Analyzing Relationship Between Female Labor Supply and Number of Children Using IV Approach

陈三强

西安交通大学 金禾经济研究中心 710049

i. Introduction

The link between women's labor supply and children has long been an interesting topic in economics. Various models have been developed to disentangle the causal mechanisms linking fertility and labor supply. For example, the household production model we learned from Labor Economics which suggests that parents derive utility from number of children and quality of children, which are complementary goods of each other.

Various empirical studies show negative correlation between fertility and female labor supply. But a difficulty in estimating this relationship is the endogeneity problem. Some other unobserved factors may affect both labor supply and fertility, such as delay of marriage, decline of childbearing, and divorce. Strong theoretical reasons also believe that labor supply and fertility are jointly determined. Hence by various reasons, OLS is flawed in such estimation. This paper uses IV approach to estimate this relationship using sibling sex mix of first two children as instrument with data from 1990 and 1980 PUMS from America. Unlike Chinese traditions of which people prefer boys than girls, American people prefers a composition of mix genders of children, under which context parents having first two children as same sex are extremely likely to have an additional child. On the other hand, gender of children is randomly designed and thus is independent from any other factors, so we can use a dummy variable of mix of sex of first two children as an instrument of childbearing. Since this instrument is only available among women with two or more children, we will limit the sample in next section.

ii. Data, Descriptive Statistics

A. Data and Descriptive Statistics

The data we use is from American Census Public Use Micro Samples (PUMS).

Table 1 reports the labor force participation rate and information of childbearing among women aged 21-35 and aged 36-50 using data from 1980 and 1990, 2015 and 2014. Our main focus is the data from 1980 and 1990 since in nowadays the phenomenon of sub-replacement fertility has made it difficult to identify the impact of fertility on female labor supply, while focusing data from 1980 and 1990 when post-war increase in female labor supply has been observed may make it easier to identify the impact running from fertility to female labor supply. Table 1 reports labor force participation rate and reported fertility of women. We cut the sample into two groups, the first group is women aged between 21 and 35 with oldest kids less than 18 and the second group is women aged between 36 and 50. Because kids older than 18 may form a new family so we limit the age of kids and by also limiting age of mother to be less than 35 we try to assure less sample selection problem. Data from 1980 and 1990 shows decline in number of children and increase in labor force participation rate, which may intuitively correspond to the result of previous works that labor supply and fertility are negatively correlated.

We then perform a T-test on the fertility and labor force participation rate between 1980 and 1990. Table 3 summarizes the levels and changes in fertility and labor force participation rates by year in women aged between 21 and 35 and in women aged between 36 and 50. Both groups shows significant positive change in labor force participation rate and negative change in fertility, which are significant at 1% level.

We then limit the sample to women aged between 20 and 60 with two or more children and drop those aged between 21 and 35 with oldest kids older than 18. And by doing Stata operation with command “collapse” and “reshape” we attach information of children to their parents according to family serial numbers. Then we drop all observations whose reported number of children did not match the actual

number we count.

Table Ⅰ— Fertility and Labor Force Participation Measures

Sample	1980 PUMS	1990 PUMS	2014 PUMS	2015 PUMS
Women aged 21-35				
mean children ever born	1.11	1.03	0.78	0.77
percent with 2 or more children	0.36	0.33	0.24	0.24
percent worked last year	73.58%	79.53%	79.25%	79.52%
Observations	981,251	985,614	180,019	361,871
Women aged 36-50				
mean children ever born	1.76	1.38	1.39	1.40
percent with 2 or more children	0.54	0.44	0.46	0.46
percent worked last year	66.73%	78.84%	77.38%	77.56%
Observations	634,287	834,044	198,327	395,084
Women aged 21-35 with 2 or more children				
mean children ever born	2.53	2.50	2.56	2.56
percent with more than 2 children	0.37	0.36	0.39	0.39
percent worked last year	55.20%	65.56%	66.79%	67.08%
Observations	352,015	325,840	43,615	86,482
Married women aged 21-35 with 2 or more children				
mean children ever born	2.52	2.49	2.54	2.54
percent with more than 2 children	0.37	0.35	0.38	0.38
percent worked last year	53.72%	66.37%	63.63%	63.72%
Observations	290,066	259,834	31,000	61,614

Table 2 reports the T-test results.

Table 2—Difference in Fertility and Labor Supply

Variable	Group by year		
	1980	1990	Difference, 1980-1990
Number of children ever born,women aged 21-35	1.11	1.03	0.08
	(0.0012)	(0.0012)	(0.0017)
Percentage worked last year,women aged 21-35	0.74	0.80	-0.06
	(0.0004)	(0.0004)	(0.0006)
Number of children ever born,women aged 36-50	1.76	1.38	1.55
	(0.0019)	(.0014)	(0.0011)
Percentage worked last year,women aged 36-50	0.67	0.79	-0.12
	(0.0006)	(0.0004)	(0.0007)

Table 3 summarizes descriptive statistics of independent variables, dependent variable and instruments. The summarization is conducted based on two subsamples,

the first is women aged between 21 and 35 and have at least 2 children, the second subsample is women aged between 21 and 35 having at least 2 children and is married at the time of census.

Table 3—Descriptive Statistics, Women Aged 21-35 with 2 or More Children

Variable	Mean and Standard Deviations			
	1980 PUMS		1990PUMS	
	All women	Wives	All women	Wives
Children ever born	2.54	2.54	2.45	2.45
	(0.83)	(0.82)	(0.76)	(0.75)
More than 2 children	0.383	0.383	0.337	0.333
	(0.486)	(0.486)	(0.473)	(0.471)
First child was boy	0.512	0.514	0.514	0.516
	(0.500)	(0.500)	(0.500)	(0.500)
Second child was boy	0.511	0.513	0.511	0.512
	(0.500)	(0.500)	(0.500)	(0.500)
Two boys	0.264	0.265	0.264	0.265
	(0.441)	(0.442)	(0.441)	(0.441)
Two girls	0.241	0.239	0.240	0.238
	(0.427)	(0.427)	(0.427)	(0.426)
Same sex	0.505	0.505	0.504	0.503
	(0.500)	(0.500)	(0.500)	(0.500)
Age	29.65	29.75	30.19	30.39
	(3.72)	(3.67)	(3.58)	(3.47)
Age at first birth	22.21	22.45	23.62	24.06
	(3.93)	(3.89)	(4.47)	(4.38)
Worked last year	0.573	0.558	0.695	0.695
	(0.495)	(0.497)	(0.461)	(0.460)
Weeks worked last year	21.34	20.45	28.09	28.06
	(22.35)	(22.13)	(22.69)	(22.62)
Labor income	6,998	6,558	9,478	9,545
	(10,691)	(10,380)	(12,983)	(13,134)
Family income	41,865	45,878	42,539	48,249
	(26,423)	(25,599)	(34,514)	(34,626)
Non-wife income	35,092	39,552	33,477	39,163
	(25,488)	(24,257)	(32,092)	(32,110)
Observations	474,622	406,310	425,59	346,185

The instrumental variable for *More than 2 children* is *Same sex*. The labor supply variables are *Worked last year*, *Weeks worked last year, Labor income, Family income and Non-wife income*. *Labor income* is income earned by women

and $Non-wife\ income$ is computed from $Family\ income$ minus $Labor\ income$. Incomes reported to be 0 or negative are set to 1 when computing logarithm and are all converted to 1995 dollar. The deflator coefficient for 1980 dollars is 2.0992, the deflator coefficient for 1990 dollars is 1.2290.

The descriptive statistics also show that women's labor force participation rates, weeks worked, and income have risen substantially from 1980 to 1990.

B. Sex Mix and Fertility – Availability of Instrument

The preference of mixed children gender composition from parents has been studied by plenty of works. We now briefly summarize the theory of impact of sex mix on fertility from previous works by Becker and Gregg H. Lewis (1973)[错误!未找到引用源。] and Rosenzweig and Wolpin (1980). According to these models, parents derive utility from "child quality" as a complementary good of "child quantity". And sex mix is something described by Ben-Porath and Welch (1980)[3] that determines child quality. Suppose a mother who already has at least one child, if she prefers a mixed gender composition then a same sex composition would reduce the utility derived from current children which then increases the marginal utility of the additional child and thus increases the possibility of having an additional child.

Table 4 reports the fraction of women that had a third child (or more children) conditional on the sex of first two children conducted on sample of all women and sample of married women. The last row of this table performs a T-test to show the difference of fraction of women that had a third child between same-sex and mixed-sex groups.

Both subsamples show significant difference of fraction of women that had a third child between same-sex group and mixed-sex group. The result that women with first two kids with same sex are significantly likely to have a third child is correspondent with the previous theory and thus the correlation between $same\ sex$ and $fertiliy$ is proved to be significant.

Table 4—Fraction of Women That Had Third Child by Sex Composition of First Two Children

	All women				Married women			
	1980 PUMS		1990 PUMS		1980 PUMS		1990 PUMS	
Sex of first two children	Fraction of sample	Fraction that had third child	Fraction of sample	Fraction that had another child	Fraction of sample	Fraction that had third child	Fraction of sample	Fraction that had third child
Two boys	0.264	0.403	0.264	0.362	0.265	0.403	0.265	0.360
		(0.002)		(0.001)		(0.002)		(0.001)
Two girls	0.241	0.419	0.240	0.373	0.239	0.420	0.238	0.373
		(0.001)		(0.001)		(0.001)		(0.001)
(1) Mixed sex	0.495	0.356	0.496	0.306	0.495	0.354	0.497	0.300
		(0.002)		(0.001)		(0.002)		(0.001)
(2) Same sex	0.505	0.410	0.504	0.367	0.505	0.411	0.503	0.366
		(0.001)		(0.001)		(0.001)		(0.001)
Difference, (2)-(1)		0.055		0.061		0.058		0.066
		(0.001)		(0.001)		(0.001)		(0.001)

To prove the availability of $same\ sex$ we next need to prove that $same\ sex$ is uncorrelated with other unobserved factors. And demographic variables show no distinction between different groups divided by $same\ sex$.

Table 5—Difference in Means of Demographic Variables by Same sex

	Mixed sex - Same sex	
Variable	1980 PUMS	1990 PUMS
Age	0.0228	-0.0174
	(0.0139)	(0.0112)
Age at first birth	-0.0007	0.0074
	(0.0097)	(0.0114)
Black	-0.0002	-0.0021
	(0.0008)	(0.0011)
White	0.0002	0.0006
	(0.0009)	(0.0013)
Other races	0.0000	0.0014
	(0.0004)	(0.0009)
Hispanic	0.0000	0.0007
	(0.0004)	(0.0010)
Length of education	0.0085	-0.0100
	(0.0064)	(0.0074)

Table 5 reports a T-test of demographic variables by $same\ sex$ groups. None of the

differences within each contrast is significant at even 10% significance level. Thus, we have confidence to believe that $same\ sex$ is randomly assigned and thus it is available to be our instrument.

iii. Estimation

A. 2SLS Estimation

As we have learned in the *problem 15.3* from homework of Chapter 15 in textbook, when using dummy as instrumental variable in a simple regression model,

$$y = \beta_0 + \beta_1 x + \mu$$

the IV estimator $\hat{\beta}_1$ can be written as

$$\hat{\beta}_1 = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0}$$

where $\bar{y}_0$ are the sample averages of y_i over the part of the sample with $z_i = 0$, and $\bar{y}_1, \bar{x}_0$ and $\bar{x}_1$ are similarly defined. This estimator, known as a *grouping estimator*, was first suggested by Wald (1940).

While the Wald estimates can provide a simple demonstration of how the covariates, that is, the fertility variables affect the dependent variables, that is, the labor supply variables. We still embark on the estimation by using 2SLS estimates and compare it with OLS estimates because during the time when this term paper was being written we have numerous tasks including preparation of four subjects and three term papers, as a result of which we have no enough time to proceed this Wald estimator as other economic papers would do.

However, before doing the regression we have something interesting to notice. The instrumental variable can be written as

$$same\ sex = b_1 b_2 + (1 - b_1)(1 - b_2)$$

Where b_j denotes the j-th birth to be boy. Then we have

$$same\ sex = 2b_1 b_2 - b_1 - b_2 + 1$$

Regress $same\ sex$ on b_1 we have the coefficient to be $2b_2 - 1$. Thus the sex of first birth does not affect the $same\ sex$ only if $2b_2 - 1$ is significantly different from

zero. Which means b_2 should be significantly different from 1/2. And this is, although satisfied from result in table 3, we still concern that the sex of each birth will have positive correlation with $same\ sex$. Another concern is that although this is data from America, parents and society may still treat boys and girls differently, otherwise there will not be feminism prevalent in this country. Hence, if we drop b_1 and b_1 there might be potential omitted variable problem. Thus, we will still include b_1 and b_2 into regression model.

The following regression model is what we want to estimate. The dependent variable y represents the labor supply variables. The available covariates that can represent the fertility we use is a dummy $More\ than\ 2$ which equals to 1 if women have more than 2 children and. And other exogenous demographic variables are represented by a vector $\boldsymbol{d}$ including: mother's age, mother's age at first birth, racial dummies. And b_1 and b_2 are dummies which equal 1 if the j-th birth is boy.

$$y = \alpha_0 \boldsymbol{d} + \alpha_1 s_1 + \alpha_2 s_2 + \beta\, x + \mu$$

The first stage equation using $same\ sex$ as instrument is

$$x = \pi_0 \boldsymbol{d} + \pi_1 s_1 + \pi_2 s_2 + \gamma(same\ sex) + \varepsilon$$

B. 2SLS Results

The first stage results regressing fertility on $same\ sex$ are reported in Table 6. In the top half of the panel we use data from 1980 PUMS and in the bottom half of the panel we use 1990 PUMS. Women in 1980 with same sex children are estimated to be 5.59% more likely to have a third child and 5.91% for married women. And both estimates are very significant with t-statistics almost equals to 50.

Next step is using $same\ sex$ as instrumental variable and performing the second stage regression. Table 7 reports comparison between OLS estimates and 2SLS estimates of different labor supply estimators.

Table 6—FIRST STAGE RESULT OF MORE THAN 2

	All women		Married Women	
Covariates	(1)	(2)	(4)	(5)
1980 PUMS				
First child was boy	-	-0.0066	-	-0.0078
		(0.0011)		(0.0012)
Second child was boy	-	-0.0074	-	-0.0089
		(0.0011)		(0.0012)
Same sex	0.0547	0.0559	0.0577	0.0591
	(0.0012)	(0.0011)	(0.0013)	(0.0012)
Other covariates	no	yes	no	yes
R-squared	0.0032	0.1177	0.0035	0.1208
1990 PUMS				
First child was boy	-	-0.0080	-	-0.0097
		(0.0015)		(0.0017)
Second child was boy	-	-0.0002	-	-0.0011
		(0.0015)		(0.0017)
Same sex	0.0613	0.0617	0.0663	0.0703
	(0.0012)	(0.0015)	(0.0013)	(0.0017)
Other covariates	no	yes	no	yes
R-squared	0.0037	0.1099	0.0044	0.1209

OLS estimates reports negative impact of fertility on labor supply with all estimators to being significant. For all women, having an additional third child would reduce labor force participation probability of women for 15.9% and reduce working duration for about 8 weeks. The labor income would reduce for $3,682.45 and family income will drop by 11.7%. For married women the result is similar.

2SLS reports results with same sign as OLS estimates but are much smaller in magnitude with greater variance but are still significant except for models of which ln(*Family income*) is dependent variable, which was significant in OLS estimates but is insignificant in 2SLS estimates.

TABLE 7—OLS AND 2SLS ESTIMATES USING 1980 PUMS

	All women		Married women	
	(1)	(2)	(3)	(4)
Instrumental Variable	-	Same sex	-	Same sex
Dependent variable:				
Worked last year	-0.159	-0.085	-0.151	-0.094
	(0.001)	(0.021)	(0.001)	(0.022)
Weeks worked	-8.04	-4.37	-7.45	-4.61
	(0.06)	(0.95)	(0.06)	(0.96)
Labor income	-3682.45	-1225.87	-3281.25	-1417.49
	(30.47)	(501.93)	(31.90)	(495.17)
ln(*Family income*)	-0.117	-0.042	-0.127	-0.033
	(0.003)	(0.055)	(0.003)	(0.047)
ln(*Non-wife income*)	-	-	-0.046	0.039
	-	-	(0.004)	(0.062)

The 2SLS estimates suggest that women with a third children will have 9% less possibility of participating in labor force, and weeks worked will decline by about 4 weeks. The labor income of women will decline by 1225.87 dollars for all women group and 1417.49 dollars for married women. It is not surprising that married women will face a greater decline in labor income because they have spouse to depend on and thus can reduce more labor supply.

In contrast to the female labor-supply estimates, the 2SLS shows insignificant relationship between fertility and family income which is out of expectation, and since *Non-wife income* is in significant, it seems like the compensated wage increase of spouse when mother takes care of children is not the case. And this part remains unclear for us.

iv. Conclusion

The results are consistent as predicted by theoretical models, female labor supply and fertility are negatively correlated. Traditional OLS estimates reports correct direction of the impact of fertility on female labor supply but tend to exaggerate the magnitude. Using IV approach we first prove that sex composition of first two children can be a valid instrument and the estimators show smaller effect.

This paper is inspired by a problem in a homework of Labor Economics which asks

a few intuitive questions about paper Angrist and Evans (1978). This is not a simple replication because AER website does not store the periodical of this paper and the data set. We perform all the process from data collection to the data manipulation by ourselves and since we do not have enough knowledge to perform data clean, the results have a little difference from the original paper.

v. Reference

[1]. Becker, G. S. , & Lewis, H. G. . (1973). On the interaction between the quantity and quality of children. Journal of Political Economy, 81(1-2), 113.

[2]. Wolpin, R. K. I. . (1980). Testing the quantity-quality fertility model: the use of twins as a natural experiment. Econometrica, 48(1), 227-240.

[3]. Welch, B. P. . (1976). Do sex preferences really matter?. The Quarterly Journal of Economics, 90(2), 285-307.