Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis BSc^2 Econometrics and Economics

Assessing the Relationship Between Fertility and family Labour Supply: An Investigation Using Lewbel's Method for Estimating Instrumental Variables With Heteroscedasticity

Quinten van der Schaaff (523193)

Abstract

This study introduces an IV approach, utilising Lewbel (2012)'s method, to investigate the causal effect of fertility on female labour supply. Drawing from census data, the relationship between the number of children and women's labour market participation is examined. Unlike conventional IV strategies, such as the same-sex instrument used by J. Angrist & Evans (1996), Lewbel's method offers a unique advantage as it does not rely on external instruments. Combined with an external IV, this feature can generate more accurate and robust estimates. Our results consistently reveal a strong negative relationship between fertility and female labour supply in all labour supply outcomes. The effectiveness of Lewbel's method, both independently and in combination with the same-sex instrument, outperforms the same-sex instrument alone. The study underscores the promise to enhance the precision of models reflecting real-world phenomena.

Key words: Fertility, Female Labour Supply, Causal effect, Lewbel's methods, Same-sex instrument

Supervisor:	J. Klooster
Second assessor:	A. Naghi
Date final version:	2nd July 2023

1 Introduction

This thesis delves into a question in econometrics: Can an instrument based on heteroscedastic error terms improve a linear model that estimates the effect of the number of children on labour supply, particularly when only a weak same-sex instrumental variable (IV) is available? This question is not only theoretically exciting but also has practical implications in the field of labour economics and policy-making.

In the past century, social and policy changes have led to increased participation in female labour (FLP) (Eckstein & Lifshitz 2011). These changes, such as expanded educational opportunities for women and expanded maternity leave policies, have significantly promoted FLP (Grant 2023; Low & Sánchez-Marcos 2015). However, women continue to bear the primary responsibility for childcare, which can potentially counteract the increase in FLP (Coltrane 2000). This tension between career and childcare has sparked interest in understanding the impact of childbearing on women's labour outcomes. Thus, it is helpful to analyse the potential effect of having a child on women's labour supply and FLP.

When analysed, this mutual relationship between the number of children and the female labour supply forms an obstacle. Due to this obstacle, endogeneity arises, creating space for inaccurate estimates. These deviations from accurate estimations become more prominent in a real-world setting, where the data differ from statistical normality and suffer from measurement error. This can lead to a false understanding of the supply of female labour. As a policy and, maybe, women's decisions to have children are based on knowledge of how a child affects their labour position, it is beneficial to have correct estimates about these supply outcomes. A suitable way to address endogeneity and measurement error is to implement a two-stage least squares (2SLS) approach using an IV.

This research partly replicates the findings described by J. Angrist & Evans (1996). An IV is used based on parents' preference for children of both sexes. According to J. Angrist & Evans (1996), this preference among parents for children of both sexes leads to an external shock that affects the number of children in a family, but does not directly affect the results of the labour supply. Although this preference among parents still manifests itself, it does have flaws (Miranda et al. 2018). First, it is the fact that it is not possible to examine changes differently from going from two to more children. Second, Öberg (2021) contests the use of the instrument as the use of the instrument would only give an external shock if the parents did not plan to have a third child before having the second child. Therefore, the instrument is considered weak, and other analysis options should be considered.

In real-world settings, the assumption of equal variance across the sample (i.e. homoscedasticity) may not hold, leading to heteroscedasticity. Recognising this issue, Lewbel (2012) proposed a novel method that turns this statistical "problem" into an advantage. This method relies on the premise that mismeasured variables across the sample could cause a common error, leading to heteroscedasticity. Using this common error, Lewbel's method creates a new instrument to address endogeneity and provide more accurate estimates. This innovative approach offers a compelling alternative when traditional IVs are insufficient. On the basis of this common error, an instrument that is supposed to be random and, more importantly, exogenous is created. Therefore, this instrument can counteract endogeneity in a similar 2SLS approach with an external IV. The Lewbel instrument can work independently of other instruments but can also be bundled with an external instrument, such as the same-sex instrument.

The present thesis uses the method proposed by Lewbel (2012) to investigate the impact of having children on women's employment factors. To assess the effectiveness of Lewbel's proposed method, we compare the results with those obtained using an external IV. This approach-based comparison forms the basis of our primary research question:

How effective is Lewbel's proposed method in establishing the effects of the number of children on women's labour supply using a comparison and a combination with an external IV approach?

In this thesis, we compare different methodological approaches to estimating instrumental variables with heteroscedasticity by thoroughly examining the coefficients of fertility and its impact on six distinct outcomes related to labour supply. This is done in two different settings: one where the external instrument plays a role, and one where this instrument is not fitted for that regression. The combination is only carried out for the first setting.

To address this question, we examine several aspects. Primarily, we explore Lewbel's proposed method for determining the impact of the number of children on women's labour supply. This method uses heteroscedasticity, a common statistical issue, to generate instrumental variables.

Furthermore, we dive deeper into the assumptions underpinning Lewbel's method. Understanding these assumptions is crucial as they form the method's foundation and significantly influence its applicability and effectiveness. We also compare these assumptions with those inherent in the external IV approach. This comparative analysis allows us to understand the unique strengths and potential limitations of each method.

By integrating these elements into our investigation, we aim to provide a comprehensive answer to our research question. This approach enhances our understanding of the impact of childbearing on women's labour supply and contributes to the broader discussion on econometric methods and their application in labour economics.

The thesis is structured as follows. After this introduction, we will present a comprehensive literature review, laying the groundwork for our research. We will then detail the data used in our study, discussing its sources, nature, and the collection and cleaning process. The methodology section will outline the theoretical underpinnings of the Lewbel method and the IV approach and how these methods are applied in our study. The results section will present the findings of our study, providing a detailed analysis and interpretation. We will conclude the thesis by summarising our key findings, discussing their implications, and suggesting avenues for future research.

2 Literature

2.1 Standard 2SLS Approach

In this part, we look at the issues that arise when analysing the effect that fertility has on the outcomes of labour supply and their solutions. As stated in the introduction, the central challenge in this thesis is addressing endogeneity. This issue arises due to the possible simultaneous determination of fertility and labour supply, a hypothesis established by Goldin & Katz (2002) and supported by research of J. Angrist & Evans (1996); Rosenzweig & Wolpin (1980); Lundborg et al. (2017). This mutual determination violates the Gauss-Markov theorem assumption of $E(X\epsilon) = 0$. Here, X are the explanatory variables used in the regression and ϵ is the resulting error term. According to the assumption, these should not be correlated; thus, the expectation should be zero (Theil 1971). In this theory, the assumption of homoscedasticity is also included. When these assumptions are compromised, this leads to biased and inefficient estimators.

A 2SLS approach using an IV can be a solution to endogeneity. However, to be a suitable IV, certain criteria must be met. In their research paper, J. D. Angrist et al. (1996) cover three assumptions that must hold for using an IV. For clarity, the instrumented variable is the predicted variable constructed by regressing the instrumental variables and additional control variables on the original endogenous variable. The first is the random assignment or independence assumption, which implies the ideal condition that the instrument is uncorrelated with the unobserved variables in the study; hence it is considered random. The second assumption is the exclusion restriction, which tells that an instrumental variable should only affect the dependent variable via the instrumented variable. The third assumption is the relevance assumption. The instrumental variable should have a nonzero average effect on the instrumented variable (i.e., coefficient in the first stage is nonzero). As specified by Imbens & Angrist (1994), the fourth assumption is the monotonicity assumption. This assumption states that the instrumented variable should either exclusively increase or decrease in the instrumental variable. These assumptions ensure that an instrumental variable is used properly and can counter endogeneity.

2.2 Same-sex

The same-sex instrument, which utilises the gender composition of children within a family, has been a key tool in studying family dynamics and its impact on various socioeconomic outcomes. This instrument is particularly relevant to our research, as it provides a natural experiment to address the endogeneity of family size and the sex composition of siblings in models (Öberg (2021)). The theoretical basis for the same-sex instrument lies in the assumption that parents may prefer a mixed gender composition of children, which can influence family size. This preference, coupled with the random nature of the sex of a newborn child, provides a natural experiment that researchers have exploited to study the causal effects of family size on various outcomes of labour supply. For example, Frenette (2011a; 2011b) used this instrument to examine the gender division of work after child birth and the impact of larger families on parental investments in child quality. The same-sex instrument allowed Frenette (2011a) to isolate the effect of family size from other confounders, providing more robust estimates of causal effects. Similarly, Nguyen (2019) used the same-sex instrument to investigate the effect of the sex composition of siblings on parental labour supply and occupational prestige in Indonesia. The study found that the composition of siblings by sex significantly influenced parental labour supply decisions with implications for occupational prestige. The same-sex instrument has also been used to study the effect of fertility on mothers' well-being (Cáceres-Delpiano & Simonsen 2012), and the consequences of unintended pregnancy on infant, child, and parental health (Gipson et al. 2008).

However, the same-sex instrument has been subject to considerable critique. Öberg (2021) argues that the same-sex instrument, like other instrumental variables for the number of children, is not reliable or interpretable. The author contends that there are many issues with these instrumental variables, including those based on twin births, as first used by Rosenzweig & Wolpin (1980), and that results based on them should be ignored. Furthermore, the author calls for the development of new, more credible methods. In addition to these critiques, the same-sex instrument is also subject to the critique of being weak. This is because the correlation between having children of the same sex and the total number of children a family has is relatively weak. This weak correlation can lead to biased estimates and reduced statistical power of the analysis. It is based on the assumption that the sex of the first two children only affects the family's decision to have more children and does not directly affect the outcome variables. However, this exclusion restriction assumption has been challenged. For example, it has been argued that the sex composition of children can directly affect parents' decisions on labour supply and children's educational results, violating the exclusion restriction assumption Rosenzweig & Wolpin (2000). In conclusion, while the same-sex instrument has been a popular tool in demographic and economic research, it is not without limitations and critiques. Another downside of using this instrument is that the standard endogenous stated variable, such as the number of children, must be adapted. J. Angrist & Evans (1996) choose to change their endogenous variable to *More than two children* instead of the actual number of children, since the same-sex relation with fertility is only compatible with such a binary variable. Future research should aim to develop and validate new instruments or methods to better understand the causal effects of family size, which could be the method based on heteroskedastic errors.

2.3 Internal Instrument Based on a Heteroscedasticity

In the field of econometrics, the issue of endogeneity often arises, leading to biased and inconsistent parameter estimates. This issue is particularly relevant in our research context, where unobserved factors, such as parental preferences or abilities, could simultaneously affect fertility and labour supply decisions. A common solution to this problem is to use instrumental variables (IV). However, when the number of endogenous regressors is greater than the number of instruments or when no external instruments are available, the Lewbel method provides a solution by using heteroscedasticity in the error terms to construct internal instruments (Lewbel 2012). Wright et al. (1928) applied "curve shifters", or IVs, to obtain estimates of the elasticity of supply and demand in the vegetable oil market. This was the first known use of second moments for identification. New methodologies have been created not to find an external instrument but to use general restrictions on higher moments. Dagenais & Dagenais (1997) show in their paper that their estimates perform better than an OLS. Cragg (1997) and Erickson & Whited (2002) show that simple errors in variables can provide a decent improvement for the estimates. A similar approach to the approach of Lewbel (2012) is taken by Hogan & Rigobon (2003). Their proposed method controls for endogenous education, unobserved ability, and measurement error using the natural heteroscedasticity of education attainment. Furthermore, through their research article, Klein & Vella (2010) discuss the challenges associated with identification in binary endogenous models. As part of their research, the authors estimate the heteroscedasticity semi-parametrically and use the residual from the second equation as an additional regressor in the first equation as the instrument. Researchers need the assumptions to hold for the distribution of the error terms to identify the model.

In examining the relationship between labour supply factors (e.g., individual income, family income, hours worked per week) and the number of children in a family, Lewbel's method proves particularly relevant. This method allows for a better understanding of these relationships, considering the heterogeneity in the data set. This heterogeneity can arise from various sources, such as differences in family structures, socioeconomic status, regional variations, and work preferences. Breusch & Pagan (1979) invented a test that can determine whether heteroscedasticity is present in a model. The Lewbel method, with its ability to construct internal instruments based on heteroscedasticity, can handle this heterogeneity effectively, providing more reliable and robust estimates Dufour (2003).

However, Lewbel's approach relies on several key assumptions. First, Lewbel's methodological approach assumes that the error term in the structural equation is heteroscedastic and that a function of the exogenous variables in the model can explain this heteroscedasticity. This is a crucial assumption as it forms the basis for constructing the internal instruments. J. Angrist & Evans (1996) use a binary dependent variable in the first-stage result. In Lewbel (2018), heteroscedasticity is tested with an approach in which the covariance between a variable and the residual in the first stage cannot equal zero. The main reason why a Breusch-Pagan (BP) test is not only used is that it might not give a correct display of heteroscedasticity. Second, it assumes that the instruments are relevant, i.e., correlated with the endogenous regressors, and valid, i.e., uncorrelated with the error term in the structural equation. These are standard assumptions in any 2SLS regression and are crucial for the consistency and unbiasedness of IV estimates (Newey & Windmeijer (2009)).

The suitability of the Lewbel method for a specific data set, such as the 1980 PUMS data used by J. Angrist & Evans (1996), would depend on heteroscedasticity and the relevance of the constructed instruments. The PUMS data, which provide a large source of information on various demographic and economic variables, are likely to exhibit heterogeneity due to the diverse nature of the population it covers. Hence, if present in error terms, this heterogeneity can be exploited by the Lewbel method to construct relevant and valid instruments. However, it is crucial to perform appropriate diagnostic tests before implementing this method to ensure that the assumptions of the Lewbel method are satisfied.

The Lewbel method also works without the use of an external instrument. Through his research paper, Lewbel indicates that the method works best in a setting with an instrument available, weak or strong. However, it is not necessary, and the analysis could still prove useful Lewbel (2012))

In the context of labour economics, the Lewbel method provides a powerful tool to examine the impact of the number of children on factors of labour supply. This method, which leverages heteroscedasticity to construct internal instruments, allows researchers to address endogeneity issues that often arise in such analyses. Consequently, it allows for a more nuanced understanding of the complex interaction between family size and labour supply factors.

However, the Lewbel method is not without assumptions. It assumes that the error term in the structural equation is heteroscedastic and that a function of the exogenous variables in the model can explain this heteroscedasticity. This assumption is crucial as it forms the basis for constructing the internal instruments. Furthermore, the method assumes that the instruments are relevant and valid (i.e., correlated with the endogenous regressors and uncorrelated with the error term in the structural equation). These are standard assumptions in any 2SLS regression and are crucial for the consistency and unbiasedness of the IV estimates.

3 Data

3.1 Descriptive Statistics

The data are collected from the 1980 Census Public Use Micro Samples (PUMS). These data are the same data that are used by J. Angrist & Evans (1996) for the sake of comparison. The census contains various information on households in the US for 1980. For this thesis, information on labour supply, the sex of the mother's first two children, an indicator of multiple births, and other demographic variables is used. The sex of the oldest children is used to define the pairs of same-sex siblings and to construct the IV. Therefore, only women between the ages of 21 and 35 with two children are selected from the total population of women in that age range.

As the census does not include retrospective fertility information other than the number of children, children are not matched to a household in the data set. J. Angrist & Evans (1996) attached individuals labelled as children in a household to a female householder or the spouse of a male householder. They removed any mother whose household size did not correspond to the total number of children ever born. Furthermore, relationship codes and subfamily identifiers matched children with their mothers in households with multiple families. Because the census does not track children across households, the sample is limited to mothers aged 21-35 whose oldest child was less than 18 years of age at the time of the census (1980). The main reason for imposing this limitation on the sample is that women under 21 are unlikely to have two children. Also, for children older than the age of 17 it is more likely to have moved to a different household.

In their paper on instrumental variables (IV) estimates J. Angrist & Evans (1996) indicate that these restrictions do not lead to a highly selected sample. The indication uses a comparison between this sample and the Current Population Survey (CPS). The CPS shows that women 35 years of age and with two or more children, at least 93% have an oldest child younger than 18. This sample, women aged 21-35 with at least two children, may appear as an unusually high-fertility group; however, for the group of women aged 21-27, at least one quarter qualifies, and for the group aged 28-35, over half of the entire population of women in that age group.

In Table 1 the descriptive statistics are given for the 1980 PUMS data. Statistics include information on variables about children's sex composition, dependent variables, and an IV. The primary variable of interest is the more than 2 children variable as an endogenous indicator of fertility. When the same-sex instrument is used, the average number of children is around 2.5. For the full data set where no criteria other than having one child have to be met, the average is around 2. The same-sex external instrumental variable is a composition of "two boys" and "two girls" variables. For the "All women" sample, the number of observations is 398,835 and for the "married" sample, the number is 254,654. The fourth column represents the sample that does not impose the restrictions that J. Angrist & Evans (1996) need for the 2SLS with the same-sex instrument. In addition, the fourth column of the tabular representation includes all data on women with at least one child, with 927,264 observations. For the present analysis, the number of children is the endogenous indicator of fertility.

In the bottom half of Table 1, demographic and labour supply variables are described, including measures of mother's age, age at first birth, years of education, and indicators of race and ethnic origin. For married women's spouses, demographic and labour supply data is also used and partially displayed in Table 1. The labour supply variables are derived from the 1997 census questions about employment. These variables measure whether the respondents worked for compensation, the number of weeks worked, their typical hours per week, and their annual labour income. When an individual did not work for remuneration, the last three are set to zero. The last three labour supply variables are actual variables for comparison in the research of J. Angrist & Evans.

In general, more boys were born as first or second child. It is derived from the labour supply variables that, in the 1980 data, men are the primary earners for families. In general, married women tend to work and earn less than the sample, including all women. This research looks mainly at the first four real labour supply factors. Therefore, it is important that around half of the women work in the different samples. They work around 20 weeks per year and between 15 and 20 hours per week, which generates an income of around $7,000,-$ yearly.

After following the exact steps from J. Angrist & Evans (1996), there was still a discongruity in the data set. J. Angrist & Evans removed five observations from the data set. However, the authors did not indicate the observations or the reason for removing five observations from the dataset. Nevertheless, the sample used still represents the researchers' data set; however, the removal of potential outliers does have an existing impact on the results. The fourth column, which includes all women with at least one child, does not deviate much from the sample that is restricted by same-sex specifications. The average number of children decreases slightly, since women with only one child are also included.

3.2 Data Suitability for Instruments Based on Heteroscedastic Errors

Lewbel's method uses, instead of an identification based on the standard minimal regression assumption (i.e. $E(\epsilon X) = 0$), a method that relies on restricting correlation $\epsilon\epsilon'$ with the set of independent variables (i.e. X). However, these restrictions do not automatically provide identification. These restrictions can lead to identification given some heteroscedasticity. Therefore, the key assumption that must hold in a setting, where one applies this proposed method, is that heteroscedasticity exists in the estimated model.

In order to establish heterogeneity in the models used in this thesis, Breusch-Pagan (BP) tests are performed. The Breusch-Pagan (BP) test was performed with H_0 of homoscedasticity, indicating the amount of heteroscedasticity in error terms. Together with the additional assumptions in Lewbel (2018) and some more standard regression assumptions, the test of the assumptions is made clear in Section 4 and is performed in Section 5.

Table 1

Descriptive Statistics, Women Aged 21-35 With More Than One Child and All Women With Children

	Data Specific for Same-sex Instrument			
Variable	All Women	Wives	Husbands	All Women
Children Ever Born	2.55 (0.81)	2.51 (0.77)	-	2.09 (0.99)
More Than 2 Children (1 if mother had more than 2 children, 0 if else)	0.402 (0.490)	0.381 (0.486)	-	-
Boy 1^{st} (s_1) (1 if the first child is a boy)	0.511 (0.500)	0.514 (0.500)	-	-
Boy 2^{nd} (s_2) (1 if the second child is a boy)	0.511 (0.500)	0.512 (0.500)	-	-
Two Boys (1 if the first two children were boys)	0.264 (0.441)	0.266 (0.442)	-	-
Two Girls (1 if the first two children were girls)	0.242 (0.428)	0.239 (0.427)	-	-
Same-sex (1 if first two children were the same-sex)	0.505 (0.500)	0.505 (0.500)	-	-
Age	30.1 (3.5)	30.4 (3.4)	33.0 (4.6)	31.5 (6.2)
Age at First Birth (Parent's age in years when the first child was born)	20.5 (2.9)	21.2 (2.9)	24.3 (4.0)	22.2 (4.4)
Worked for Pay (1 if worked for pay in the year prior to the census)	0.565 (0.496)	0.528 (0.499)	0.977 (0.150)	0.610 (0.488)
Weeks Worked (weeks worked in the year prior to the census)	20.8 (22.3)	19.0 (21.9)	48.0 (10.5)	23.2 (22.6)
Hours/Week (average hours worked per week)	18.8 (18.9)	16.7 (18.3)	43.5 (12.3)	20.6 (19.0)
Labour Income (labour earnings in the year prior to the census, in 1995 dollars)	7,160 (10,804)	6,250 (10,211)	38,919 (25,014)	8,689 (12,171)
Family Income (family income in the year prior to the census, in 1995 dollars)	42,342 (26,563)	47,646 (25,821)	-	44,710 (28,924)
Non-wife Income (family income minus wife's labour income, in 1995 dollars)	-	41,635 (24,734)	-	
Number of Observations	394,840	254,652	254,652	927,267

Note: The samples include women aged 21-35 with two or more children except for women whose second child is less than a year old in the first three columns. In the 1980 PUMS, the married women sample refers to women who were married at the time of their first birth, married at the time of the survey, and married once. The fourth column represents all women with at least one child.

3.3 Same-sex Instrument

J. Angrist & Evans (1996) make use of the same-sex instrument. This is a binary variable that combines data on families that first have two boys or two girls. It is built on the concept that women and men tend to want both girls and boys as their child composition. In Table 2, the fractions of families are given for a certain composition of children who had a third child. A clear difference is visible between families that have two children of the same sex or different sexes. Indeed, there is a visible tendency that couples' families are more likely to have a third child if they do not already have a boy and a girl. This effect increases slightly when this family consists of a married man and woman. The reason for this increase is beyond the scope of this research.

Table 2

Fractions of Families that Had a Third Child

	All Women		Married Women	
Sex of First Two Children in Families With Two or More Children	Fraction of Total	Fraction that had Another Child	Fraction of Total	Fraction that Had Another Child
One Girl One Boy	0.49461	0.371971	0.494652	0.346742
Two Boys	0.263719	0.422686	0.266073	0.403831
Two Girls	0.24167	0.441150	0.239276	0.424637
Both Same-sex	0.50539	0.431515	0.505348	0.413683
Difference Same-sex - Both Sexes	-	0.059544	-	0.066941

4 Methodology

In this thesis, the main goal is to establish whether an instrument based on heteroscedastic error terms can improve a linear model that estimates the effect of the number of children on the labour supply, where only weak same-sex IV is available. Labour supply measures consist of 6 different variables that are all independently estimated. The six labour supply variables are: whether an individual worked or not, the number of weeks worked, hours/ week worked, labour income, and the log of family income and nonwife income for the married sample.

In addition to the comparison, this thesis also considers a setting in which the instrument is not useful and, therefore, only Lewbel (2012) is used as IV. This additional consideration offers new opportunities, and also the direct relation between an extra child and the effect on labour supply can be estimated.

4.1 Lewbel's Method

In this section, Lewbel's method for a triangular or 2SLS design is set forth for general cases. In Lewbel (2012), the author considers a system where two observed variables Y_1 and Y_2, a set of exogenous variables called X, and two error terms are included:

$$Y_1 = X'\beta_1 + Y_2\gamma_1 + \epsilon_1 \tag{1}$$

$$Y_2 = X'\beta_2 + Y_1\gamma_2 + \epsilon_2 \tag{2}$$

Lewbel (2012) identifies the equation in both the simultaneous design, where the γ values are not equal to zero, and the triangular design, where γ_2 equals zero. When γ_2 is zero, this system forms a 2SLS design. Therefore, we are focused on that setting.

$$Y_1 = X'\beta_1 + Y_2\gamma_1 + \epsilon_1 \tag{3}$$

$$Y_2 = X'\beta_2 + \epsilon_2 \tag{4}$$

Lewbel first performs a regression in equation 4 and following on that the equation 5 is regressed again, however including the Lewbel instrument. Z is defined as a (sub)set of the exogenous variables in X:

$$Y_2 = X'\beta_2 + \gamma(Z - E(Z))\epsilon_2 + v_i \tag{5}$$

The estimated values for Y_2 are then used in model 3. The method is explained in more detail in Appendix A, where a Monte Carlo simulation is performed following this structure. A data-generating process creates the information, after which the estimates of the coefficients are derived.

The assumptions underlying the proposed method mainly focus on heteroscedasticity in the first-stage model. For this standard Breusch-Pagan (BP) tests can be performed. Lewbel (2012) needs $cov(Z, \epsilon_2^2) \neq 0$ to hold, which can be tested with the BP test. In the results section, these statistics are given and, for every model, the BP test statistic is listed. These tests work best when the dependent variable is continuous. The variable *Number of children* takes on the values of natural numbers, however, the variable *More than two children* is binary. Lewbel (2018) makes another assumption for this special case. For the first stage model, the following notation is used, where Y_2 is binary.

$$Y_2 = g(X) + \epsilon_2 \tag{6}$$

The additional assumptions that must hold is $cov(X, g(X)(1 - g(X)) \neq 0$. As Y_2 only takes values of 0 and 1, $g(X)(1 - g(X))$ equals $\epsilon_2(1 - \epsilon_2)$ for both values of Y_2. Another assumption is that $Y = (Y_1, Y_2)$' and X are random vectors. $E(XY'), E(XY_1Y'), E(XY2Y')$, and $E(XX')$ are finite and identified from data. $E(XX')$ is non-singular. These assumptions are related to the data used and can, quite straightforwardly, be tested by looking at the outcomes for the given multiplications and their rank. Another assumption states that $E(X\epsilon_1) = 0, E(X\epsilon_2) = 0$, and, for some random vector $Z, cov(Z, \epsilon_1\epsilon_2) = 0$. Z is a subset of X or the full set of X. These assumptions are also tested in the results and can be considered standard OLS assumptions as described in the Gauss-Markov Theorem.

Standard BP tests are performed to derive the heteroscedasticity for the first-stage models. For the second stage, testing for heteroscedasticity is done by means of the Pagan-Hall test. Pagan & Hall (1983) created a test that relaxes the homoscedasticity requirement in the other structural equations. The BP test does require this and, therefore, in this setting, where heteroscedasticity in the first stage is needed, does not fit. Pagan-Hall tests are, thus, performed for the second-stage models.

4.2 2SLS Models

To establish whether the external instrument can improve a linear model, where only a weak instrument is available in this setting, we make a comparison between various models. The four models that are considered for three different sample groups are (1) an OLS regression, (2) a 2SLS regression with only the use of the same-sex IV, (3) a 2SLS regression with only the use of Lewbel (2012) IV, and (4) a 2SLS regression with the use of both instruments.

A generalization of the four models using notation from J. Angrist & Evans (1996) is

$$Y = \alpha_1 W + \alpha_2 S + \beta X + \epsilon_1, \tag{7}$$

where Y = {Worked for pay, Weeks worked, Hours/week, Labour income, ln(Family income), ln(Non-wife income)}, W = {Age, Age at birth first child, Race}. S gives the sex of the first and second born child. X = endogenous variable (i.e., more than 2 children). The variables on race, age, and age when the first child was born are clearly exogenous demographic variables. In addition, the variables to indicate the sex of the child are incorporated in this model, and these are exogenous. They are included in the model to reduce the likelihood of omitted variable biases. The OLS regression uses the exact formula 7 to estimate its parameters.

To estimate the other three models, a 2SLS (two-stage least squares) is used with an IV or multiple IVs. This method is used when there is an endogenous variable, such as *More than 2 children*. In this setting, equation 7 would be the second stage of the 2SLS, the first stage equation looks like

$$X = \pi W + \gamma Z + \epsilon_2. \tag{8}$$

Again, the controls are included in this stage. In equation 8, an IV is included. In this thesis, three kinds of IV were used, producing three different models. The standard for a 2SLS, first, the first stage is regressed by means of OLS and with the estimates for the endogenous variable, the second stage is regressed also with OLS. The IVs used are the same-sex, Lewbel and a combination of both.

4.2.1 Same-sex

This binary instrument is constructed by a combination of families that have either first two sons or two daughters. The variable is 1 if the first two children are either both female or both male. If the data on the sex of the first two children are available, the instrument can be created. In section 2.1, the four IV assumptions are set forth. As the same-sex instrument consists of the sexes of children, which is random and has nothing to do with labour supply parameters, the instrument is random. The sex composition does affect the fraction of families that had a third child, as can be seen in Table 2. The fourth assumption cannot be verified as the instrument does not strictly increase the *More than 2 children* variable. This means that having two same-sex children is not only going to increase *More than 2 children*. A mixed-sex composition could also lead to an increase in fertility. This implies that the instrument does not fully meet the criteria to be a fit instrument. The first stage with this instrument looks as follows.

$$X = \pi W + \gamma(Samesex) + \epsilon_2. \tag{9}$$

4.2.2 Instrument Based on Heteroscedastic Errors

This instrument can be constructed as a simple function of the data set. The 2SLS with this IV can identify the parameters in the regression model by means of variables that are correlated with heteroskedastic error terms. To use this method, one assumes the following assumptions. The first two are the expectation of the product of the set W with the error terms in the first and second stages being zero. This assumption is also in the standard OLS assumptions. The third is assumption $cov(W, \epsilon_1\epsilon_2) = 0$. The key assumption $cov(W, \epsilon_2) \neq 0$ must hold in order for this method to work. Structural parameters are identified by a 2SLS regression of Y_1 on the controls $= \{W, S\}$ and the endogenous variable $= X$, using both these sets and a product of the demeaned set of controls and the error terms of the first stage as an instrument. Therefore, for this instrument, three regressions need to be performed. First equation 10 in order to perceive the error terms, and after this, these error terms are used for the instrument in 11.

$$X = \omega W + \epsilon_2 \tag{10}$$

$$X = \pi W + \gamma(W - E(W))\epsilon_2 + v, \tag{11}$$

If the covariance between the product and the error term increases, then so does the strength of the instrument (Lewbel (2012)). Thus, before using this method, one first has to establish heterogeneity in the model. Then the parameters in the model are identified.

4.2.3 Same-sex and Lewbel

The fourth model estimated in this thesis is a combination of both IVs. As Lewbel writes about combining the instrument based on heteroscedastic errors with an external instrument in a model in his working paper: "The resulting identification is based on higher moments and so is likely to provide less reliable estimates than identification based on standard exclusion restrictions, but may be useful in applications where traditional instruments are not available or could be used along with traditional instruments to increase efficiency." The first stages would have the same structure as Section 4.2.2, but the same-sex instrument is added.

$$X = \omega W + \epsilon_2 \tag{12}$$

$$X = \pi W + \gamma_1(W - E(W))\epsilon_2 + \gamma_2(Samesex) + v \tag{13}$$

Again, the same assumptions hold as in the two separate models. γ_1 is a vector of multiple numbers that correspond to the number of elements in W. This implies that every variable in W creates its own instrument.

4.3 Intensive margin

This thesis also looks at how specifically the number of children affects the outcomes of the labour supply. J. Angrist & Evans (1996) use the variable *More than 2 children* as a measure of fertility, although it is interpretable, it is better to have a simple measure, such as the number of children. The main reason for the researchers to use this variable *More than 2 children* is that their instrument requires this specification for the data and instrumented variable. It is better to have a simple measure such as the number of children, where J. Angrist & Evans (1996) their criteria do not have to be met. The Lewbel method gives this opportunity. The system of equations looks as follows:

$$Y = \alpha W + \beta X + \epsilon_1, \tag{14}$$

$$X = \pi W + \gamma(W - E(W))\epsilon_2 + v, \tag{15}$$

$$X = \omega W + \epsilon_2 \tag{16}$$

Now, X represents the number of children for all observations. The controls remain the same as in the previous settings. The additional assumption mentioned in Lewbel (2018) does not have to hold here as the endogenous variable is nonbinary. Therefore, assumption $cov(W, \epsilon_2^2) \neq 0$ must hold, which can also be tested with a BP test.

In the analysis, we examine the linearity of the relationship between the number of children and the outcomes of the labour supply. In an OLS model, this linearity is assumed, and by checking linearity, it can give insights into whether a linear model is appropriate and whether the number of children should not be estimated with a (semi)parametric model.

5 Results

In this section, the IV estimates of fertility on labour market outcomes of women, married or not, are captured by the models described in Section 4. First, all assumptions are tested for the models used. After that, the estimates of the IV 2SLS model per model are shown in Section 5.2. As the 2SLS model based on heteroscedastic errors uses the key assumption of heterogeneity in the first stage, a heterogeneity analysis is performed on different groups of husband income and education. In the last part of this section, we depart from the same-sex instrument with its criteria (as specified by J. Angrist & Evans (1996)) on the data set and look at the effect on all women with at least one child.

5.1 Assumptions

As Lewbel (2012) his method is restricted by some assumptions, these must be tested. According to the model used in the present thesis, the method for testing is given in the methodology. In this section, the actual testing takes place. Apart from the standard BP and PH tests that are included per model, the covariances between the controls and error terms in the first stage are tested separately. In Table 3, these covariances are given for all three different first stage results used in this thesis. The results for the age of the mother and the age of the firstborn child appear to differ only from zero. For covariances of *more than 2 children*, these values are not expected to be large, as $\epsilon_2(1-\epsilon_2)$ may never exceed a value of 0.25. Also, the values in parameter W (i.e., exogenous control variables) are mostly binary, making a higher covariance unlikely. Lewbel (2012) writes that if one of these covariances is close to a value of zero, then this results in a weak or useless instrument. Extensive trials were performed to explore whether leaving out race variables improves the model; however, improvements in the model as a result of leaving out race variables was not found to be the case. The strength of the Lewbel method comes mainly from the first two variables. Furthermore, the other three variables did not weaken the models. Therefore, in this thesis, all five variables are used to create instruments using Lewbel's method.

Table 3

Assumptions Heteroscedasticity

Endogenous Variable	more2k		KIDCOUNT
	All Women	Married Women	All Women
	$cov(W, \epsilon_2(1-\epsilon_2))$	$cov(W, \epsilon_2(1-\epsilon_2))$	$cov(W, \epsilon_2^2)$
Age Mother	-0.02	-0.03	1.08
Age First Time Mother	0.05	0.06	-0.54
Mother Hispanic	0.00	0.00	0.01
Mother Black	0.00	0.00	0.03
Mother Other Race	0.00	0.00	0.03
Number of Observations	394,840	254,652	927,267

Further tests were performed on the suitability of the data. Data set W is not correlated with the error terms for both stages for all models. Also, $E(XX')$ is non-singular. The rank is equal to the number of variables, making the expectation full rank.

5.2 2SLS Results

The results of the first stage for all different models are given in Table 4. The first four columns are used for the setting in which the same-sex instrument is used. The same-sex instrument only works for the margin from 2 to more children. The coefficients for same-sex in the first and third columns suggest a positive effect of this instrument on *More than two children*, used as a measure of fertility. The five other controls are used in the Lewbel method by creating instruments based on these separate controls. All the first-stage models appear to fail the homoscedasticity test. This is deduced by looking at the BP test statistics.

To establish the effect of *More than two children* on measures of labour supply in the 1980 PUMS data, the coefficients for this parameter are given in Tables 5 and 6. In these Tables, the coefficients for *More than two children* are given for four different methods: OLS, 2SLS using only the same-sex instrument, 2SLS using Lewbel's method based on heteroscedastic error terms and a combination of the two. Then different dependent variables or measures for the labour supply are taken to establish the effect of *More than two children.* The first measure is the *worked for pay* binary variable that indicates if an individual worked in the year before the census. The next variables display the *number of weeks worked* in the prior year for the census and the third indicator is the *average hours per week.* The remaining dependent variables display the individual's *labour income* and the last two variables show the *log of family -and non-wife income.*

As J. Angrist & Evans (1996) already established in their work, where only the same-sex instrument was included, the effect of *More than two children* is generally negative for labour supply outcomes. This is visible in columns (2) and (6) in Table 5 and 6 for the sample with all women and only married women. When comparing the models that use the same-sex instrument with the external instrument, a significantly decreasing shift is visible when using the external instrument. In most cases, the coefficient for *More than two children* becomes even more negative

Table 4

First Stage Results

Independent Variables	All Women Used in Same-sex Setting		All Married Women Used in Same-sex Setting		
	more2k	more2k	more2k	more2k	KIDCOUNT
Same-sex	0.06^{***} (0.00)	-	0.07^{***} (0.00)	-	-
Age Mother	-	0.03^{***} (0.00)	-	0.03^{***} (0.00)	0.9^{***} (0.00)
Age First Time Mother	-	-0.05^{***} (0.00)	-	-0.04^{***} (0.00)	-0.12^{***} (0.00)
Mother Hispanic	-	0.16^{***} (0.00)	-	0.16^{***} (0.01)	0.32^{***} (0.01)
Mother Black	-	0.00 (0.00)	-	0.00 (0.01)	-0.18^{***} (0.01)
Mother Other Race	-	0.9^{***} (0.00)	-	0.06^{***} (0.01)	0.17^{***} (0.01)
R^2	0.0037	0.0807	0.0048	0.0740	0.2213
F-stat	$1,461^{***}$	$6,928^{***}$	$1,216^{***}$	$4,069^{***}$	$52,700^{***}$
Breusch-Pagan Test	$1,360^{***}$	$10,482^{***}$	$1,149^{***}$	$8,327^{***}$	$44,177^{***}$
Number of Observations	394,840	394,840	254,652	254,652	927,267

Note: The first four columns are used for the results in tables 5 and 6. The fifth is used in Table 8.
$^{*}p < 0.1;^{**} p < 0.05;^{***} p < 0.01$

than the OLS estimates. The coefficients in (3) and (7) become more robust, compared to (2) and (6), even though the results are more negative. Specifically, for the *income variables and the weeks worked*, the results indicate an even more negative relationship between fertility and those dependent variables.

Lewbel indicates in his paper that the use of IV based on heteroscedastic errors, in the best scenario, can be done also including an external instrument. Therefore, in columns (4) and (8), a combination of the two instruments is used. For all the estimates in these columns, the coefficients are more robust than in either of the models with one instrument. In line with the expectation when combining models, the combination gives estimates that are between the ones from the two separate models. However, the estimates submitted through the model combination are skewed more towards the estimates for the coefficients of the model where only the external instrument is used. This could be a result of the more robust estimates that the model with the external instrument has for all the dependent variables.

Linking the above-stated to the research question of how effective Lewbel's method is in establishing the effect of the number of children on labour supply outcomes, it can be considered that Lewbel's method proves to be a more accurate method than the standard 2SLS with the same-sex instrument. It is true that the combination provides the most accurate estimates and, thus, again Lewbel's method provides a valuable addition to the general analysis of changes in female labour supply induced by having children. We can even state that the external instrument is an addition to Lewbel's method when we look at the standard errors. These prove to be lower than in the model where the external same-sex instrument is used. Although the covariances in Table 3 showed results that indicated moderate relations between the error terms and the set of data, the estimates still improve.

The Pagan-Hall statistics are added next to the model's coefficient estimates. The statistics indicate strong heteroscedasticity. This heteroscedasticity increases generally when also the same-sex instrument is included in the model specification. All the coefficients, including Lewbel's IV, are significantly relevant on a one-percent level. Also, more relevant than the models that only use the same-sex instrument.

Table 5

OLS and 2SLS Estimates of Labour-supply Models using Census Data

	All Women						
	(1)	(2)	F-stat (2)	(3)	PH (3)	(4)	PH(4)
Estimation Method	OLS	2SLS	-	2SLS	-	2SLS	-
Instrument for More than 2 children	-	Same-sex		Lewbel		Both	
Dependent Variable:							
Worked for Pay	-0.18^{***} (0.00)	-0.12^{***} (0.03)	1,382	-0.13^{***} (0.01)	8,846	-0.13^{***} (0.00)	8,817
Weeks Worked	-9.0^{***} (0.1)	-5.5^{***} (1.1)	2,288	-12.3^{***} (0.6)	5,418	-10.7^{***} (0.5)	6,056
Hours/Week	-6.7^{***} (0.1)	-4.5^{***} (1.0)	2,060	-7.1^{***} (0.5)	2,840	-6.5^{***} (0.5)	2,989
Labour Income	$-3,768^{***}$ (35)	$-1,900^{***}$ (546)	1,897	$-7,147^{***}$ (306)	1,267	$-5,920^{***}$ (266)	1,283
ln(Family Income)	-0.14^{***} (0.00)	-0.03 (0.07)	2,177	-0.34^{***} (0.04)	1,185	-0.27^{***} (0.03)	1,192

Note: The number of observations is 394,840. Non-wife income is not included, as this variable only plays a role in a couple. $^{*}p < 0.1;^{**} p < 0.05;^{***} p < 0.01$

For analysing the differences between the married and the total sample for women in Tables 5 and 6, we first look at the coefficients for the *log of non-wife income*. This generally shows that the link between fertility and the dependent variable is negative. The 2SLS coefficient for the combined models indicates that having a third child reduces the non-wife income by around 12.5%. This implies that the husband also earns less after having a child. This finding could tell that the husband does some of the work for raising children that would otherwise be done by the woman alone if a woman is not married. Therefore, the coefficients suggest that married women can work more and earn more after having a child. Also, generally, the family income for married couples decreases less than in the all-women sample.

5.3 Heterogeneity Analysis

This section provides a description of the heterogeneity analysis performed. As part of this thesis, we performed a heterogeneity analysis by grouping the sample into three different groups for both the husband's income and the woman's education. The income distribution for the husband is not explained in the paper by J. Angrist & Evans (1996), while the income distribution for the husband is indicated in their data. Therefore, this thesis uses the income group indicator in their data. In addition, J. Angrist & Evans do not specifically mention which outliers they remove from their sample, making it difficult to replicate. Still, the same-sex model and the combination of IVs show surprising differences, as seen in Table 5.3. Again, the coefficients are more robust in the model with both instruments compared to the one where only the same-sex

Table 6

OLS and 2SLS Estimates of Labour-supply Models Using Census Data for Married Couples

	Married Women						
	(5)	(6)	F-stat (6)	(7)	PH(7)	(8)	PH(8)
Estimation Method	OLS	2SLS	-	2SLS	-	2SLS	-
Instrument for More Than 2 Children	-	Same-sex		Lewbel		Both	
Dependent Variable:							
Worked for Pay	-0.12^{***} (0.03)	-0.10^{***} (0.02)	932	-0.15^{***} (0.02)	6,052	-0.14^{***} (0.01)	6,165
Weeks Worked	-8.0^{***} (0.1)	-5.3^{***} (1.2)	1,382	-10.7^{***} (0.7)	3,461	-9.5^{***} (0.6)	3,643
Hours/Week	-6.0^{***} (0.1)	-4.8^{***} (1.0)	1,394	-6.5^{***} (0.6)	1,974	-6.1^{***} (0.5)	2,025
Labour Income	$-3,148^{***}$ (42)	-1271^{**} (573)	1,041	$-5,087^{***}$ (312)	874	$-4,227^{***}$ (275)	884
ln(Family Income)	-0.14^{***} (0.05)	-0.05 (0.05)	949	-0.29^{***} (0.03)	246	-0.24^{***} (0.03)	248
ln(Non-wife Income)	-0.06^{***} (0.01)	0.03 (0.07)	854	-0.18^{***} (0.04)	357	-0.13^{***} (0.04)	359

Note: The number of observations is 254,652. $^{*}p < 0.1;^{**} p < 0.05;^{***} p < 0.01$.

instrument is used. From section 5.2, it was determined that the combined model outperforms the model where only Lewbel's instrument is used. Together with the fact that Lewbel (2012) indicates that the generated instrument is best used together with another external IV, the heterogeneity analysis only uses the same-sex and the combined model.

As the external instrument needs heteroscedasticity in the first stage, it is especially interesting to look at the differences between cohorts. For the first three rows in Table 7, the three labour supply variables are considered: *Worked for Pay, Weeks / Year, and Individual Income.* The cohorts are grouped according to the husband's income distribution. The model with only the same-sex instrument tends to underestimate the effect that an increase in *more than 2 children* has on the labour supply outcomes for the bottom and upper groups. This aligns with the findings of Section 5.2. However, now it is visible that this is mainly the case for the more divergent cohorts. Although the estimates for the lower husband's income group are lower for the combined, the estimates for the higher income group are even worse. The following three rows are grouped by educational level. This is believed to be a proxy for income, so the sample is grouped in this manner. As Gronau (1973) states in his research: The impact of having a child on labour supply factors might be greater for more educated women. Based on the results displayed for the same-sex instrument J. Angrist & Evans reasoned that this was not the case. However, the results for the combined model show strongly that for more educated women, the effect of having a third child is severe. This would then confirm the findings of Gronau.

Again, linking this with the main research question, including the Lewbel instrument, provides more precise estimates in all models. It gives new insight into how having children affects labour supply outcomes for different groups. The effectiveness is derived from the more accurate estimates, significantly changing results for the different cohorts and improving the estimates.

Table 7

2SLS Estimates of Labour-supply Models for Different Income and Education Cohorts for Married Couples

		Worked for Pay			Weeks per Year			Income Individual		
Cohorts	% of Total	Mean of Dep. Var.	Same-sex	Both	Mean of Dep. Dar.	Same-sex	Both	Mean of Dep. Var.	Same-sex	Both
Husband's Income										
Bottom Third	21	0.55	-0.11^{**} (0.07)	-0.15^{***} (0.05)	20.1	-6.7^{*} (3.4)	-11.1^{***} (2.2)	6,210	$-1,397$ (1,553)	$-4,455^{***}$ (989)
Middle Third	37	0.59	-0.20^{***} (0.05)	-0.10^{***} (0.02)	21.9	-8.5^{***} (2.1)	-8.1^{***} (1.0)	7,191	$-1,461^{**}$ (716)	$-4,359^{***}$ (345)
Upper Third	42	0.46	-0.07^{*} (0.04)	-0.17^{***} (0.02)	15.9	-3.1^{*} (1.6)	-9.7^{***} (0.8)	7,615	-388 (801)	$-4,149^{***}$ (384)
Education										
Not Graduated	18	0.47	-0.12^{*} (0.06)	-0.10 (0.06)	16.1	-7.2^{**} (2.7)	-3.9^{**} (2.2)	4,747	$-3,249^{***}$ (1,081)	$-1,557$ (926)
High School Graduated	49	0.52	-0.15^{***} (0.04)	-0.12^{***} (0.02)	19.2	-6.2^{***} (1.7)	-8.95^{***} (0.8)	5,865	$-1,461^{**}$ (716)	$-4,359^{***}$ (345)
More Than Graduated	33	0.57	-0.07 (0.05)	-0.19^{***} (0.02)	20.4	-2.6 (2.3)	-10.4^{***} (0.8)	7,615	141 (1,278)	$-5,402^{***}$ (443)

Note: This table estimates the variable *More than 2 children* for different cohorts for education and husband's income on three labour supply outcomes. The (control) variables are the same as noted in the methodology section. For the models in the first three rows, two variables (i.e., *hispm* and *otherracem*) were not found to be heteroskedastic at a 5% level, thus indicating a weak instrument for those variables.
$^{*}p < 0.1;^{**}p < 0.05;^{***}p < 0.01$.

5.4 Intensive Margin

In this part of the results, we look at how the number of children affects the labour supply proxies. Previously, a variable introduced by J. Angrist & Evans (1996) *More than 2 children* was used as a fertility measure. The main reason for this usage was the applicability of the same-sex instrument and the suitability to combine with the Lewbel instrument. The same-sex instrument works only at the margin for more than two children. However, the instrument, as generated through the seminal paper by Lewbel, also works without an external instrument. As other external instruments are unavailable, Lewbel could provide exciting insights into labour supply outcomes for all changes in the number of children. In Table 8 the outcomes are given for an OLS and 2SLS regression. Also, the R^2 and the Breusch-Pagan statistics are expressed in the Table 8. The coefficients for the child count still indicate a strong negative relation for all the labour supply outcomes. The 2SLS uses the Lewbel instrument.

Table 8

OLS and 2SLS Estimates of Labour-supply Models With the Lewbel Instrument

Dependent Variable	OLS	2SLS	OLS R^2	2SLS R^2	PH Test
Worked for Pay	-0.11^{***} (0.00)	-0.04^{***} (0.00)	0.05	0.04	15,000
Weeks Worked	-6.1^{***} (0.0)	-2.9^{***} (0.1)	0.08	0.06	9,587
Hours/Week	-4.8^{***} (0.0)	-1.9^{***} (0.0)	0.07	0.05	2,714
Labour Income	$-2,840^{***}$ (14)	$-2,012^{***}$ (30)	0.07	0.07	5,583
ln(Family Income)	-0.04^{***} (0.00)	-0.09^{***} (0.00)	0.05	0.05	2,679

Note: $^{*}p < 0.1;^{**}p < 0.05;^{***}p < 0.01$.

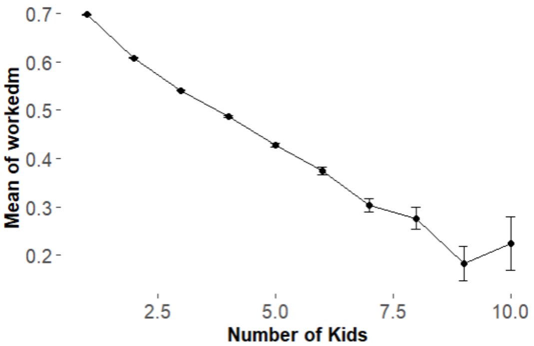
Considering the results presented in Tables 5 and 6, the OLS and the 2SLS only using Lewbel give the most negative estimates. Therefore, it is probable that the estimates in Table 8 are negatively overestimated. Yet, the estimates are well below the OLS results, which did not occur in section 5.2. The estimated effect of having a child on the first five labour supply outcomes indicates a negative relation between having an additional child and labour supply factors. Comparing the results with Tables 5 and 6, we see that the estimates in the current are approximately a fraction of 2 to 4 lower than the estimates in Tables 5 and 6. The estimates in Tables 5 and 6 seemed to be the most robust and efficient for the model using two instruments. Unfortunately, some of this strength is lost in the current models as the external instrument is unavailable. These results do, very straightforwardly, state what an additional child costs the mother and family. A woman is approximately 4% less likely to work when having an additional child, keeping all the other control variables the same. Also, her income drops with \$2012,- on average per child.

Although an external instrument would be a helpful addition, this method does already allow looking at the linearity of the labour supply outcomes on the number of children. This helps to determine whether the variable for child count can be used most appropriately in a linear model. This linearity is displayed in Figure 1 for four outcomes of the female labour supply: whether one worked, average hours/week, the average number of weeks/year and the individual income.

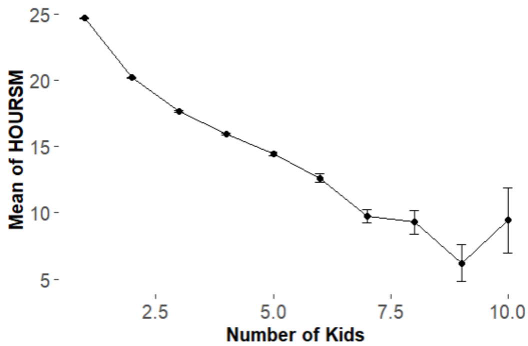
In Figure 1, the mean of four labour supply outcomes is plotted on the number of kids together with the standard errors of the means. When looking at the graphs, no visible differences appear in marginal increases in children. However, the step from 1 to 2 does indicate some more negative results compared to the rest of the values and it does not implicate nonlinearity.

Figure 1

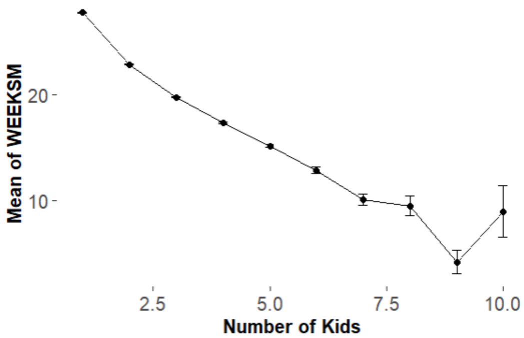
A Plot of Labour Supply Outcomes' Means on Numbers of Children



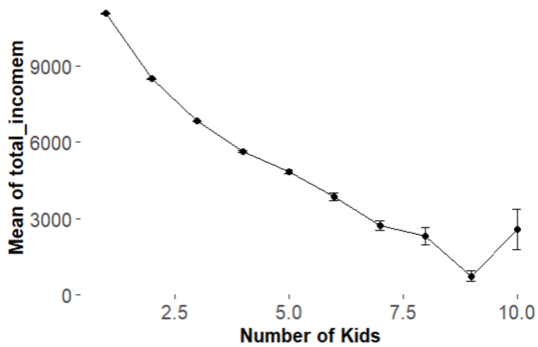
(a) Workedm



(b) HOURSM



(c) WEEKSM



(d) total_incomem

Note. This Figure gives a representation of the means for the dependent variables: Whether a woman worked in the year before the census, how many hours on average per week, how many weeks in the last year and her yearly income and the standard errors of the means for all different values of the variable *KIDCOUNT*. All observations for more than 10 children were left out.

6 Conclusion

This thesis aimed to answer two main questions: What is the causal effect of fertility on female labour supply, and how effective is Lewbel's method in determining this relationship compared to the traditional same-sex instrument? By taking advantage of various econometric methodologies such as 2SLS for the same-sex and Lewbel IV, we provided different insights into the relationship between fertility and labour supply outcomes. Our findings consistently suggest a serious negative relationship between fertility and female labour supply, across all variables tested. Having more children reduces female labour participation, the number of weeks worked, hours per week, and labour income. This supports existing research suggesting that increased fertility reduces women's participation in the labour market and strengthens and supports even more clearly J. Angrist & Evans (1996) their conclusion that the relationship is negative. These results have substantial implications, potentially influencing women's decision-making regarding family and career. Our study went a step further by comparing the effectiveness of different models. We found that the Lewbel instrument provided better estimates independently and when accompanied by the same-sex instrument. Linking this conclusion to the main research question provides us with the answer that Lewbel's instrument is very effective and more accurate in establishing the relationship between fertility and female labour supply than when using the lacking same-sex instrument. The combination is an improvement over both models alone since Lewbel (2012) supports. Even though not all constructed instruments proved to be helpful, Lewbel's method is a constructive addition to this research field, where an appropriate external instrument is not accessible yet.

In response to our research questions, we found a significant negative relationship between fertility and female labour supply, with each additional child reducing a woman's likelihood of working by 4% and her income by an average of $2012. Furthermore, we discovered that Lewbel's method, due to its ability to provide robust estimates without the need for an external instrument, was more accurate in establishing this relationship compared to the traditional same-sex instrument. The Lewbel instrument was also more effective when used in combination with the same-sex instrument, enhancing its accuracy.

In summary, this study builds on existing literature by giving new insights into the effect of fertility on female labour supply through various econometric methods, including the Lewbel instrument based on heteroscedastic error terms. It highlights the potential of innovative approaches and advocates for continuous exploration and exploitation of new methods to refine models for real-life phenomena. The revelations of this study could inspire future research and policy formation, encouraging society to mitigate the adverse effects of fertility on women's labour supply.

7 Discussion

This thesis used Lewbel's method for creating an instrumental variable. The main issue with applying this method was that the all the variables used as instrument did not show very heteroscedastic errors. This can lead to useless instruments, however in this thesis, the estimates still greatly improved. Also, the models still rejected homoscedasticity in the second-stage, which potentially indicates that the estimates are not without error. The reason to apply the Lewbel method is, because strong external IV's are scarce in this research field. If there was no question of endogeneity or mismeasurement then the use of Lewbel would probably not give such differences. Therefore, a more appropriate IV is necessary. Lundborg et al. (2017) found a strong instrument based on IVF, however the use of this instrument reduces the number of observations to all the women that attempt IVF. As long as the perfect IV has not been found, the Lewbel method proves to be a valuable addition to models into female labour supply outcomes. To improve the models used in this thesis, an increase in the number of exogenous controls can be considered. Also, a different method than 2SLS can be considered. For instance, as many binary variables are used in the model such as the endogenous *More than two children* and also one dependent *Worked for pay*, a probit regression could be a more fitting method for models including those variables. As Lewbel (2018) mentions his method works also for a binary variable, however a probit model is a more econometric, enhanced method for dependent binary variables. For further model improvements, it should be determined whether a linear relation between the number of children is actually correct. These are only three suggestions for model specification.

As the results in this thesis mainly compare the results of J. Angrist & Evans (1996), their data set is used. The 1980 PUMS data is likely not representational for the current situation. A newer data set would thus give more precise information, however this privacy sensitive information including an IV is not freely available. In summary, the model could be further optimised for different settings combined with a more contemporary data set.

References

Angrist, J. & Evans, W. N. (1996). *Children and their parents' labor supply: Evidence from exogenous variation in family size.* National bureau of economic research Cambridge, Mass., USA.

Angrist, J. D., Imbens, G. W. & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, *91*(434), 444–455.

Breusch, T. S. & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, 1287–1294.

Coltrane, S. (2000). Research on household labor: Modeling and measuring the social embeddedness of routine family work. *Journal of Marriage and family*, *62*(4), 1208–1233.

Cragg, J. G. (1997). Using higher moments to estimate the simple errors-in-variables model. *Rand Journal of Economics*, S71–S91.

Cáceres-Delpiano, J. & Simonsen, M. (2012). The toll of fertility on mothers' wellbeing. *Journal of Health Economics*, *31*, 752-766.

Dagenais, M. G. & Dagenais, D. L. (1997). Higher moment estimators for linear regression models with errors in the variables. *Journal of Econometrics*, *76*(1-2), 193–221.

Dufour, J.-M. (2003). Identification, weak instruments, and statistical inference in econometrics. *Canadian Journal of Economics/Revue canadienne d'économique*, *36*(4), 767–808.

Eckstein, Z. & Lifshitz, O. (2011). Dynamic female labor supply. *Econometrica*, *79*(6), 1675–1726.

Erickson, T. & Whited, T. M. (2002). Two-step gmm estimation of the errors-in-variables model using high-order moments. *Econometric Theory*, *18*(3), 776–799.

Frenette, M. (2011a). How does the stork delegate work? childbearing and the gender division of paid and unpaid labour. *Journal of Population Economics*, *24*, 895-910.

Frenette, M. (2011b). Why do larger families reduce parental investments in child quality, but not child quality per se? *Review of Economics of the Household*, *9*, 523-537.

Gipson, J. D., Koenig, M. A. & Hindin, M. J. (2008). The effects of unintended pregnancy on infant, child, and parental health: A review of the literature. *Studies in Family Planning*, *39*, 18-38.

Goldin, C. & Katz, L. F. (2002). The power of the pill: Oral contraceptives and women's career and marriage decisions. *Journal of political Economy*, *110*(4), 730–770.

Grant, B. M. (2023). Dismantling the father's house?: Women as doctoral supervisors. *Learning and Teaching*, *16*(1), 1–30.

Gronau, R. (1973). The effect of children on the housewife's value of time. *Journal of Political Economy*, *81*(2, Part 2), S168–S199.

Hogan, V. & Rigobon, R. (2003). Using unobserved supply shocks to estimate the returns to education. *Unpublished manuscript*, 7.

Imbens, G. W. & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica: journal of the Econometric Society*, 467–475.

Klein, R. & Vella, F. (2010). Estimating a class of triangular simultaneous equations models without exclusion restrictions. *Journal of Econometrics*, *154*(2), 154–164.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, *30*(1), 67–80.

Lewbel, A. (2018). Identification and estimation using heteroscedasticity without instruments: The binary endogenous regressor case. *Economics Letters*, *165*, 10–12.

Low, H. & Sánchez-Marcos, V. (2015). Female labour market outcomes and the impact of maternity leave policies. *IZA Journal of Labor Economics*, *4*, 1–22.

Lundborg, P., Plug, E. & Rasmussen, A. W. (2017). Can women have children and a career? iv evidence from ivf treatments. *American Economic Review*, *107*(6), 1611–1637.

Miranda, V., Dahlberg, J. & Andersson, G. (2018). Parents' preferences for sex of children in sweden: Attitudes and outcomes. *Population research and policy review*, *37*, 443–459.

Newey, W. K. & Windmeijer, F. (2009). Generalized method of moments with many weak moment conditions. *Econometrica*, *77*(3), 687–719.

Nguyen, G. (2019). Sibling-sex composition, childbearing and female labour market outcomes in indonesia. *Journal of Population Research*, *36*, 13-34.

Pagan, A. R. & Hall, A. D. (1983). Diagnostic tests as residual analysis. *Econometric Reviews*, *2*(2), 159–218.

Rosenzweig, M. R. & Wolpin, K. I. (1980). Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica: journal of the Econometric Society*, 227–240.

Rosenzweig, M. R. & Wolpin, K. I. (2000). Natural "natural experiments" in economics. *Journal of Economic Literature*, *38*(4), 827-874.

Theil, H. (1971). Principles of econometrics.

Wright, P. G. et al. (1928). Tariff on animal and vegetable oils.

Öberg, S. (2021). The casual effect of fertility: The multiple problems with instrumental variables for the number of children in families.

A Replication simulation

Lewbel (2012) included a simulation study in his paper. This thesis used the 2SLS, which coincides with the triangular system in his simulation. This simulation aims to elaborate on the method and prove that method works. The R code for the simulation is included in the code file. The simulation is based on equations A3 and A4. Monte Carlo simulations draw data from the reduced form of the structural model

$$Y_1 = \beta_{11} + X\beta_{12} + Y_2\gamma_1 + \varepsilon_1, \qquad \varepsilon_1 = U + e^X S_1 \tag{A1}$$

$$Y_2 = \beta_{21} + X\beta_{22} + Y_1\gamma_2 + \varepsilon_2, \qquad \varepsilon_2 = U + e^{-X} S_2 \tag{A2}$$

where X, U, S_1, and S_2 are independent standard normal scalars and $\beta_{11} = \beta_{12} = \beta_{21} = \beta_{22} = \gamma_1 = 1$. The triangular design sets $\gamma_2 = 0$ and $Z = X$. With these choice of Z the model parameters in each design are exactly identified by Theorems 1 and 2 in in Lewbel (2012). The parameters in equation A3 for the triangular design are not identified using traditional exclusion assumptions. Table 1 reports results of 10,000 simulations of each design, with sample size $n = 500$. Thus, the triangular design the DGP looks as follows:

$$Y_1 = 1 + X + Y_2\gamma_1 + \varepsilon_1, \qquad \varepsilon_1 = U + e^X S_1 \tag{A3}$$

$$Y_2 = 1 + X + \varepsilon_2, \qquad \varepsilon_2 = U + e^{-X} S_2 \tag{A4}$$

After generating the data, we first estimate $\widehat{\beta}_2 = (\widehat{\beta}_{21}, \widehat{\beta}_{22})$

$$\widehat{\beta}_2 = \overline{XX'} - \overline{XY_2}, \quad \widehat{\varepsilon}_2 = Y_2 - X'\widehat{\beta}_2$$

The triangular design is estimated using the two stage least squares estimator and $\widehat{\varepsilon}_2$

$$\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\gamma}_1 \end{pmatrix} = \left(\widehat{\Psi}'_{ZX}\widehat{\Psi}^{-1}_{ZZ}\widehat{\Psi}_{ZX}\right)^{-1}\widehat{\Psi}'_{ZX}\widehat{\Psi}^{-1}_{ZZ}\left(\frac{\overline{XY_1}}{\overline{(Z-\bar{Z})\widehat{\varepsilon}_2 Y_1}}\right)$$

where $\widehat{\Psi}_{ZX}$ function is defined as

$$\Psi_{ZX} = E\left[\begin{pmatrix} X \\ [Z - E(Z)]\varepsilon_2 \end{pmatrix}\begin{pmatrix} X \\ Y_2 \end{pmatrix}'\right]$$

and where $\widehat{\Psi}_{ZX}$ function is defined as

$$\Psi_{ZZ} = E\left[\begin{pmatrix} X \\ [Z - E(Z)]\varepsilon_2 \end{pmatrix}\begin{pmatrix} X \\ [Z - E(Z)]\varepsilon_2 \end{pmatrix}'\right]$$

The coefficients of interest are β_1, β_2 and γ_1. For this simulation specifically, the β is a 2x1 vector and X is a 500x1 matrix. Y_1, Y_2, ϵ_1 and ϵ_2 are of the same dimension. The Ψ functions are 2x2. This results in the following in the simulation results. As expected, the results are very near the true values.

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
β_{11}	1	1.002	0.139	0.909	1.001	1.093	0.139	0.11	0.092
β_{12}	1	1.002	0.275	0.83	0.999	1.171	0.275	0.212	0.17
β_{21}	1	1.001	0.137	0.911	1.002	1.094	0.137	0.109	0.092
β_{22}	1	1	0.277	0.831	1.003	1.174	0.277	0.213	0.172
γ_1	1	0.999	0.035	0.98	1	1.019	0.035	0.026	0.019