This effect is present both on the extensive margin, whether or not a woman has paid work, and on the intensive margin, for how many weeks a woman works and for how many hours a week. They also find a negative effect of increased fertility on women's labour income for both years, irrespective of these women's marital status. They find no significant effects on the family income, the labour supply of husbands or the labour income of husbands. An important result of their additional analysis is that the significant negative effects of increased fertility on labour supply are absent for women with more than a high-school degree. All these results correspond to the predictions of the theoretical framework.

Cools, Markussen and Strøm (2017) investigate the effects of fertility on labour supply and career outcomes by applying the 'Same-sex' instrumental variable approach to Norwegian administrative data. This means their sample also includes only parents with at least two children. They divide their sample into groups of different sexes and into groups with different educational attainments to investigate whether the effects differ. The smallest subsample they use consists of 12,000 individuals. They also follow the parents in their sample for up to 40 years so as to investigate how the effects develop over time. The authors use the number of children a parent has as a treatment variable, and find that for Norwegian women labour supply decreases in the short and in the medium run if fertility increases. This effect is present both on the extensive margin and the intensive margin. These effects on labour supply disappear if the authors focus on the period beginning twenty years after the birth of the second child. They also find a negative effect of having an additional child on the quality of women's careers. Here career quality is measured by the probability of being employed by one of the higher paying employers and by the earnings rank within a firm. These effects also fade out in the long run and are stronger for college-educated women. The latter is a difference with the conclusions of the theoretical papers and the results from Angrist and Evans (1998). In these papers the effects of fertility decrease with the female's educational attainment. A possible explanation for this difference is the strong focus on redistribution in Norway, which might weaken the effect higher human capital has on post-taxation wages. This would weaken the incentive of college-educated women to keep working on the labour market after childbearing. The authors find no significant effects of increased fertility on male labour market outcomes.

Lundborg, Plug and Rasmussen (2017) use Danish data to analyse the effects of fertility on labour market outcomes. They create a sample that consists of about 18,500 women who underwent IVF-treatment and who had no children prior to this treatment. The authors use whether or not the IVF-treatment was successful as an instrument for whether or not an individual has a child. They argue this is possible, since whether or not IVF-treatment is successful is likely to be uncorrelated with labour market outcomes and mechanisms, besides fertility, influencing them. They find that having a first child significantly reduces the labour earnings of mothers. This effect declines over time, but does remain. The authors also find negative effects on labour participation, both on the extensive and intensive margin. These effects disappear six to ten years after the birth of a child. Wages follow an opposite pattern. In the short run the authors find no effect on wages, but in the medium and long run, two to ten years after the birth of the child, the authors find a significant, negative effect. These type of causal movements can be fitted into the theoretical framework.

Cruces and Galiani (2007) investigate the effects of fertility on labour supply in Mexico and Argentina. They also use the 'Same-sex' variable as an instrument for whether or not a parent has more than two children. They restrict their sample to women between 21 and 35 years old, whose youngest child is at most 18 years old and whose second child is at least a year old. In the end they have almost 600,000 observations for Argentina and almost 460,000 observations for Mexico. This sample is divided into two subsamples; a sample containing all women and a sample containing solely married women. The authors find that increased fertility has a significant negative effect on the labour supply of women in both countries. This corresponds to the hypotheses provided by the discussed theories.

Iacovou (2001) also uses the 'Same-sex' variable as an instrument for whether or not an individual has more than two children. She separately uses two separate British datasets; the British Household Panel Survey and the National Child Development Study. She only uses the second wave of the British Household Panel survey, which has data concerning the year 1992. She restricts her samples to mothers between 21 and 49 years old. Over all, her regressions include between about 1,300 and 3,200 observations. Her point estimates of the effect of having additional children on the probability of employment are positive. Her point estimates of the effect that increased fertility has on the amount of hours a woman works each weeks are also positive. None of her point estimates are statistically significant, which is a difference between her results and the results of the other discussed papers. Iacovou ascribes her lack of significant estimates to the limited size of her samples.

My research resembles the work already done by Iacovou, which makes discussing the differences useful. A more elaborate discussion of my exact method and why this method is valid will follow later. The first difference is that Iacovou doesn't estimate the effects of fertility on labour market outcomes for men, I will do so. She also doesn't divide her sample of women into college educated women and women without a college education. I will do so, for both the men and women in my sample. Furthermore, she doesn't obtain estimates of the effect fertility has on the monthly labour incomes of parents. This is one of the three relevant dependent variables in my analysis. Another difference lies in the fact that she only uses the second wave of the British Household Panel Survey. Instead, I will be pooling data from three separate waves to increase the size of my sample. Furthermore, I also use the exact amount of children a parent has as a treatment variable, as opposed to solely using a dummy indicating whether or not a parent has more than two children. This makes it easier to discern what the effects on labour market outcomes are per additional child. I also perform slightly different sample restrictions. Where Iacovou focusses on women between 21 and 49 years old, I focus on parents whose youngest child is at most 17 years old. The academical relevance of my analysis therefore lies in the ways it expands upon the work already done by Iacovou.

III. Data

III.A. British Household Panel Survey

I use data gathered by the British Household Panel Survey to provide an answer to my research question. This survey was produced by the Institute for Social and Economic Research, University of Essex with the support of the Economic and Social Research Council. The research data are distributed by the UK Data Service. The sample of this survey contained approximately 10,000 individuals who were interviewed each year over the period 1991-2008. The composition of the sample was aimed to be representative of the British population. Observations concern individuals, not households. This means that it is visible in the data if one partner in a household has a value for a certain variable that differs from the value of the other partner, for example if one has a job and the other doesn't. Some questions were not asked each year, which means that some variables are only available in a couple of individual waves. Examples of such variables are variables containing the sex and sequence of birth of the children of an individual. Data on these variables is only available for waves two, eleven and twelve of the survey. These waves correspond to the years 1992, 2001 and 2002, respectively. These specific variables are necessary for the method I use in my research, which will become clearer when I discuss my method. For this reason I only use data from waves two, eleven and twelve.

III.B. Sample selection and restrictions

Iacovou (2001) faced problems caused by small sample sizes when she used the data I use in my analysis. I try to resolve these problems by pooling all observations of the mentioned three waves into one larger sample. This doesn't imply that there are parents in my sample of whom I use multiple observations from different periods. This has to do with how some questions from the survey were asked. For my analysis I use variables denoting the sex of a parent's children and the sequence in which they were born. Each parent only answered the questions used to derive the values for these variables once. For example, parents who answered these specific questions in 1992 didn't do so in 2001 and 2002. As a consequence, I have only wave in which the value of the variables derived using these questions is non-missing for a specific parent. I therefore only use one observation for each parent.

I restrict my sample to certain kinds of parents. Firstly, a consequence of the specific method I use is that all parents in both the control and treatment group have at least two children, therefore my sample is restricted to this type of parents. Secondly, I want to avoid including parents whose children are all adults. For my main analysis I therefore only include parents whose youngest child is at most 17 years old. I check whether my results are robust to this choice by performing part of my main analysis on separate samples containing only parents whose youngest child is at most 12, 14 or 19 years old.

III.C. Variables

III.C.I Treatment variables and instrumental variable

I will now introduce the variables that I use in my research, starting with my treatment variables and my instrument. The first treatment variable is the 'Amount of children' variable, which indicates the amount of children a parent has. The average amount of children parents in my sample have is 2.68. This may seem high, but remember that my sample only includes parents with at least two children. Note that the averages I mention in this section can also be found in Table 1. The second treatment variable is the 'Three or more children' variable which is a dummy that has value one if an individual has three or more children and zero otherwise. 43.1 percent of my sample has three or more children, meaning 56.9 percent has two children. Using data on the birth sequence and sex of a parent's children I created the 'Same-sex' variable. This is a dummy variable with value one for parents with two or more children of whom the two born first are of the same sex. For parents whose first two children have different sexes this variable has value zero. This variable will function as an instrument for the two treatment variables. 49.4 percent of the individuals in my sample has value one for this instrument.

This is what you would expect if the sex of children is randomly and independently assigned; in that case two boys or two girls are two out of four possible sex combinations of which each is equally likely.

Table 1: Descriptive statistics

	Full	Women	Men	College		Non-college	
				Women	Men	Women	Men
Amount of children	2.677 (0.999)	2.673 (0.967)	2.681 (1.039)	2.475 (0.756)	2.609 (0.895)	2.709 (0.996)	2.699 (1.071)
Three or more children	0.431	0.438	0.423	0.344	0.416	0.455	0.425
Same-sex	0.494	0.493	0.495	0.481	0.487	0.494	0.496
Employed	0.702	0.601	0.832	0.772	0.932	0.573	0.807
Weekly hours worked	23.452 (20.046)	14.983 (15.605)	34.348 (19.867)	20.620 (15.791)	36.045 (16.514)	14.048 (15.381)	33.964 (20.589)
Gross labour income	795.03 (1049.73)	439.86 (720.25)	1251.98 (1212.21)	979.97 (1019.40)	1874.11 (1351.35)	350.13 (613.40)	1103.91 (1133.75)
Female	0.563	1.00	0.00	1.00	0.00	1.00	0.00
College	0.166	0.144	0.194	1.00	1.00	0.00	0.00

Notes. This table provides descriptive statistics for my dataset. The columns present different groups within my sample, and the rows present different variables. Cells provide the average of the row variable for the column group. For binary variables this group average is the proportion of that group that has value one for that variable. Standard deviations are reported within brackets for non-binary variables. 'Amount of children' is a non-binary variable that indicates the amount of children an individual has. 'Three or more children' is a binary variable with value one if an individual has three or more children. 'Same-sex' is a binary variable with value one if an individual's first two children are of the same sex, the variable is missing for individuals with less than two children. 'Employed' is a binary variable with value one if an individual is employed, self-employed or on maternity leave. 'Weekly hours worked' is a non-binary variable, it indicates the amount of hours worked each week, the variable has value zero for unemployed individuals. 'Gross labour income' is a non-binary variable, for employed individuals it indicates the gross labour income earned each month, the variable has value zero for unemployed individuals. 'Female' is a binary variable that takes value one if an individual is of the female sex. 'College' is a binary variable that takes value one for individuals who obtained my proxy for a college degree.

III.C.II Outcome variables

In my analysis three separate outcome variables are of interest. The first is 'Employed', a dummy variable with value one if a parent is either employed, self-employed or on maternity leave. 70.2 percent of my sample indicated to belong to one of these three groups. My second outcome variable is 'Weekly hours worked', which contains the amount of hours a parent indicated to usually work each week, it has value zero for unemployed parents. In my sample the average value for this variable is about 23.5 hours a week. My final outcome variable is 'Monthly labour income', which contains the gross labour income, denoted in pounds, a parent indicated to earn each month. This variable also has value zero for unemployed individuals. Some self-employed individuals only reported gross monthly profit, in those cases I use this value as a proxy for labour income. In my sample the average value for this variable is about 795 pounds a month.

III.C.III Other variables

Several other variables are useful for my analysis. The first of these is a created variable indicating the age of the youngest child of a parent, which I use to perform sample restrictions. The second is a variable indicating the sex of a parent. I use this variable to separately investigate effects for men and women, as I will describe in my methodology section. Note that 56.3 percent of my sample indicated to be female. This might indicate an overrepresentation of women in my sample. I also generated a variable that functions as a proxy for whether or not a parent obtained a college degree. To do so I used a variable indicating a parent's highest academical achievement. If an individual stated to have obtained a '1st degree', 'higher degree', 'HNC qualification' or 'HND qualification', then I classify this individual as having a college degree. 16.6 percent of my sample has value one for this dummy variable.

III.D. Descriptive statistics

Table 1 presents descriptive statistics for my dataset. The rows each present a different variable and the columns each present a different subsample. Each cell shows the group mean for the specific variable of that row. For dummy variables this average can be interpreted as a proportion. I would like to point the reader to the apparent gender gaps in the outcome variables. The percentage of men that is employed is higher than the corresponding percentage of women. This can also be said about the average monthly labour income and the average amount of hours worked each week. The fraction of men that went to college is also higher than the corresponding fraction for women. These gender gaps provide a rationale for separately investigating treatment effects for men and women.

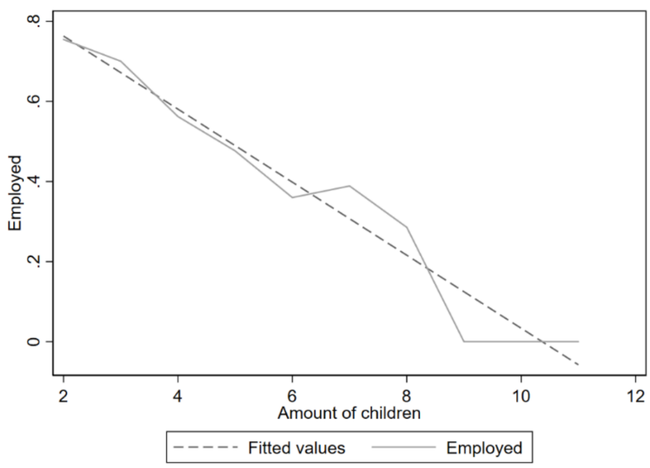

Figure 1: Fraction of all parents with a certain amount of children that is employed

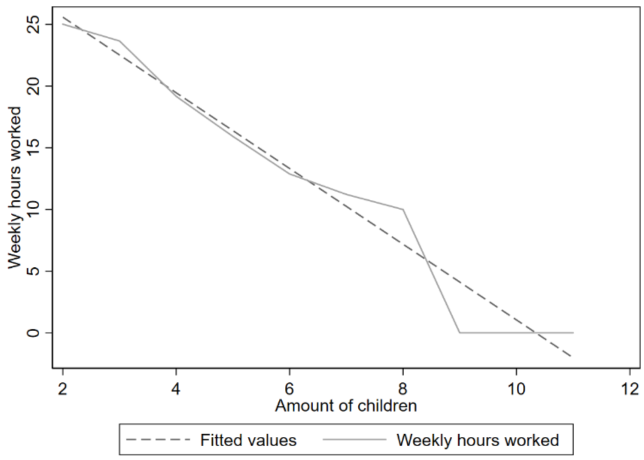

Figure 2: Average amount of hours worked per week, by all parents with a certain amount of children

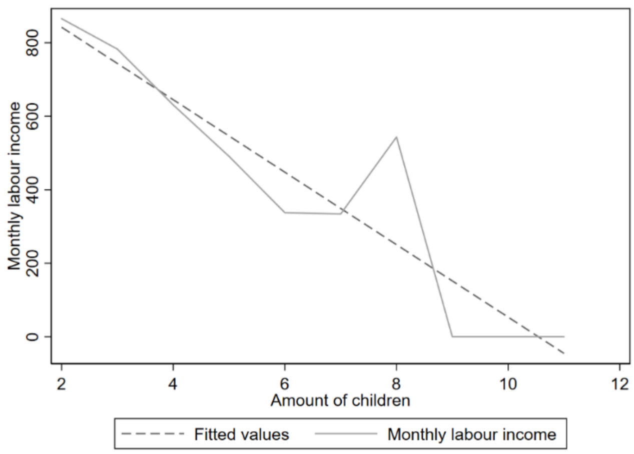

Figure 3: Average gross labour income earned per month, by all parents with a certain amount of children

Figure 1 plots the fraction of all parents with a certain amount of children that is employed, for each amount of children. The fitted line shows us a negative correlation between the amount of children parents have and the probability that they have a job. Figure 2 plots the average value of the 'Weekly hours worked' variable for each value of the 'Amount of children' variable. Figure 3 does the same but uses the 'Monthly labour income' variable as an outcome variable instead. The fitted lines in both figures also indicate a negative correlation between both outcome variables and the treatment variable. Note that these negative correlations are no evidence of negative causal effects, since we do not know whether or not there are cofounders or problems with reverse causality.

IV. Methodology

IV.A. Description of method and regression equations

The structure of my dataset makes several methods feasible. I could investigate the effect of fertility on a particular labour market outcome by estimating a linear regression with the labour outcome as a dependent variable and the amount of children as a treatment variable. This could lead to problems with confounders, both observable and unobservable. The observable confounders, for example education, I could then control for. The unobservable confounders would, however, remain a problem. It is not unlikely that there are unobservable confounders, potential examples are someone's upbringing, intelligence and character. All of these are hard to measure and might affect both the number of children someone has and their labour market outcomes. Reverse causality, if labour market outcomes influence the amount of children someone has, might also be problematic. A simple linear regression cannot deal with this problem. I could instead use the individual fixed effect method, which uses the panel nature of the dataset. This method requires enough within-individual variation in fertility. This would get rid of time-invariant unobserved confounders. However, time-varying unobservable confounders and potential reverse causality would remain problematic.

I therefore use the instrumental variable method in my analysis. I copy the approach of Angrist and Evans (1998); using whether or not the first two children a parent had have the same sex as an instrument for the amount of children that parent has. The preferences of parents are a potential mechanism through which the instrument affects the amount of children a parent has. Parents may prefer to have both a boy and a girl. In that case parents with either two boys or two girls have an additional rationale to have a third child as opposed to parents who have a boy and a girl, assuming that the types of parents are equal in all other aspects affecting fertility choices. This would make it more likely that parents whose first two children were of the same sex have a third child.

The first step of my methodology is to derive the value for the 'Same-sex' variable for the parents remaining after I perform my sample restrictions. This variable is used to predict values for the 'Amount of children' and 'Three or more children' variables. These predictions are then used as values for these two variables in separate regressions where one of the two functions as a treatment variable. Both variables have their relative merits as a treatment variable; 'Amount of children' gives estimates of the effect per additional child and 'Three or more children' often has a stronger first stage. The latter claim will become clearer when we discuss the validity of the instrument.

Besides using the instrumental variable approach to counter potential endogeneity, I also control for the effects of the following factors: age of the parent; whether the parent is disabled; sex of the first child of a parent and the marital status of the parent. I add the sex of a parent's first child as a control variable because excluding it may lead to biased estimates. If the probability that a child is male is higher or lower than the probability that a child is female, then parents who had a boy first are more or less likely to have two children of the same sex. This leads to bias if the sex of a parent's first child influences labour market outcomes. Because I pool data from three waves, there will be some parents who answered the relevant questions in 1992, some who answered them in 2001 and some who answered them in 2002. To my regressions I also add a dummy for each of these years, any one of which takes value one if an individual answered the questions in that specific year.

I separately estimate my equations for several subsamples: women; men; college-educated women; college-educated men; women without a college degree and men without a college degree. Obtaining separate estimates for men and women is useful, because Becker's theory suggests that the treatment effects might differ between men and women. Doing so might also help explain the gender wage gap. Obtaining separate estimates for college educated parents and parents without a college education is useful, because Willis's theory might suggest that treatment effects differ with educational attainment. Testing the validity of their predictions is a way to partially test the validity of these theories. This is important, because these models may point out mechanisms behind potential treatments effects.

The described approach can be summarized by the following two equations:

1. $Y_i = \alpha + \beta\hat{T}_i + \Gamma X_i + \varepsilon_i$
2. $\hat{T}_i = \mu + \delta Z_i + \Gamma X_i + \sigma_i$

The first equation represents the second stage and the second equation represents the first stage. Y_i is the value of the outcome variable, for individual i. This outcome variable can be any of the three explained outcome variables. Both α and μ are non-individualized constants. Z_i is the value of the 'Same-sex' instrument, for individual i. $\hat{T}_i$ is the predicted value of the treatment variable, T, as predicted by the instrument, for individual i. X_i is a vector including the described control variables. If 'Three or more children' is used as a treatment variable, then δ can be interpreted as the percentage point change in the probability that a parent has more than two children if the first two children of an individual are of the same sex instead of different sexes. If 'Amount of more children' is used as a treatment variable, then δ can be interpreted as the change in the number of children that a parent has if his or her first two children are of the same sex instead of different sexes. β is the coefficient of interest. Its interpretation differs with the used outcome variable, with the used treatment variable and with the used subsample. If 'Employed' is used as dependent variable, then the interpretation is as such: if a parent has an additional child or additional children (depending on the treatment variable) because his or her first two children are of the same sex, then the probability that this parent has a job increases or decreases (depending on the sign of the coefficient) by $\beta * 100$ percentage points, conditional on the parent being part of the particular subsample to which β is applicable. The interpretation is similar if either 'Weekly hours worked' or 'Monthly labour income' is used as an outcome variable. However, the interpretation of β is then the increase or decrease in weekly hours worked or gross monthly labour income, respectively.

IV.B. Assessing the Instrument

IV.B.I Strong first stage

In this section I discuss the validity of the assumptions underlying the instrumental variable approach for my analysis. Firstly, the instrument must have a strong first stage, a strong causal effect on the treatment variable. I test whether there is at least a strong correlation by running regressions with the treatment variable as a dependent variable and the instrument as a treatment variable. The results of these regressions can be found in Table 2. Note that I ran separate first stage regressions for the separate subsamples I will use in my actual analysis and for my separate treatment variables. The regressions in Table 2 also include the control variables that my second stage regressions will include.

Table 2: First stage regressions

Panel A, Dependent Variable: Amount of children

	Full Sample	Women	Men	College		Non-College	
				Women	Men	Women	Men
Same-sex	0.150*** (0.030)	0.161*** (0.038)	0.126*** (0.045)	0.065 (0.080)	0.138 (0.096)	0.174*** (0.043)	0.117** (0.051)
Controls	YES	YES	YES	YES	YES	YES	YES
Observations	4,169	2,346	1,823	336	353	1,999	1,463
F-statistic instrument	25.90	17.70	7.78	0.67	2.06	16.82	5.23

Panel B, Dependent Variable: Three or more children

	Full Sample	Women	Men	College		Non-College	
				Women	Men	Women	Men
Same-sex	0.104*** (0.015)	0.118*** (0.020)	0.082*** (0.022)	0.036 (0.051)	0.113** (0.052)	0.131*** (0.022)	0.070*** (0.025)
Controls	YES	YES	YES	YES	YES	YES	YES
Observations	4,169	2,346	1,823	336	353	1,999	1,463
F-statistic instrument	48.14	34.92	13.25	0.48	4.71	36.45	7.88

Notes. This table provides the results of first stage regressions. In Panel A of the Table 'Amount of children' is the dependent variable. This is a non-binary variable indicating the amount of children an individual has. In Panel B of the table 'Three or more children' is the dependent variable. This is a binary variable with value one if an individual has more than three children. In both panels 'Same-sex' serves as the treatment variable, this is a binary variable with value one for individuals whose first two children were of the same sex. The variable is missing for individuals with less than two children. The different columns show results obtained using different subsamples. The regressions in all columns include a full set of controls. Standard errors of estimates are reported within brackets. The interpretation of the significance stars is as such: *p<0.10, **p<0.05 and ***p<0.01

There is no hard definition of a 'strong causal effect', however, Stock, Wright and Yogo (2002) recommend that the F-statistic of the coefficient of the instrument in the first-stage regression should exceed ten. In Table 2 it can be seen that this is the case when I use my whole sample, my subsample containing only women or my subsample containing only women without a college degree. In these cases it doesn't matter whether I use 'Three or more children' or 'Amount of children' as a dependent variable. The F-statistic also exceeds ten if I use my subsample containing only men and use 'Three or more children' as my dependent variable. In all other cases the F-statistic doesn't exceed ten although the coefficient is still significant in some other estimations. Namely if I use the subsample containing men and use 'Amount of children' as a dependent variable; if I use the subsample containing only men without a college education or if I use the sample containing only college educated men and use 'Three or more children' as a dependent variable. In the remaining three specifications the coefficient of the instrument isn't significant at any of the conventional levels. One should take these first stage results into account when interpreting the estimates made using those specifications for which the first stage is weak. This is important, because a weak instrument may lead to inaccurate estimates of treatment effects and problems with statistical inference, according to Stock et al. (2002).

All of the point estimates in Table 2 are positive. This is according to expectations if the preferences of parents are used to explain the instrument's influence. All significant point estimates are either slightly smaller or larger than ten percentage points which is, arguably, an economically significant increase in the probability that a parent has an additional child. The point estimates are also smaller for men, if only statistically significant estimates are considered. The point estimates for men without a college education and women without a college education aren't even within each other's 95 percent confidence interval if three or more children is used as a dependent variable. It is likely that these gaps are the result of the disproportionate absence of certain types of fathers in the dataset, since children have both a biological mother and father.

IV.B.II Uncorrelated with the error term

The instrument must also be uncorrelated with the error term of equation 1. This implies that the 'Same-sex' variable must be uncorrelated with other determinants of the relevant labour market outcomes. Note that 'other determinants' doesn't include mechanisms through which fertility affects labour market outcomes. We cannot test this assumption. However, one can argue that children's sex at birth is as good as randomly determined. This would imply that a parent's value of the instrument is as good as randomly determined, excluding correlations with labour market outcomes' other determinants. The latter claim is only true if another assumption, the exclusion restriction, also holds.

Sex-selective abortions might cause non-random assignment of sex to a child. These are abortions where it is decided to abort based on an infant's sex. This might cause a correlation between a parent's characteristics and the sex combination of this parent's children, possibly causing a correlation between the instrument and the error term. I suspect that sex-selective abortion isn't very prevalent and that this is a problem in only a few cases. However, I don't have data on whether and why parents had abortions performed previously. Therefore, I cannot prove my suspicion, which means caution is necessary when interpreting results.

The exogeneity assumption cannot be tested. We can, however, test for a correlation between the instrument and certain observables in the dataset. The results of such a 'balance test', which was performed using the same sample as is used in my actual analysis, can be found in Table 3. The results give reason for concern, although the coefficients aren't jointly significant as can be derived from the F-statistic. The reason for concern is that the coefficient of a variable containing the year of birth of a parent is significant at a significance level of ten percent. This doesn't immediately imply a correlation between instrument and error term, since I add this year of birth variable as a control variable to my regressions. However, this result shows us that it is possible that the instrument is correlated with variables that are not influenced by the instrument. The found correlation therefore raises concern that there might be non-observable confounders which bias estimates. This balance test and its results should therefore make me more cautious when I interpret the results of my main analysis.

Table 3: Balance test

Female	College	Disabled	Marital status	Year of birth	F-statistic	Observations
0.004 (0.016)	-0.011 (0.021)	-0.055 (0.038)	0.001 (0.006)	-0.002* (0.001)	1.15	4,151

Notes. This table provides the results of a balance test. 'Same-sex' is used as a dependent variable. This is a binary variable that takes value one for individuals whose first two children were of the same sex. This binary variable is not defined for individuals with less than two children. The balance test includes several independent variables, these are: 'Female', a binary variable with value one for individuals who are female; 'College', a binary variable with value one for individuals who obtained my proxy for a college degree; 'Disabled', a binary variable with value one for individuals who are disabled; 'Marital status', a binary variable indicating whether an individual is married and 'Year of birth', a non-binary variable indicating the year of birth of an individual. For lay-out purposes rows and columns are inversed. This means the first five columns each present a different independent variable. The coefficients of these variables are reported in the second row. The sixth and seventh column present the F-statistic of a test for joint significance and the amount of observations used, respectively. Standard errors of estimates are reported within brackets. The interpretation of the significance stars is as such: *p<0.10, **p<0.05 and ***p<0.01

It is unclear how a correlation between the instrument and a variable that cannot be causally influenced by the instrument is possible. A potential explanation is that sex-selective abortions are more prevalent in my dataset than I suspected. A second potential explanation lies in measurement error. My dataset was developed using questionnaires, meaning all variables contain self-reported values for individuals. This could be prone to errors, which might cause a correlation between the instrument and, for example, a parent's age. A potential reason might be that older parents sometimes state that their first two children have differing sexes, even when this is untrue. If younger parents do not make such mistakes, then this could lead to the kind of correlation detected by the balance test. I will discuss potential problems caused by measurement error more rigorously in my discussion section.

IV.B.III Exclusion restriction

The third assumption underlying the instrumental variable approach is the exclusion restriction. In the context of this analysis it entails that the 'Same-sex' variable may have no direct effect on the outcome variables. Such an effect would otherwise be included in the estimate, thereby leading to bias. The instrument affecting the outcome variable via the treatment variable is, of course, allowed under this assumption, as this would be an indirect effect. This assumption is also untestable. However, it can be falsified in some situations, if there is a group for whom the instrument has no or a very small first stage effect. According to the exclusion restriction, the estimate of the treatment effect obtained using the 'Same-sex' variable as an instrument, should not significantly differ from zero for these groups. The exclusion restriction is falsified if the estimate is significant. Remember that failing to falsify the exclusion restriction is not the same as proving that it holds.

Several of my subsamples have first stages that aren't strong, which I discussed above. The least significant estimates are those made using the sample containing only college educated women and using either of the two treatment variables. This can be checked using Table 2, which shows that the F-statistics of tests for statistical significance of these coefficients are 0.67 and 0.48. Estimates of the treatment effect made using this subsample and using the 'Same-sex' instrument should therefore be statistically insignificant, according to the exclusion restriction. However, the sample containing solely college educated women contains only 336 observations. This makes it unclear whether the weak first stage estimates result from the actual absence of a strong first stage for college educated women, or because the small subsample leads to imprecise estimates. The exclusion restriction remains untestable and caution when interpreting results remains necessary. However, I think it is not an unreasonable assumption to make in this specific context. It is hard to think of reasons why the sex combination of an parents' first two children would directly influence their labour market outcomes.

IV.B.IV Monotonicity assumption and further remarks

The final assumption that needs to be met is the monotonicity assumption. For my analysis this assumption holds if there are no individuals who chose not to have a third child because their first two children were of the same sex and if there are no individuals who chose to have a third child because their first two children did not have the same sex. The instrumental variable approach does not lead to unbiased estimates of the local average treatment effects if the monotonicity assumption is not fulfilled.

The instrumental variable approach has a final limitation. It doesn't necessarily provide an estimate of either the average treatment effect or the average treatment effect on the treated, even if all discussed assumptions hold. My approach tries to provide an estimate of the treatment effect by comparing two groups: a group of parents whose first two children have the same sex and a group of parents whose first two children have different sexes. Simply put, it compares the average amount of children (or the proportion of the group with more than three children) of the two groups and compares the average value of the outcome variable for the two groups. Then it divides the difference in the average value of the outcome variable by the difference in the average value of the treatment variable. In doing so it obtains an estimate of the average change in the outcome variable resulting from a unit change in the treatment variable.

There is, however, only one specific group of parents whose amount of children differs between the two groups with different values for the instrument. This means the estimate is based on one specific group of parents. These are the so-called 'compliers'. Compliers are the parents who have a third child because their first two children are of the same sex and who would not have a third child if their first two children were of differing sexes. The difference in the average value of the treatment variable between the two groups with different values for the instrument and the difference in the average value of the outcome variable between the two groups with different values for the instrument are both caused by this group of compliers. The estimate of the treatment effect is therefore an average of the individual effects for these 'compliers'. This is not necessarily the same as the average effect for every individual in the population or even sample. It might, for example, be the case that the treatment effect is larger for compliers, which means that the estimate made using an instrumental variable approach would be larger than the average treatment effect. When interpreting the results it is therefore important to keep in mind that these are local average treatment effects, 'LATE's'.

V. Results

V.A. The effect on the probability of being employed

Table 4 shows estimates of the effect that increased fertility has on the value of the 'Employed' variable. These can also be interpreted as estimates of the effect increased fertility has on the probability that a parent is employed. The table consists of two panels. The estimates in the first panel were made using 'Three or more children' as a treatment variable and the estimates in the second were made using 'Amount of children' as a treatment variable. Different columns within Table 4 present estimates made using these different subsamples. There are two columns with estimates made using the full sample, one including and one excluding control variables.

The following is an example interpretation of one of the estimates: if a woman in my sample has additional children because her first two children are of the same sex, this on average decreases the probability that this woman is employed by 10.1 percentage points. The estimates in the different columns of the first panel have similar interpretations, except that the subsamples to which the results apply differ. This also holds for the results in the second panel of the table, with the addition that these estimates are estimates of the effect per additional child. Note that I emphasize the fact that a change in treatment level must be caused by the instrument, in order for the estimate to be an accurate estimate of the treatment effect. This is important because of the potential difference between the local average treatment effect and the average treatment effect.

Table 4: Effect on the probability that an individual is employed

Panel A, Treatment variable: Three or more children								
	Full Sample	**Full Sample**	**Women**	**Men**	**College**		**Non-College**	
					Women	**Men**	**Women**	**Men**
Three or more children	-0.088 (0.130)	-0.125 (0.129)	-0.101 (0.161)	-0.029 (0.201)	-0.211 (1.252)	0.033 (0.243)	-0.088 (0.160)	-0.050 (0.274)
Controls	NO	YES	YES	YES	YES	YES	YES	YES
Observations	4,173	4,168	2,346	1,822	336	353	1,999	1,462
Panel B, Treatment variable: Amount of children								
	Full Sample	**Full Sample**	**Women**	**Men**	**College**		**Non-College**	
					Women	**Men**	**Women**	**Men**
Amount of children	-0.059 (0.087)	-0.086 (0.088)	-0.074 (0.116)	-0.019 (0.130)	-0.116 (0.674)	0.027 (0.201)	-0.066 (0.118)	-0.030 (0.163)
Controls	NO	YES	YES	YES	YES	YES	YES	YES
Observations	4,173	4,168	2,346	1,822	336	353	1,999	1,462

Notes. This table presents estimates made using the instrumental variable method and using 'Employed', a binary variable indicating whether or not an individual has paid work, as a dependent variable. The 'Same-sex' variable, a binary variable indicating whether an individual's first two children were of the same sex, is used as an instrument. In Panel A 'Three or more children' serves as treatment variable. This is a binary variable indicating whether or not an individual has three or more children. In Panel B 'Amount of children' serves as a treatment variable. This is a non-binary variable indicating the amount of children an individual has. Each panel has several columns, each of these columns presents estimates made using a separate subsample. In both 'Full Sample' columns the same sample is used, however, in the first no controls are included. All estimates in columns other than the second were made using a full set of controls. Standard errors of estimates are reported within brackets. The interpretation of the significance stars is as such: *p<0.10, **p<0.05 and ***p<0.01

Most of the point estimates are negative and larger in magnitude for women. Both of these findings fit into the theoretical framework. None of the estimates is statistically significant at any of the conventional levels. This lack of significant estimates is in agreement with Iacovou's (2001) results. The other discussed empirical papers did find statistically significant negative effects of (increased) fertility on the extensive margin of female labour participation. Iacovou ascribes these differences to the differences in the size of her sample and the sizes of the samples of other authors. She argues that the small size of her sample leads to imprecise, and because of that, insignificant estimates.

The limited size of my sample is not an unlikely explanation for the differences in statistical significance of my results and the results of the empirical papers in which significant effects are found for women. After all, each of these papers use samples which are several times larger than my largest sample. The smallest subsample found in these papers still consists of about 12,000 observations (Cools et al., 2017). Furthermore, it is not the case that I obtain accurate estimates of a null effect. This becomes clear when looking at the estimate made using my full sample, which use as many observations as possible, and made using 'Amount of children' as a treatment variable. The 95 percent confidence interval of this estimate still ranges from a decrease of 25.9 percentage points to an increase of 8.6 percentage points. This cannot be considered a narrow confidence interval, seeing as the average, minimum and maximum value of the 'Employed' variable are 70.2 percent, 0 percent and 100 percent, respectively. Furthermore, one could argue that economically significant effects are within the confidence interval. For example, if having an additional child decreases the probability of being employed by 20 percentage points then this would be economically significant given that employment rate is 70.2 percent in my sample. This increases my suspicion that my lack of significant estimates is caused by my small sample size, and not because treatment effects are actually zero. Differences between my results and those obtained by Lundborg et al. (2017) may also be caused by the fact that they estimate the effect that having a first child has on labour market outcomes. I obtain estimates of the effects that having additional children or an additional child has, instead. Finally, I must mention that a lack of statistically significant estimates of an effect on the extensive margin of female labour participation doesn't necessarily mean that my results do not correspond with the predictions made in the theoretical framework. After all lower female labour participation might still be visible on the intensive margin, the amount of hours worked each week by mothers.

It might also be the case that the reason that my results differ from the results of the discussed empirical papers is that the treatment effects differ between different countries. However, the discussed papers focus on several types of countries: more capitalist societies like the United States; countries with strong welfare systems like Norway and developing countries like Argentina and Mexico. The fact that significant negative effects for women were found in each of these different types of countries does raise suspicion that effects are also present in the United Kingdom. In the following sections I discuss estimates of the treatment effect on my other outcome variables. My results will again partly deviate from those of other papers. A lack of external validity of these other results may again provide a rationale for these differences. However, this remains an unlikely explanation. Therefore, I won't repeat this potential explanation.

V.B. The effect on the amount of hours worked each week

Table 5 contains estimates of the effect of increased fertility on the 'Weekly hours worked' variable. The layout of the table is exactly equal to the layout of Table 4; with the same setup of panels and columns. The interpretation of the results is also similar to the interpretation of the results in Table 4, with exception of the different outcome variable. The interpretation also differs in the same way when switching between panels and columns. The following is an example interpretation of the results in Table 5: if parents in my sample have an additional child because their first two children are of the same sex, this on average decreases the amount of hours these parents work each week by 2.7 hours. In this interpretation I used the estimate made using the full sample and including controls.

Table 5: Effect on the amount of hours worked each week

Panel A, Treatment variable: Three or more children								
	Full Sample	**Full Sample**	**Women**	**Men**	**College**		**Non-College**	
					Women	**Men**	**Women**	**Men**
Three or more children	-2.059 (5.740)	-3.912 (5.730)	-0.274 (5.208)	-2.303 (10.987)	32.345 (69.937)	7.365 (15.816)	-1.362 (5.023)	-3.515 (14.632)
Controls	NO	YES	YES	YES	YES	YES	YES	YES
Observations	4,173	4,168	2,346	1,822	336	353	1,999	1,462
Panel B, Treatment variable: Amount of children								
	Full Sample	**Full Sample**	**Women**	**Men**	**College**		**Non-College**	
					Women	**Men**	**Women**	**Men**
Amount of children	-1.387 (3.850)	-2.700 (3.925)	-0.200 (3.803)	-1.495 (7.097)	17.815 (37.035)	6.058 (13.533)	-1.019 (3.735)	-2.104 (8.700)
Controls	NO	YES	YES	YES	YES	YES	YES	YES
Observations	4,173	4,168	2,346	1,822	336	353	1,999	1,462

Notes. This table presents estimates made using the instrumental variable method and using the 'Weekly hours worked' variable as a dependent variable. This is a non-binary variable indicating the amount of hours worked each week by an individual, the variable has value zero for unemployed individuals. The 'Same-sex' variable, a binary variable indicating whether an individual's first two children were of the same sex, is used as an instrument. In Panel A 'Three or more children' serves as treatment variable. This is a binary variable indicating whether or not an individual has three or more children. In Panel B 'Amount of children' serves as a treatment variable. This is a non-binary variable indicating the amount of children an individual has. Each panel has several columns, each of these columns presents estimates made using a separate subsample. In both 'Full Sample' columns the same sample is used, however, in the first no controls are included. All estimates in columns other than the second were made using a full set of controls. Standard errors of estimates are reported within brackets. The interpretation of the significance stars is as such: *p<0.10, **p<0.05 and ***p<0.01

Most estimates in Table 5 are negative, which can be fitted into the theoretical framework. None of the estimates differ significantly from zero at any of the conventional levels, which is in accordance with the results obtained by Iacovou (2001). It is partly in contrast with Angrist and Evans (1998), Cools et al. (2017) and Lundborg et al. (2017), as in these papers significant negative effects were found for women on the intensive margin of labour supply. The lack of significant estimates also doesn't correspond to the predictions made in the theoretical framework; the kind of specialization described there should lead to a negative effect of fertility on hours worked for at least some subgroups.

A potential explanation for the explained differences is that my smaller sample size leads to less precise estimates. An indication of this is that I don't obtain accurate estimates of a null effect either. The confidence intervals of my estimates are again broad and include economically significant effects. This is clearest if I focus on my subsample containing only women, in which case the confidence intervals deviate the least from zero either way. Even in this specification the 95 percent confidence interval of the treatment effect ranges from a decrease of approximately 10.5 hours worked a week to an increase of approximately 9.9 hours worked a week, if 'Three or more children' is used as a treatment variable. The equivalent confidence interval if 'Amount of children' is used as treatment variable ranges between a decrease of about 7.7 hours a week and an increase of about 7.3 hours a week. These confidence intervals include both economically significant increases and decreases, considering the fact that in my sample the average amount of hours worked each week is approximately 23.5. The difference between my results and those of Lundborg et al. could also be caused by the fact that they use a different treatment variable.

V.C. The effect on monthly labour income

Table 6 shows estimates of the effect of increased fertility on the 'Monthly labour income' variable. The layout of the table is exactly equal to the layout of Table 4; with the same setup of panels and columns. The interpretation of the results is also similar to the interpretation of the results in Table 4, with exception of the different outcome variable. The interpretation also differs in the same way when switching between panels and columns. The following is an example interpretation of the estimates: if parents in my sample have an additional child because their first two children are of the same sex, this on average decreases the labour income these parents earn each month by 318.96 pounds. In this interpretation I used the estimate made using the full sample and including controls.

Table 6: Effect on the gross labour income earned each month

Panel A, Treatment variable: Three or more children

	Full Sample	Full Sample	Women	Men	College		Non-College	
					Women	Men	Women	Men
Three or more children	-350.61 (301.38)	-462.09 (303.04)	-244.26 (230.48)	-591.70 (693.85)	-3580.21 (5518.66)	101.91 (1244.28)	-97.560 (194.56)	-833.21 (863.85)
Controls	NO	YES	YES	YES	YES	YES	YES	YES
Observations	4,173	4,168	2,346	1,822	336	353	1,999	1,462

Panel B, Treatment variable: Amount of children

	Full Sample	Full Sample	Women	Men	College		Non-College	
					Women	Men	Women	Men
Amount of children	-236.16 203.09	-318.96 (209.43)	-178.61 (167.84)	-384.05 (449.89)	-1971.96 (2634.20)	83.822 (1028.81)	-73.000 (144.84)	-498.70 (518.32)
Controls	NO	YES	YES	YES	YES	YES	YES	YES
Observations	4,173	4,168	2,346	1,822	336	353	1,999	1,462

Notes. This table presents estimates made using the instrumental variable method and using 'Gross labour income' as a dependent variable. This is a non-binary variable indicating the gross labour income earned each month by an individual, the variable has value zero for unemployed individuals. The 'Same-sex' variable, a binary variable indicating whether an individual's first two children were of the same sex, is used as an instrument. In Panel A 'Three or more children' serves as treatment variable. This is a binary variable indicating whether or not an individual has three or more children. In Panel B 'Amount of children' serves as a treatment variable. This is a non-binary variable indicating the amount of children an individual has. Each panel has several columns, each of these columns presents estimates made using a separate subsample. In both 'Full Sample' columns the same sample is used, however, in the first no controls are included. All estimates in columns other than the second were made using a full set of controls. Standard errors of estimates are reported within brackets. The interpretation of the significance stars is as such: *p<0.10, **p<0.05 and ***p<0.01

Most of the estimates in Table 6 are negative, which can be explained using the theoretical framework. However, none are statistically significant. This is partly in contrast with the results obtained by Angrist and Evans (1998). They do find significant negative effects on women's labour income. It is also difficult to fit my lack of significant results into the theoretical framework, which does predict negative effects on the labour income of some subgroups. My small sample size once more provides a potential explanation for these differences between my results and the results of other papers, since I don't obtain accurate estimates of a null effect either. This is clearest from the 95 percent confidence interval of the estimate made using my sample of only women without a college education and using 'Amount of children' as a treatment variable. This is the confidence interval which deviates the least from zero either way, but still it ranges from a decrease of about 357 pounds a month to an increase of about 211 pounds a month. This is a relatively broad interval containing both economically significant increases and decreases, considering the fact that the average monthly labour income is only about 350 pounds for my sample containing only women without a college education and only about 795 pounds for my full sample.

V.D. Robustness and discussion exclusion restriction

I mentioned before that the exclusion restriction is falsified if I obtain statistically significant estimates using a subsample for which the first stage is weak or non-existent. It is clear that I have no such significant estimates, since none of my estimates are statistically significant. Therefore, I cannot falsify the exclusion restriction. This is not the same as proving it holds.

The estimates I discussed before were made using a sample restricted to parents whose youngest child is at most 17 years old. In this section I investigate whether estimates made using my full sample and including control variables are robust to this sample restriction. I do so by separately re-estimating these regression equations using a sample restricted to parents whose youngest child is at most 12, 14 or 19 years old. The results of this robustness analysis can be found in Table 7, Table 8 and Table 9. Each of these tables presents estimates for a different outcome variable. The point estimates are all negative, which was also the case in the original sample specification. The coefficients remain statistically insignificant and the estimates made using different samples are all within each other 95 percent confidence intervals. The point estimates do differ with which sample is used and it might be the case that the actual coefficients do as well. It is unwarranted to make a more definitive conclusion as the problem of imprecise estimates is also apparent here.

Table 7: Robustness test of effect on the probability that an individual is employed

	Panel A, Treatment variable: Three or more children			Panel B, Treatment variable: Amount of children		
	Under 13	**Under 15**	**Under 20**	**Under 13**	**Under 15**	**Under 20**
Three or more children	-0.061 (0.158)	-0.016 (0.141)	-0.130 (0.141)	-	-	-
Amount of children	-	-	-	-0.041 (0.106)	-0.011 (0.097)	-0.093 (0.100)
Controls	YES	YES	YES	YES	YES	YES
Observations	3,160	3,582	4,554	3,160	3,582	4,554

Notes. This table presents estimates made using the instrumental variable method and using 'Employed', a binary variable indicating whether or not an individual has paid work, as a dependent variable. The 'Same-sex' variable, a binary variable indicating whether an individual's first two children were of the same sex, is used as an instrument. In Panel A 'Three or more children' serves as treatment variable. This is a binary variable indicating whether or not an individual has three or more children. In Panel B 'Amount of children' serves as a treatment variable. This is a non-binary variable indicating the amount of children an individual has. Each panel has several columns, each of these columns presents estimates made using a separate sample. In the second and fifth columns only parents whose youngest child is younger than thirteen years old are taken into account. In the third and sixth columns only parents whose youngest child is younger than fifteen years old are taken into account. The fourth and seventh columns report estimates made using only those parents whose youngest child is younger than 20 years old. All reported regressions include a full set of controls. Standard errors of estimates are reported within brackets. The interpretation of the significance stars is as such: *p<0.10, **p<0.05 and ***p<0.01

Table 8: Robustness test of effect on the amount of hours worked each week

	Panel A, Treatment variable: Three or more children			Panel B, Treatment variable: Amount of children		
	Under 13	**Under 15**	**Under 20**	**Under 13**	**Under 15**	**Under 20**
Three or more children	-5.489 (6.799)	-2.231 (6.126)	-5.845 (6.279)	-	-	-
Amount of children	-	-	-	-3.693 (4.541)	-1.533 (4.188)	-4.188 (4.470)
Controls	YES	YES	YES	YES	YES	YES
Observations	3,160	3,582	4,554	3,160	3,582	4,554

Notes. This table presents estimates made using the instrumental variable method and using 'Weekly hours worked' as a dependent variable. This is a non-binary variable indicating the amount of hours worked each week by an individual, it has value zero for unemployed individuals. The 'Same-sex' variable, a binary variable indicating whether an individual's first two children were of the same sex, is used as an instrument. In Panel A 'Three or more children' serves as treatment variable. This is a binary variable indicating whether or not an individual has three or more children. In Panel B 'Amount of children' serves as a treatment variable. This is a non-binary variable indicating the amount of children an individual has. Each panel has several columns, each of these columns presents estimates made using a separate sample. In the second and fifth columns only parents whose youngest child is younger than thirteen years old are taken into account. In the third and sixth columns only parents whose youngest child is younger than fifteen years old are taken into account. The fourth and seventh columns report estimates made using only those parents whose youngest child is younger than 20 years old. All reported regressions include a full set of controls. Standard errors of estimates are reported within brackets. The interpretation of the significance stars is as such: *p<0.10, **p<0.05 and ***p<0.01

Table 9: Robustness test of effect on the gross labour income earned each month

	Panel A, Treatment variable: Three or more children			Panel B, Treatment variable: Amount of children		
	Under 13	**Under 15**	**Under 20**	**Under 13**	**Under 15**	**Under 20**
Three or more children	-614.58 (377.22)	-507.04 (325.83)	-480.84 (328.80)	-	-	-
Amount of children	-	-	-	-413.44 (286.76)	-348.53 (224.83)	-344.51 (237.32)
Controls	YES	YES	YES	YES	YES	YES
Observations	3,160	3,582	4,554	3,160	3,582	4,554

Notes This table presents estimates made using the instrumental variable method and using 'Gross labour income' as a dependent variable. This is a non-binary variable indicating the gross labour income earned each month by an individual, it has value zero for unemployed individuals. The 'Same-sex' variable, a binary variable indicating whether an individual's first two children were of the same sex, is used as an instrument. In Panel A 'Three or more children' serves as treatment variable. This is a binary variable indicating whether or not an individual has three or more children. In Panel B 'Amount of children' serves as a treatment variable. This is a non-binary variable indicating the amount of children an individual has. Each panel has several columns, each of these columns presents estimates made using a separate sample. In the second and fifth columns only parents whose youngest child is younger than thirteen years old are taken into account. In the third and sixth columns only parents whose youngest child is younger than fifteen years old are taken into account. The fourth and seventh columns report estimates made using only those parents whose youngest child is younger than 20 years old. All reported regressions include a full set of controls. Standard errors of estimates are reported within brackets. The interpretation of the significance stars is as such: *p<0.10, **p<0.05 and ***p<0.01

V.E. Relationship results and research question

My analysis aimed to provide an answer to the following research question: What is the effect of having additional children on the probability of employment, the amount of hours worked per week and the gross monthly labour income for parents in the United Kingdom? I failed to provide a definitive answer to this question, since I obtained neither statistically significant estimates nor accurate estimates of a null effect. I have to make the latter claim, since economically significant effects were within the 95 percent confidence intervals for all separate outcome variables.

VI. Discussion and threats to identification

VI.A. Problems with methodology and dataset

In this subsection I discuss how the dataset and method I used could lead to problems. A major flaw of my dataset lies in the limited sample size. This lead to imprecise estimates and uncertain results. This could be resolved by using a larger dataset, for example one assembled by a tax agency. My choice of dataset could also lead to problems with measurement error, since individuals have to report their values for variables themselves. This could lead to accidental errors, for example because it might be hard for individuals to remember what their gross monthly labour income was. Errors could also be of a non-accidental nature. For example, because parents might not want to report that they are unemployed and lie instead. It could be that this measurement error is non-random, if specific groups of individuals make specific types of errors. It is unclear in what way this would bias results as it differs per situation. The errors could also be of a random kind. This type of measurement error leads to an understatement of the treatment effect, as it weakens the correlation between the treatment variable and the dependent variable.

Another potential problem with the used data source is caused by attrition. My results may not be applicable to the population of interest if only specific types of individuals keep filling in the surveys. This is because such types of attrition can cause a difference between the composition of the sample and the composition of the population of interest. An indication of such a mismatch can be found in the overrepresentation of women in my sample.

The dataset also didn't allow me to add many control variables to my models, since most available variables could be mechanisms or colliders. With an exogenous instrument this isn't a problem. However, this exogeneity assumption might not hold, since I found a correlation between a parent's age and their value for the instrument. Adding more control variables could then be necessary. The data I used is also quite old. It might be the case that estimates made using data from 1992, 2001 and 2002 aren't relevant anymore. All these problems could be resolved by using a larger, more recent dataset, with precise measurements and more variables to add as controls. Most other papers into the subject that I discussed use data gathered by tax agencies.

Another limitation is that my estimates are estimates of the local average treatment effect. This is not necessarily the same as the average treatment effect if treatment effects are heterogeneous. Furthermore, estimates in this paper are not estimates of the effect that having a first or second child has on the outcome variables. Instead, I obtain estimates of the effect that having an additional child or having additional children has for parents who already have two children. A paper with a different approach might be more suited for readers interested in the effect of a first or second child. For example the paper by Lundborg et al. (2017), in which they use the 'IVF-instrument'.

VI.B. Results and Policy Implications

I cannot give strong policy implications, because my research didn't provide a definitive answer to my research question. Therefore I only want to emphasize that policy makers should not take my lack of statistical estimates as proof of an absence of economically significant effects. After all, I didn't obtain precise estimates of a null effect either, since my confidence intervals were broad and included economically significant estimates. Further research is needed to conclude whether effects are absent or present. It is likely that a better dataset is necessary for such research. I described the characteristics of such a dataset in the previous subsection.

VII. Conclusion

In this paper I obtained estimates of the effects increased fertility has on three different labour market outcomes, for individuals in the United Kingdom. These labour outcomes are: whether or not a parent is employed, self-employed or on maternity leave; for how many hours a week a parent works, and the gross monthly labour income a parent earns. I investigate these effects for several different subgroups using an instrumental variable. The variable I use as an instrument for the fertility of a parent is a dummy variable indicating whether the parent's first two children are of the same sex. This variable is only defined for individuals with at least two children, meaning I am estimating the effect of having additional children on labour market outcomes. I use two separate treatment variables; whether an individual has three or more children and the exact amount of children an individual has.

None of my estimates are statistically significant at any of the conventional levels. This is a difference between my paper and several other investigations into the same effects, where significant effects on certain labour market outcomes are found for specific subgroups. A potential explanation for this difference is the fact that my sample was relatively small. I don't obtain accurate estimates of a null effect either, instead obtaining estimates with wide confidence intervals. Herein lies my main suggestion for future research; use a dataset with a large amount of observations to obtain precise estimates. This would hopefully lead to more definitive conclusions. I used a dataset constructed using surveys, which may lead to measurement error and thereby bias. My second suggestion would therefore be to use a secure data source, for example data from a tax agency. Preferably future research would also use data from more recent years than I did in my research.

Bibliography

Angrist, J., & Evans, W. (1998). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *The American Economic Review, 88*(3), 450-477.

BBC News. (2020, February 26). Childcare costs: Parents of children under two pay 5% more. *BBC News*. https://www.bbc.com/news/uk-51638329

Becker, G. (1981) *A treatise on the family.* National Bureau of Economic Research.

Cools, S., Markussen, S., & Strøm, M. (2017). Children and careers: How family size affects parents' labor market outcomes in the long run. *Demography, 54*(5), 1773-1793. https://doi.org/10.1007/s13524-017-0612-0

Cruces, G., & Galiani, S. (2007). Fertility and female labor supply in Latin America: New causal evidence. *Labour Economics, 14*(3), 565-573. https://doi.org/10.1016/j.labeco.2005.10.006

Iacovou, M. (2001). Fertility and female labour supply (No. 2001-19). ISER Working Paper Series.

Lundborg, P., Plug, E., & Rasmussen, A. (2017). Can Women Have Children and a Career? IV Evidence from IVF Treatments. *The American Economics Review, 107*(6), 1611-1637.

Petter, O. (2020, November 20). Equal Pay Day 2020: What is the gender pay gap and how is it different from equal pay? *The Independent*. https://www.independent.co.uk/life-style/women/gender-pay-gap-equal-pay-women-paid-less-motherhood-a8856121.html

Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics, 20*(4), 518-529. https://doi.org/10.1198/073500102288618658

University of Essex. Institute for Social and Economic Research. (2018). British Household Panel Survey: Waves 1-18, 1991-2009. [data collection]. 8th Edition. UK Data Service. SN: 5151, http://doi.org/10.5255/UKDA-SN-5151-2

Willis, R. J. (1973). A new approach to the economic theory of fertility behavior. *Journal of political Economy, 81*(2), S14-S64. https://doi.org/10.1086/260152