Fertility and Weeks Worked per Year

Khalil A.

load("C:/Users/KhalilA/Downloads/fertility.RData")  
library(ivreg)

## Warning: package 'ivreg' was built under R version 4.2.3

library(car)

## Loading required package: carData

reg1 = lm(formula= weeksm1 ~ morekids, data= fertility)  
summary(reg1)

##   
## Call:  
## lm(formula = weeksm1 ~ morekids, data = fertility)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -21.07 -21.07 -13.68 24.93 36.32   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 21.06843 0.05466 385.4 <2e-16 \*\*\*  
## morekids -5.38700 0.08861 -60.8 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 21.71 on 254652 degrees of freedom  
## Multiple R-squared: 0.01431, Adjusted R-squared: 0.0143   
## F-statistic: 3696 on 1 and 254652 DF, p-value: < 2.2e-16

This rather simple linear model predicts that women with more than two children tend to work less than women who do not, specifically they work about 5.4 weeks less.

However the issue with this model is that it does not take into account women who have less than two children. This is an omitted variable bias error because such women have an effect on what we want to study and they are not involved. Also the R-squared values

Let’s examine the variable “samesex”

kid\_ivreg = lm(morekids ~ samesex, data = fertility)  
summary(kid\_ivreg)

##   
## Call:  
## lm(formula = morekids ~ samesex, data = fertility)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.4139 -0.4139 -0.3464 0.5860 0.6536   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.346425 0.001365 253.79 <2e-16 \*\*\*  
## samesex 0.067525 0.001920 35.17 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4844 on 254652 degrees of freedom  
## Multiple R-squared: 0.004835, Adjusted R-squared: 0.004831   
## F-statistic: 1237 on 1 and 254652 DF, p-value: < 2.2e-16

It appears that couples whose first two children are of the same sex are more likely to have a third child.

We may be able to make use of this as an instrument. But first we need to test for its relevance and its strength.

cor(fertility$morekids, fertility$samesex)

## [1] 0.06953403

The variable “samesex” has a nonzero correlation with “morekids” making it relevant.

reg2 = glm(morekids~ samesex, data= fertility)  
reg2.u = resid(reg2)  
cor(reg2.u, fertility$samesex)

## [1] 2.669621e-13

While the correlation with our instrument and u is not literally zero, it is close enough to say that it’s exogenous.

This is a valid instrument.

reg3 = lm(morekids~samesex,data=fertility)  
summary(reg3)

##   
## Call:  
## lm(formula = morekids ~ samesex, data = fertility)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.4139 -0.4139 -0.3464 0.5860 0.6536   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.346425 0.001365 253.79 <2e-16 \*\*\*  
## samesex 0.067525 0.001920 35.17 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4844 on 254652 degrees of freedom  
## Multiple R-squared: 0.004835, Adjusted R-squared: 0.004831   
## F-statistic: 1237 on 1 and 254652 DF, p-value: < 2.2e-16

Our instrument is “strong” because its F-statistics is 1237 which is > 10.

tsls1 = ivreg(weeksm1~morekids|samesex, data=fertility)  
summary(tsls1)

##   
## Call:  
## ivreg(formula = weeksm1 ~ morekids | samesex, data = fertility)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -21.42 -21.42 -13.42 24.89 36.89   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 21.421 0.487 43.988 < 2e-16 \*\*\*  
## morekids -6.314 1.275 -4.953 7.3e-07 \*\*\*  
##   
## Diagnostic tests:  
## df1 df2 statistic p-value   
## Weak instruments 1 254652 1237.219 <2e-16 \*\*\*  
## Wu-Hausman 1 254651 0.531 0.466   
## Sargan 0 NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 21.71 on 254652 degrees of freedom  
## Multiple R-Squared: 0.01388, Adjusted R-squared: 0.01388   
## Wald test: 24.54 on 1 and 254652 DF, p-value: 7.296e-07

The instrument seems to have a pretty noticeable effect as the model says that increasing fertility by a unit decreases the weeks worked by 6.314 weeks.

reg4 = lm(weeksm1~morekids+agem1+black+hispan+othrace, data=fertility)  
summary(reg4)

##   
## Call:  
## lm(formula = weeksm1 ~ morekids + agem1 + black + hispan + othrace,   
## data = fertility)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -36.62 -17.79 -10.72 22.86 45.47   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -4.83451 0.38540 -12.544 < 2e-16 \*\*\*  
## morekids -6.23042 0.08813 -70.696 < 2e-16 \*\*\*  
## agem1 0.83788 0.01262 66.389 < 2e-16 \*\*\*  
## black 11.66424 0.19217 60.697 < 2e-16 \*\*\*  
## hispan 0.46609 0.17937 2.599 0.00936 \*\*   
## othrace 2.14213 0.20304 10.550 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 21.38 on 254648 degrees of freedom  
## Multiple R-squared: 0.04376, Adjusted R-squared: 0.04374   
## F-statistic: 2331 on 5 and 254648 DF, p-value: < 2.2e-16

tsls2<- ivreg(weeksm1~morekids+agem1+black+hispan+othrace | samesex+agem1+black+hispan+othrace, data=fertility)  
summary(tsls2)

##   
## Call:  
## ivreg(formula = weeksm1 ~ morekids + agem1 + black + hispan +   
## othrace | samesex + agem1 + black + hispan + othrace, data = fertility)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -36.34 -17.66 -10.99 22.72 45.15   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -4.79189 0.40657 -11.786 <2e-16 \*\*\*  
## morekids -5.82105 1.24631 -4.671 3e-06 \*\*\*  
## agem1 0.83160 0.02289 36.336 <2e-16 \*\*\*  
## black 11.62327 0.22893 50.772 <2e-16 \*\*\*  
## hispan 0.40418 0.25986 1.555 0.12   
## othrace 2.13096 0.20586 10.352 <2e-16 \*\*\*  
##   
## Diagnostic tests:  
## df1 df2 statistic p-value   
## Weak instruments 1 254648 1279.811 <2e-16 \*\*\*  
## Wu-Hausman 1 254647 0.108 0.742   
## Sargan 0 NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 21.38 on 254648 degrees of freedom  
## Multiple R-Squared: 0.04368, Adjusted R-squared: 0.04366   
## Wald test: 1335 on 5 and 254648 DF, p-value: < 2.2e-16

The effect fertility has on the amount of weeks worked increases from -6.31 to -5.82 when we add these variables. These results seem very much more credible since these variables are significant for the most part.