R: Assignment 5
Group 4, FIN 560

Cindy Lu

April 8, 2021

1   1

1.1   a

Compared to women with 2 or less kids, women with more kids worked 6 weeks less on average. The estimate was -6.0082.

1.2   b

The OLS is very inappropriate because there is an equilibrium situation where weeks worked and number of children are endogeneously determined via labor supply and demand. Endogeneity in OLS causes the error term to be correlated with the regressors, which causes biased and inconsistency.

1.3   c

Based on the 1st stage regression, the likelihood of having another child goes up by an expected 0.06682 if the variable samesex = 1 compared to the group where samesex=0. This effect is statistically significant because the t-stat reported is larger tha 11 (11.96), which is high evidence against the null and beats the critical values at conventional levels. The effect is sizeable given that it is a linear probability model.

1.4   d

To be a valid instrument, the F test and the t-statistic show that the null hypotheses corresponding to these tests can be rejected at significance levels commonly used. This variable is sufficiently correlated with the endogenous variable and satisfies the relevance condition via these statistical inference tests. The F test suggests that because the null that says that all coefficients are jointly not different from 0 statistically can be rejected with high confidence. It is valid if the F-statistic is larger than 10. Here the F-statistic was 143.1. The t-statistic was 11.96, and p-value much smaller than 0.01.

1.5   e

Examining the F-statistic, it is much larger than 10 in this 1st stage regression. Therefore this test suggests that because the null that says that all coefficients are jointly not different from 0 statistically can be rejected with high confidence.

1.6   f

-6.033 is the coefficient estimate produced by TSLS. Compared to women with 2 or less kids, women with more kids worked again around 6 weeks less on average.

1.7 g

Now the coefficient estimate produced by TSLS is -5.780746. It means that ceteris paribus, compared to women with 2 or less kids, women with more kids worked again around 5.780746 weeks less on average.

2   Appendix R code

```
#####################################################################
# R Assignnment 5
# Group 4, FIN 560
#
# Cindy Lu
#
# Last modified: 6-Apr-2021
#####################################################################

###############
# Load packages
###############

library(openxlsx)
library(AER)
library(sandwich)
library(lmtest)
library(xtable)

##################
# Question 1 part a
##################

# Read data
fer <- read.xlsx("C://Users/cindy.lu/Downloads/fertility_small.xlsx")

# Make sure categorical
#fer$morekids <- as.factor(fer$morekids)

# OLS
ols1 <- lm(weeksm1 ~ morekids, data = fer)
print(summary(ols1))

##################
# Question 1 part c
##################

# Change categorical
#fer$samesex <- as.factor(fer$samesex)

# First stage, regress endogenous on instrument candidate
stage1 <- lm(morekids ~ samesex, data = fer)
print(summary(stage1))

##################
# Question 1 part f
```

```
###################

# TSLS
fitTsls1 = ivreg(weeksm1 ~ morekids | samesex, data = fer)
print(summary(fitTsls1, diagnostics = TRUE))

# Adjust standard errors
print(coeftest(fitTsls1, vcov = vcovHC, type = 'HC1'))

###################
# Question 1 part f
###################

# TSLS with more exogenous variables
fitTsls2 = ivreg(weeksm1 ~ morekids +
                    agem1 + black + hispan + othrace |
                    samesex + agem1 + black + hispan + othrace,
                 data = fer)
print(summary(fitTsls2,
              diagnostics = TRUE))

# Adjust standard errors
print(coeftest(fitTsls2,
               vcov = vcovHC,
               type = 'HC1'))
```