ECON 7110 Empirical Analysis

Feng Zixiao
21440328

April 17, 2022

Abstract: This paper aims to replicate the main results of another paper "Children and Their Parents' Labor Supply: evidence from Exogenous Variation in Family Size" by Joshua D. Angrist and William N. Evans (Angrist & Evans, 1998). The original paper discussed the causal relationship between childbearing and women's labor supply in the U.S.. The paper solved the endogeneity problem caused by the fertility variables by using sibling sex mix as a new Instrumental Variable to estimate the causal relation running from fertility to work effort.

I. INTRODUCTION

The relationship between fertility and women's labor supply has been a well discussed topic in the field of Labor Economics, to study this relationship, a variety of theoretical models and empirical works have been developed. Childbearing is believed to be a strong influencing factor of how family composition can affect the labor market. Finding this relationship can also shed light on many other relevant and important issues such as the low fertility rate in the U.S. and the drop in the women's labor participation between the 1970s and 80s.

Many efforts have been made to find the causal relationship between fertility and women's labor supply, however, the conventional method often failed to capture the relationship. The conventional OLS model which has the labor outcome as the dependent variable and the fertility measures as the independent variable often suffers from a critical drawback, that most of the fertility variables are endogenous. Fertility itself is involved in a very complicated mechanism and it may correlate with unobservable factors that are also correlated with women's labor outcome. If this happens, the fertility and labor outcome are jointly determined by the unobservable error, making the OLS model biased.

To overcome this endogenous fertility variable issue, we adopted the sex-mix estimation strategy used by Angrist and Evans (1998), which is to develop a new Instrumental Variable (IV) based on the sibling sex composition. The design behind this strategy is to take advantage of parents' preference over children of mixed sex. Empirical evidence has shown that parents with children of the same sex are more likely to give birth to another child. Because the sex of child is largely a random selection, and there is little evidence of sex selection problem in the U.S., therefore, the sibling sex composition is exogenously determined in our study.

In this paper we only focused on a subsample where only married women with more than two children and are between 21-35 years old are considered. We first estimate the impact of having more than two children and having one additional child on labor outcomes using OLS model, then we will use a binary indicator of whether the first two children are of the same sex as our first instrument to estimate the Wald estimators, at last, we will add more exogenous regressors to the regression and use Two-Stage Least Squares method to find more precise results. Our Two-Stage Least Squares estimation has two identifications, the first one is a just identified model with one instrument for one endogenous variable and the alternative one is an overidentified model where we arranged two instruments for that one endogenous variable. The benefit of such arrangement will be discussed later.

The primary purpose of this paper is to find the causal link between fertility and women's labor supply, we try to find the effect of having more than two children on mothers' labor participation and her labor income. Section II discusses the statistics summary and the description of the interested sample. Section III reports the empirical results starts from the Wald estimates and then the Two-Stage Least Squares estimates. Section IV is the conclusion.

II. DATA

To be able to implement the mix-sex estimation strategy, we obtained data from the Census Public Use Micro Sample (PUMS), which contains information on the labor supply, the number of children and the children's sex composition for each mother. To obtain a more representative studying sample, we combine the subsamples from three different years, namely the year of 2009, 2014 and 2016 to provide a sufficient sample size. And to have a variation of the children sex composition, our studying sample is further narrowed down to comprised of only married women who are aged between 21 to 35 with at least two children and who are the head of the household or the spouse of the household head. Another concern with regards to our studying sample is the possibility of confusion caused by multiple families in the same household. The reason behind this concern is that if the number children claimed by the mother is different with that of the father in any household, then the sex composition cannot be clearly defined, this situation could be caused by stepchildren or other complicated family relationships that cannot be specified in this paper. To address this concern, we got rid of the individuals whose claimed number of children is different with that of their children's fathers. After all these processing work, our studying sample contains 77175 observations. Table 1 shows some basic statistics of the key variables used in the estimation.

Table 1 - Statistics Summary, Women Aged 21-35 With At Least 2 Children

	Means and standard deviations
	PUMS Combined
Variable	Married women
More than 2 children (=1 if mother had more than 2 children, =0 otherwise)	0.43
	(0.49)
Boy 1st (= 1 if first child was a boy)	0.52
	(0.5)
Boy 2nd (= 1 if second child was a boy)	0.51
	(0.5)
Two boys (=1 if first two children were boys)	0.27
	(0.44)
Two girls (=1 if first two children were girls)	0.24
	(0.42)
Same sex (=1 if first two children were the same sex)	0.5
	(0.5)
Age	31.28
	(3.14)
Age at first birth (parent's age in years when first child was born)	22.53
	(4.12)
worked last year (=1 if worked in the previous year)	0.66
	(0.47)
Labor income (labor earnings in year prior to census, in 1995 dollars)	19159.79
	(28258.5)
Weeks worked (weeks worked in year prior to census)	3.28
	(2.73)
Hours/week (average hours worked per week)	22.67
	(19.18)
Family income (family income in year prior to census, in 1995 dollars)	78256.15
	(67340.88)
Number of observations	77175

Note: the sample incudes married women aged 21 - 35 with at least two children in 2007, 2014 and 2016 PUMS.

The table above already revealed some important demographic information. The first halve of the table reported the children sex indicators, the mothers' age and their age when giving their first birth. Within our focused sample 43% of the married mothers have more than two children, this gives us a relatively good representation if we want to test the difference between the group with more than two children with the rest of the sample. We also have indicators representing the sex of both the first and the second child, we can that the probability of having a boy is a little above 50%. The *"two boys"* and *"two girls"* are the dummy variables of whether the first children are both boys or both girls, the *"same sex"* represents whether their first two children are of the same sex, as given in the table, the mean for *"same sex"* is the 0.5, this indicates that children with each sex is equally likely to be born, and there is unlikely to have any sex selection issue in our sample. Finally, the average age of our sample is 31.28 years old and our sampled married mothers gave their first birth at an average age of 22.53. The labor outcomes are reported in the second halve of the above table, we included labor outcomes such as whether the individual has worked in the last year (66%, a majority of mothers with at least two children has worked, this is a high labor participation rate for this cohort), the mothers' labor income in 1995 dollars, the number of weeks worked in the previous year and the average number of hours worked per week, as well as, the family income.

Our focused sample includes only married mothers who have at least two children and are between 21-35 years old, and we intend to use this sample to generalize the result for the larger population. People may argue that this high fertility cohort only constitutes a minority of the whole population and hence is misrepresentative of the whole population. Angrist and Evans (Angrist & Evans, 1998) addressed this question with statistics generated from a much large sample, the 1990 Current Population Survey (CPS) and found that over half of all women aged 28 – 35 belongs to this cohort. Therefore, our sample would render a well representation of the population.

III. Empirical Results

A. Wald Estimation

After knowing the basic set up of our mix-sex estimation strategy, we first deploy a Wald estimation to identify the effect of fertility on women's labor supply. Consider the simple OLS regression shown in equation (1), the dependent variable is the labor supply of our choice and the independent variable is the fertility measure. As we discussed before, most fertility measures suffer from endogeneity problem that violate the unbiasedness assumption of the OLS model, therefore, we introduce a binary instrument *"same sex"* which indicate whether the first two children of an individual are of the same sex, this instrument is denoted as z_i . z_i is a valid instrument because it satisfied both the Exclusion Restriction and the Relevance condition.

$$y_i = \alpha + \beta x_i + \varepsilon_i \qquad (1)$$

By definition, the Wald estimator equals to the ratio of difference in mean outcome between treatment group and control group and the difference in the endogenous variable between the treatment group and the control group. In the terms of mathematics, it can be also expressed as the ratio of two conditional mean differences, as in Equation (2). The numerator is the mean difference of women's labor outcome when their first two children are of the same sex and when They have a mixed sex composition, and the denominator stands for the mean difference between women's fertility when there is no mixed sex composition versus when there is mixed sex composition. By using the Wald estimator, we are able to capture the effect of women's childbearing on their labor effort without have to violate the unbiasedness assumption in the OLS model.

$$\rho_{IV} = \frac{E[Y_i|Z_i=1] - E[Y_i|Z_i=0]}{E[X_i|Z_i=1] - E[X_i|Z_i=0]} \qquad (2)$$

The IV estimator is essentially the ratio of the coefficient of the first-stage regression which is Equation (3) and the coefficient of the reduced-form regression which is Equation (4). In this way the IV estimators are expected to render similar results as the Wald estimators.

$$x_i = \beta_i + \pi_1 z_i + \zeta_{1i} \qquad (3)$$
$$y_i = \beta_i + \pi_2 z_i + \zeta_{2i} \qquad (4)$$

In this section of our study, we use *"same sex"* to be the only instrument, and there are two fertility measures, the *"More than 2 children"* and the "Number of children" to represent women's childbearing choice to offer some comparison. Four labor outcomes are studied they are whether the individual was in the labor force in the previous year, the number of weeks worked, the hours worked per week and their labor income. Table 2 shows detailed result of our

Wald estimations. The first two rows are the first-stage regressions of both endogenous fertility measures on *"same sex"*, both yields positive coefficients meaning that mothers have the same sex composition are more likely to have additional child. The third row gives us the reduced-form of whether the individual was in the labor force or not, and it's two Wald estimators on the two fertility measures. The results show a negative correlation between women's fertility and their labor supply. Numerically, having more than two children reduced the woman's chance of participate in the work force by 23.22%, and an additional children would reduce this chance by about 17%, both coefficients are statistically significant. The last three variables are estimated in the same way and all suggest that higher fertility has a negative impact on all kind of labor participation of women as well as their labor income.

Table 2 - Wald Estimations of Labor Supply Models

	PUMS Combined		
		Wald Estimate using as covariate	
Variable	*Mean difference by Same sex*	*More than 2 children*	*Number of children*
More than 2 children	0.051691 (0.003554)	-	-
Number of children	0.070218 (0.006166)	-	-
Worked last year	-0.012003 (0.003419)	-0.2322 (0.06588)	-0.17093 (0.04869)
Weeks worked	-0.06472 (0.01969)	-1.252 (0.377)	-0.9216 (0.2781)
Hours per week	-0.56516 (0.13808)	-10.934 (2.667)	-8.049 (1.981)
Labor Income	-544.7 (203.4)	-10538 (3898)	-7758 (2876)

Note: Same sample as in Table 1, Standard Errors are reported in parentheses.

Using the method mentioned above the calculate the IV estimators of each dependent variables yield similar results as the Wald estimators.

B. 2SLS Estimation

In this section we imitate the two -Stage Least Squares Estimation developed by Angrist and Evans (1998), there are three reasons for this adoption. First, the 2SLS estimation allows us to add more exogenous covariates such as individuals' age, their race etc. into our regression and have these effects controlled for. Although empirical evidence suggests that our instrument is not related to exogenous covariates, including them into the model can still improve the accuracy of our estimates.

Second, the 2SLS estimation helps to keep any secular effects of child sex while using *"Same sex"* as instrument. These unobservable effects could be caused by parents' differential treatment or attitudes toward children of different sex or the fact that boys usually suffer from a higher disability rate than girls, and a disabled child may change the way parents raise their children. Through a 2SLS estimation we can add the *"Boy 1st"* and *"Boy 2nd"* regressors to control for errors from these effects.

The third benefit of using 2SLS estimation is that we are able to break the first instrument *"Same sex"* down into two other instruments *"two boys"* and *"two girls"* to generate an overidentified model. By using the alternative instruments we can identify the difference between the effects of both children are boys and both children are girls.

Let s_1 denotes a binary variable equals to one if the first child is a boy and let s_2 denotes another binary variable equals to one if the second child is a boy. We then include s_1 and s_2 into a multiple linear regression to account for the secular effects mentioned before:

$$y_i = \alpha'_0 \boldsymbol{W}_i + \alpha_1 s_{1i} + \alpha_2 s_{2i} + \beta x_i + \varepsilon_i \qquad (5)$$

$\boldsymbol{W}_i$ in the equation is a vector of exogenous variables that are not related to fertility such as mothers' age and the age at their first birth, s_{1i} and s_{2i} represent the sex of the first and second child of mother *i*, s_i equals to one if the child is a boy. Without adding the alternative instruments the first-stage equation is

$$x_i = \pi'_0 \boldsymbol{W}_i + \pi_1 s_{1i} + \pi_2 s_{2i} + \gamma(Same\ sex_i) + \eta_i \qquad (6)$$

Then, we construct the overidentified model, where we includes two instruments (*"two boys"* and *"two girls"*) for one endogenous variable x_i , because s_{1i} and s_{2i} has linear relationship with the instruments we drop the regressor s_{2i} to avoid the perfect multicollinearity problem. The first-stage equation now looks like:

$$x_i = \pi'_0 \boldsymbol{W}_i + \pi_1 s_{1i} + \pi_2 s_{2i} + \gamma_0(Two\ boys_i) + \gamma_1(Two\ girls_i) + \eta_i \qquad (7)$$

Table 3 reports three first-stage results discussed above with *"More than 2 children"* as the endogenous variable. Column (1) represents the multiple regression with *"Same sex"* as the explanatory variable and including additional regressors like mothers' age and age at first birth and mothers' race to keep these effects constant during our estimation. Column (2) represents Equation (6) mentioned above, which still uses *"Same sex"* as the explanatory variable, but included s_{1i} and s_{2i} as additional regressors. Column (3) is stated by Equation (7) where we decompose *"Same sex"* into *"two boys"* and *"two girls"* to distinguish the effect of sex composition by different sexes.

Table 3 - OLS Estimates Of More Than 2 Children Equations

	Married women		
Independent Variable	(1)	(2)	(3)
Boy 1st	-	-0.0059294	-0.0029397
		(0.0033284)	(0.0047098)
Boy 2nd	-	-0.0029896	-
		(0.0033287)	
Same sex	0.0498391	0.0500853	-
	(0.0033253)	(0.0033284)	
Two boys	-	-	0.0470957
			(0.0046184)
Two girls	-	-	0.053075
			(0.0047945)
With other covariates	no	yes	yes
R Square	0.1273	0.1273	0.1273

Note: Other covariates in the models are indicators for Age, Age at first birth, Black, Hispanic, and Other race. The variable Boy 2nd is excluded from columns (3) and (6). Standard errors are reported in parentheses.

Next, we tested the effect of *"More than 2 children"* on a variety of labor participation and earning outcomes with the effects of exogenous variables controlled for. Table 4 provides a comparison among the OLS estimation and 2SLS estimations using different instrument strategies. The OLS estimation suggests that having a third child decreases a woman's chance of working by 16%, women with more than two children work one week fewer in a year and works seven fewer hours per week, their labor income is also dropped by 6860 dollars on average, all of these effects are statistically significant. Column (2) reports the results of 2SLS estimation with one instrument "*Same sex"*, by contributing the effect of having children of the same sex to the effect of having a third child on labor outcomes, we are able to find a strong negative relationship between fertility and women's labor supply. This model suggests having a third child can reduce the chance of working by 23%, 1.21 fewer weeks worked per year, about 11 fewer hours worked per week and 8771 dollars lesser in labor income, again all effects are statistically significant and overall, the magnitude of the effects are larger than that of the OLS model. Column (3) represents another 2SLS model with all else same as the one in Column (2) but having two instruments instead of one. By separating the effects of having two boys and having two girls, we obtained a new set of estimates that are slightly higher than in Column (2), having more than two children reduces the chance of working by 24%, leads to 1.25 fewer weeks worked per year, 11.43 fewer hours worked per week and 9110 dollars lesser in labor income. Overall, we can conclude an obvious negative relationship between fertility and women's labor supply.

Table 4 -OLS And 2sls Estimates Of Labor-Supply Models

	Married women		
	(1)	(2)	(3)
Estimation method	OLS	TSLS	TSLS
Instrument for More than 2 children	-	Same sex	Two boys, Two girls
Dependent variable			
Worked last year	-0.1613455	-0.230548	-0.239338
	(0.0036193)	(0.067336)	(0.066846)
Weeks worked	-1.057178	-1.21771	-1.2469

	(0.0207)	(0.38436)	(0.38138)
Hours per week	-7.27387	-10.99104	-11.43209
	(0.14552)	(2.71231)	(2.6937)
Labor income	-6860.19	-8771.9	-9109.9
	(212.52)	(3946.7)	(3916.2)

Note: The table reports estimates of the coefficient on the More than 2 children variable in equations (4) and (6) in the text. Other covariates in the models are Age, Age at first birth, plus indicators for Boy 1st, Boy 2nd, Black, Hispanic, and Other race. Standard errors are reported in parentheses.

IV. Conclusion

Fertility has been considered a major factor influencing women's labor supply, and extensive previous literatures have tried to find the causal relationship between them. However, the fertility variables are often endogenous and jointly determined with the labor outcomes by some unobservable factors, making the conventional OLS model difficult to interpret any causal relationship. Aware of that, we adopted the mix sex estimation strategy used by Angrist and Evans (1998) which takes advantage of the fact that parents with two children of the same sex are more like to have a third child than their counterparts who already have a mixed sex composition.

We focused on a sample involves only married women who has at least two children and aged between 21-35. We then implemented the OLS, Wald and Two-Stage Least Squares estimations to identify the general effect of childbearing on women's labor efforts. Two different identifications are used in the Two-Stage Least Squares estimations, the first one used whether the first two children are of the same sex as the instrument for whether a mother has more than two children or not. And the second is an overidentified model where *"Same sex"* is replace by two alternative instruments *"Two boys"* and *"Two girls"* to separate the effect of *"Same sex"* when the first two children are both boys and when they are both girls.

After taking account for the effects of the sex of each of the first two children (s_{1i} and s_{2i}), and including exogenous regressors such as mothers' age and race, we found a strong negative relationship between childbearing and women's labor supply. Our Two-Stage Least Squares estimation with separate instruments for two boys and two girls suggests that having more than two children reduces the mother's chance of working by 24%, leads to 1.25 fewer weeks worked per year, 11.43 fewer hours worked per week and 9110 dollars lesser in labor income.

Reference:

Angrist, J., & Evans, W. (1998). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *American Economic Association*, *88*(3), 450–477. http://www.jstor.org/stable/116844?origin=JSTOR-pdf

Appendix A: R Results

Table 1

```
> vars <- c('morethan2kids', 'boy1st', 'boy2nd', 'twoboys', 'twogirls',
'firstsex1',  'secondsex1',  'samesex', 'AGE',  'age1st',  'worklastyr',
'INCWAGE', 'WKSWORK2', 'UHRSWORK', 'FTOTINC')
> sink("Table 2.txt")
> describe(ACS_combined[vars], fast = T)
> sink()
```

	vars	n	mean	sd	min	max	range	se
morethan2kids	1	77175	0.43	0.49	0	1	1	0.00
boy1st	2	77175	0.52	0.50	0	1	1	0.00
boy2nd	3	77175	0.51	0.50	0	1	1	0.00
twoboys	4	77175	0.27	0.44	0	1	1	0.00
twogirls	5	77175	0.24	0.42	0	1	1	0.00
firstsex1	6	77175	1.48	0.50	1	2	1	0.00
secondsex1	7	77175	1.49	0.50	1	2	1	0.00
samesex	8	77175	0.50	0.50	0	1	1	0.00
AGE	9	77175	31.28	3.14	21	35	14	0.01
age1st	10	77175	22.53	4.12	-22	35	57	0.01
worklastyr	11	77175	0.66	0.47	0	1	1	0.00
INCWAGE	12	77175	19159.79	28258.50	0	665000	665000	101.72
WKSWORK2	13	77175	3.28	2.73	0	6	6	0.01
UHRSWORK	14	77175	22.67	19.18	0	99	99	0.07
FTOTINC	15	77175	78256.15	67340.88	-19600	1253000	1272600	242.40

Table 2

```
> # Equa 1
> equa1 <- lm(NCHILD ~ samesex, data = ACS_combined)
> summary(equa1)
> # Equa 2
> equa2 <- lm(morethan2kids ~ samesex, data = ACS_combined)
> summary(equa2)
```

```
Call:
lm(formula = NCHILD ~ samesex, data = ACS_combined)

Residuals:
    Min      1Q  Median      3Q     Max
-0.6373 -0.6373 -0.5670  0.4330  6.4330

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.567048   0.004366  587.97   <2e-16 ***
samesex     0.070218   0.006166   11.39   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8564 on 77173 degrees of freedom
```

```
Multiple R-squared:  0.001678,    Adjusted R-squared:  0.001665
F-statistic: 129.7 on 1 and 77173 DF,  p-value: < 2.2e-16


Call:
lm(formula = morethan2kids ~ samesex, data = ACS_combined)

Residuals:
    Min      1Q  Median      3Q     Max
-0.4509 -0.4509 -0.3992  0.5491  0.6008

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.399194   0.002517  158.62   <2e-16 ***
samesex     0.051691   0.003554   14.54   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4937 on 77173 degrees of freedom
Multiple R-squared:  0.002733,    Adjusted R-squared:  0.00272
F-statistic: 211.5 on 1 and 77173 DF,  p-value: < 2.2e-16
```

```
> # Equa 3
> equa3 <- lm(worklastyr ~ samesex, data = ACS_combined) # reduced form
> summary(equa3)
```

```
Call:
lm(formula = worklastyr ~ samesex, data = ACS_combined)

Residuals:
    Min      1Q  Median      3Q     Max
-0.6625 -0.6505  0.3375  0.3495  0.3495

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.662500   0.002421 273.682  < 2e-16 ***
samesex     -0.012003   0.003419  -3.511 0.000447 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4749 on 77173 degrees of freedom
Multiple R-squared:  0.0001597,    Adjusted R-squared:  0.0001467
F-statistic: 12.33 on 1 and 77173 DF,  p-value: 0.0004468
```

```
> # Equa 4: TSLS
> equa4 <- ivreg(worklastyr ~ NCHILD | samesex, data = ACS_combined)
> summary(equa4)
> # Equa 5: TSLS
```

> equa5 <- ivreg(worklastyr ~ morethan2kids | samesex, data = ACS_combined)
> summary(equa5)

```
Call:
ivreg(formula = worklastyr ~ NCHILD | samesex, data = ACS_combined)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.7594 -0.5885  0.2406  0.2406  1.4371 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.10129    0.12671   8.691  < 2e-16 ***
NCHILD      -0.17093    0.04869  -3.511 0.000447 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4749 on 77173 degrees of freedom
Multiple R-Squared: 7.128e-05,	Adjusted R-squared: 5.832e-05 
Wald test: 12.33 on 1 and 77173 DF,  p-value: 0.0004471 


Call:
ivreg(formula = worklastyr ~ morethan2kids | samesex, data = 
ACS_combined)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.7552 -0.5230  0.2448  0.2448  0.4770 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)    0.75519    0.02806  26.913  < 2e-16 ***
morethan2kids -0.23220    0.06588  -3.524 0.000425 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.473 on 77173 degrees of freedom
Multiple R-Squared: 0.007741,	Adjusted R-squared: 0.007729 
Wald test: 12.42 on 1 and 77173 DF,  p-value: 0.0004248
```

Table 3

```
# Equa 6
equa6 <- lm(morethan2kids ~ AGE + age1st + samesex + factor(RACE), data = ACS_combined)
summary(equa6)
# Equa 7
equa7 <- lm(morethan2kids ~ AGE + age1st + samesex + boy2nd + boy1st + factor(RACE), data = 
ACS_combined)
summary(equa7)
# Equa 8
```

equa8 <- lm(morethan2kids ~ AGE + age1st + twoboys + twogirls + boy1st + factor(RACE), data = ACS_combined)
summary(equa8)

```
Call:
lm(formula = morethan2kids ~ AGE + age1st + samesex + factor(RACE),
    data = ACS_combined)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4675 -0.3761 -0.1955  0.4682  1.0327

Coefficients:
                Estimate Std. Error  t value Pr(>|t|)
(Intercept)    0.4550796  0.0170454   26.698  < 2e-16 ***
AGE            0.0308795  0.0005812   53.129  < 2e-16 ***
age1st        -0.0451532  0.0004484 -100.700  < 2e-16 ***
samesex        0.0498391  0.0033253   14.988  < 2e-16 ***
factor(RACE)2  0.0002814  0.0073469    0.038   0.9694
factor(RACE)3  0.0933338  0.0168442    5.541 3.02e-08 ***
factor(RACE)4 -0.1213382  0.0225715   -5.376 7.65e-08 ***
factor(RACE)5 -0.1185036  0.0507419   -2.335   0.0195 *
factor(RACE)6 -0.1059287  0.0087553  -12.099  < 2e-16 ***
factor(RACE)7  0.0011917  0.0073246    0.163   0.8708
factor(RACE)8 -0.0026934  0.0120198   -0.224   0.8227
factor(RACE)9  0.0389065  0.0356816    1.090   0.2755
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4619 on 77163 degrees of freedom
Multiple R-squared:  0.1273, Adjusted R-squared:  0.1272
F-statistic:  1023 on 11 and 77163 DF,  p-value: < 2.2e-16


Call:
lm(formula = morethan2kids ~ AGE + age1st + samesex + boy2nd +
    boy1st + factor(RACE), data = ACS_combined)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4689 -0.3747 -0.1941  0.4647  1.0342

Coefficients:
                Estimate Std. Error  t value Pr(>|t|)
(Intercept)    0.4596445  0.0172171   26.697  < 2e-16 ***
AGE            0.0308758  0.0005812   53.122  < 2e-16 ***
age1st        -0.0451514  0.0004484 -100.697  < 2e-16 ***
samesex        0.0500853  0.0033284   15.048  < 2e-16 ***
boy2nd        -0.0029896  0.0033287   -0.898   0.3691
boy1st        -0.0059294  0.0033284   -1.781   0.0748 .
factor(RACE)2  0.0002039  0.0073471    0.028   0.9779
factor(RACE)3  0.0932196  0.0168443    5.534 3.14e-08 ***
factor(RACE)4 -0.1213848  0.0225713   -5.378 7.56e-08 ***
factor(RACE)5 -0.1181782  0.0507415   -2.329   0.0199 *
```

```
factor(RACE)6 -0.1059559  0.0087556  -12.102  < 2e-16 ***
factor(RACE)7  0.0011114  0.0073247    0.152   0.8794    
factor(RACE)8 -0.0026808  0.0120196   -0.223   0.8235    
factor(RACE)9  0.0385889  0.0356815    1.081   0.2795    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4619 on 77161 degrees of freedom
Multiple R-squared:  0.1273, Adjusted R-squared:  0.1272 
F-statistic: 866.1 on 13 and 77161 DF,  p-value: < 2.2e-16


Call:
lm(formula = morethan2kids ~ AGE + age1st + twoboys + twogirls + 
    boy1st + factor(RACE), data = ACS_combined)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.4689 -0.3747 -0.1941  0.4647  1.0342 

Coefficients:
                Estimate Std. Error  t value Pr(>|t|)    
(Intercept)    0.4566549  0.0171999   26.550  < 2e-16 ***
AGE            0.0308758  0.0005812   53.122  < 2e-16 ***
age1st        -0.0451514  0.0004484 -100.697  < 2e-16 ***
twoboys        0.0470957  0.0046184   10.197  < 2e-16 ***
twogirls       0.0530750  0.0047945   11.070  < 2e-16 ***
boy1st        -0.0029397  0.0047098   -0.624   0.5325    
factor(RACE)2  0.0002039  0.0073471    0.028   0.9779    
factor(RACE)3  0.0932196  0.0168443    5.534 3.14e-08 ***
factor(RACE)4 -0.1213848  0.0225713   -5.378 7.56e-08 ***
factor(RACE)5 -0.1181782  0.0507415   -2.329   0.0199 *  
factor(RACE)6 -0.1059559  0.0087556  -12.102  < 2e-16 ***
factor(RACE)7  0.0011114  0.0073247    0.152   0.8794    
factor(RACE)8 -0.0026808  0.0120196   -0.223   0.8235    
factor(RACE)9  0.0385889  0.0356815    1.081   0.2795    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4619 on 77161 degrees of freedom
Multiple R-squared:  0.1273, Adjusted R-squared:  0.1272 
F-statistic: 866.1 on 13 and 77161 DF,  p-value: < 2.2e-16
```

Table 4

```
# Equa 9
equa9 <- lm(worklastyr ~ morethan2kids + AGE + age1st +  factor(RACE), data = ACS_combined)
summary(equa9)
# Equa 10
equa10 <- ivreg(worklastyr ~ morethan2kids + AGE + age1st +  factor(RACE)
            | samesex + AGE + age1st +  factor(RACE), data = ACS_combined)
summary(equa10)
```

Equa 11

```
equa11 <- ivreg(worklastyr ~ morethan2kids + AGE + age1st +  factor(RACE)
           | twoboys + twogirls + AGE + age1st + factor(RACE), data = ACS_combined)
summary(equa11)
```

```
Call:
lm(formula = worklastyr ~ morethan2kids + AGE + age1st + factor(RACE), 
    data = ACS_combined)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.1569 -0.5698  0.2597  0.3551  0.7771 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    0.3046145  0.0171575  17.754  < 2e-16 ***
morethan2kids -0.1613455  0.0036193 -44.579  < 2e-16 ***
AGE            0.0199462  0.0005958  33.481  < 2e-16 ***
age1st        -0.0088211  0.0004802 -18.371  < 2e-16 ***
factor(RACE)2  0.1095999  0.0073972  14.816  < 2e-16 ***
factor(RACE)3  0.0188075  0.0169628   1.109   0.2675    
factor(RACE)4 -0.0404195  0.0227299  -1.778   0.0754 .  
factor(RACE)5 -0.3113031  0.0510908  -6.093 1.11e-09 ***
factor(RACE)6 -0.1125310  0.0088235 -12.754  < 2e-16 ***
factor(RACE)7 -0.1075867  0.0073748 -14.589  < 2e-16 ***
factor(RACE)8 -0.0108408  0.0121019  -0.896   0.3704    
factor(RACE)9  0.0049822  0.0359260   0.139   0.8897    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.465 on 77163 degrees of freedom
Multiple R-squared:  0.04126,	Adjusted R-squared:  0.04112 
F-statistic: 301.9 on 11 and 77163 DF,  p-value: < 2.2e-16


Call:
ivreg(formula = worklastyr ~ morethan2kids + AGE + age1st + factor(RACE) 
| samesex + AGE + age1st + factor(RACE), data = ACS_combined)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.3295 -0.5356  0.2353  0.3569  0.8208 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)    0.337948   0.036670   9.216  < 2e-16 ***
morethan2kids -0.230548   0.067336  -3.424 0.000618 ***
AGE            0.022081   0.002159  10.228  < 2e-16 ***
age1st        -0.011948   0.003076  -3.884 0.000103 ***
factor(RACE)2  0.109614   0.007415  14.783  < 2e-16 ***
factor(RACE)3  0.025316   0.018141   1.396 0.162858    
factor(RACE)4 -0.048937   0.024240  -2.019 0.043507 *  
factor(RACE)5 -0.319479   0.051824  -6.165  7.1e-10 ***
factor(RACE)6 -0.119850   0.011349 -10.561  < 2e-16 ***
```

```
factor(RACE)7 -0.107503   0.007393 -14.542  < 2e-16 ***
factor(RACE)8 -0.011060   0.012132  -0.912 0.361963
factor(RACE)9  0.007710   0.036108   0.214 0.830924
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4661 on 77163 degrees of freedom
Multiple R-Squared: 0.03671, Adjusted R-squared: 0.03658
Wald test: 121.7 on 11 and 77163 DF,  p-value: < 2.2e-16


Call:
ivreg(formula = worklastyr ~ morethan2kids + AGE + age1st + factor(RACE)
| twoboys + twogirls + AGE + age1st + factor(RACE), data = ACS_combined)

Residuals:
    Min      1Q  Median      3Q     Max
-1.3514 -0.5308  0.2318  0.3583  0.8264

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.342182   0.036467   9.383  < 2e-16 ***
morethan2kids -0.239338   0.066846  -3.580 0.000343 ***
AGE            0.022353   0.002144  10.424  < 2e-16 ***
age1st        -0.012346   0.003055  -4.042 5.31e-05 ***
factor(RACE)2  0.109616   0.007419  14.774  < 2e-16 ***
factor(RACE)3  0.026143   0.018135   1.442 0.149425
factor(RACE)4 -0.050019   0.024233  -2.064 0.039015 *
factor(RACE)5 -0.320517   0.051847  -6.182 6.36e-10 ***
factor(RACE)6 -0.120779   0.011321 -10.669  < 2e-16 ***
factor(RACE)7 -0.107493   0.007397 -14.531  < 2e-16 ***
factor(RACE)8 -0.011088   0.012140  -0.913 0.361058
factor(RACE)9  0.008056   0.036130   0.223 0.823551
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4664 on 77163 degrees of freedom
Multiple R-Squared: 0.03549, Adjusted R-squared: 0.03535
Wald test: 121.6 on 11 and 77163 DF,  p-value: < 2.2e-16
```