

CA3bVdVeen

Ties van der Veen

15-9-2019

Rebecca Costello - Ties van der Veen - Niels van Opstal

I. Introduction to the assignment

```
library(foreign)
library(tidyverse)
library(ggdag)
library(dplyr)
library(tinytex)
library(jtools)
library(huxtable)
library(summarytools)
library(ggstance)
library(pwr)
library(knitr)
library(lemon)
library(AER)
```

```
knit_print.data.frame <- lemon_print
```

```
theUrl_ca3b <- "https://surfdrive.surf.nl/files/index.php/s/SCE35YQGR4hX6fw/download"
children <- read.dta (file = theUrl_ca3b)
```

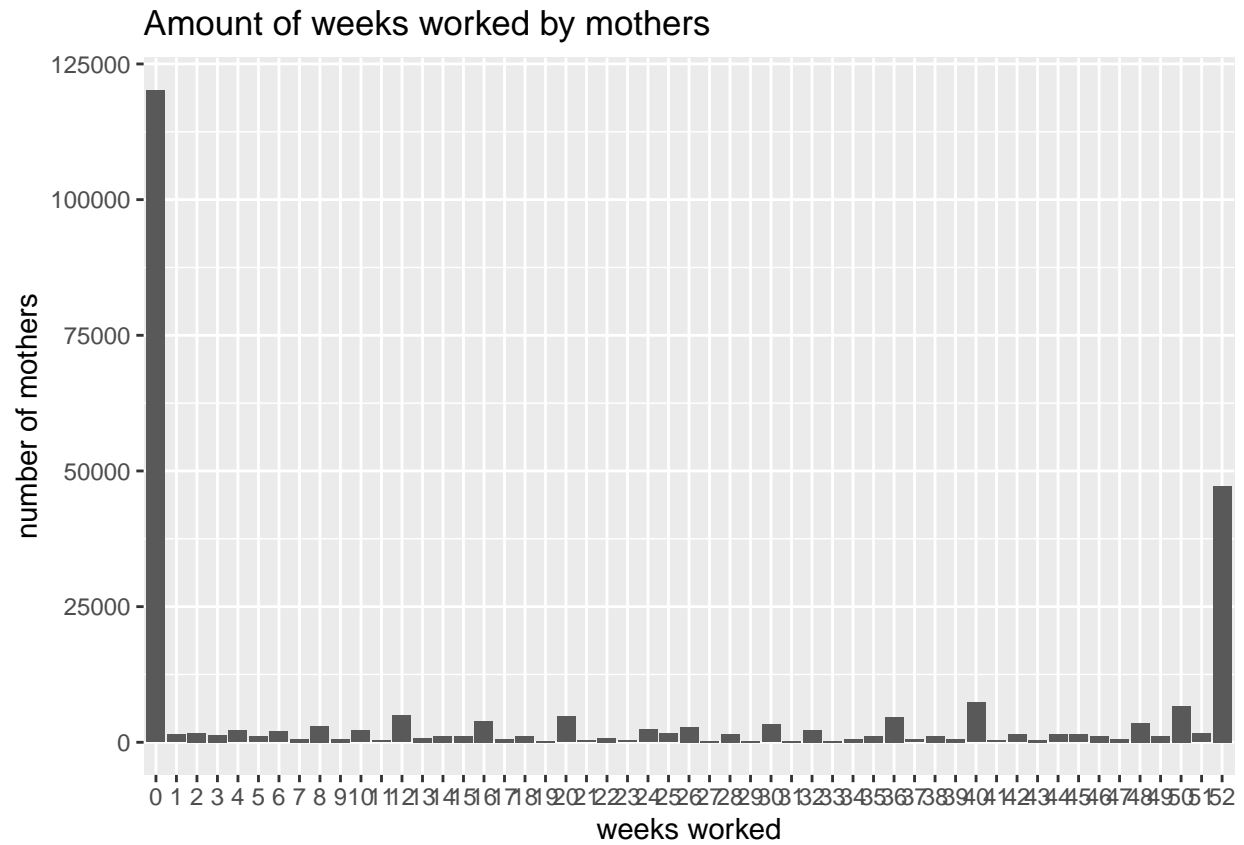
II. What to submit

- (1) $Y(0,i)$ = labour supply is not reduced (as much) $Y(1,i)$ = labour supply is reduced A reason for this additional reduction of labour supply could be that there are more kids to care for by either parent, which is generally a role that falls on the mother. This leads to having less time to spend on her career. Other reasons include physical discomfort of the mother caused by multiple birthings, discrimination by employers and non-change in the behaviour of the husband.

- (2) (a)

```
ggplot(children, aes(x=as.factor(weeksm1)))+
  geom_histogram(stat='count', binwidth = "1.0")+
  labs(x='weeks worked', y='number of mothers', title='Amount of weeks worked by mothers')
```

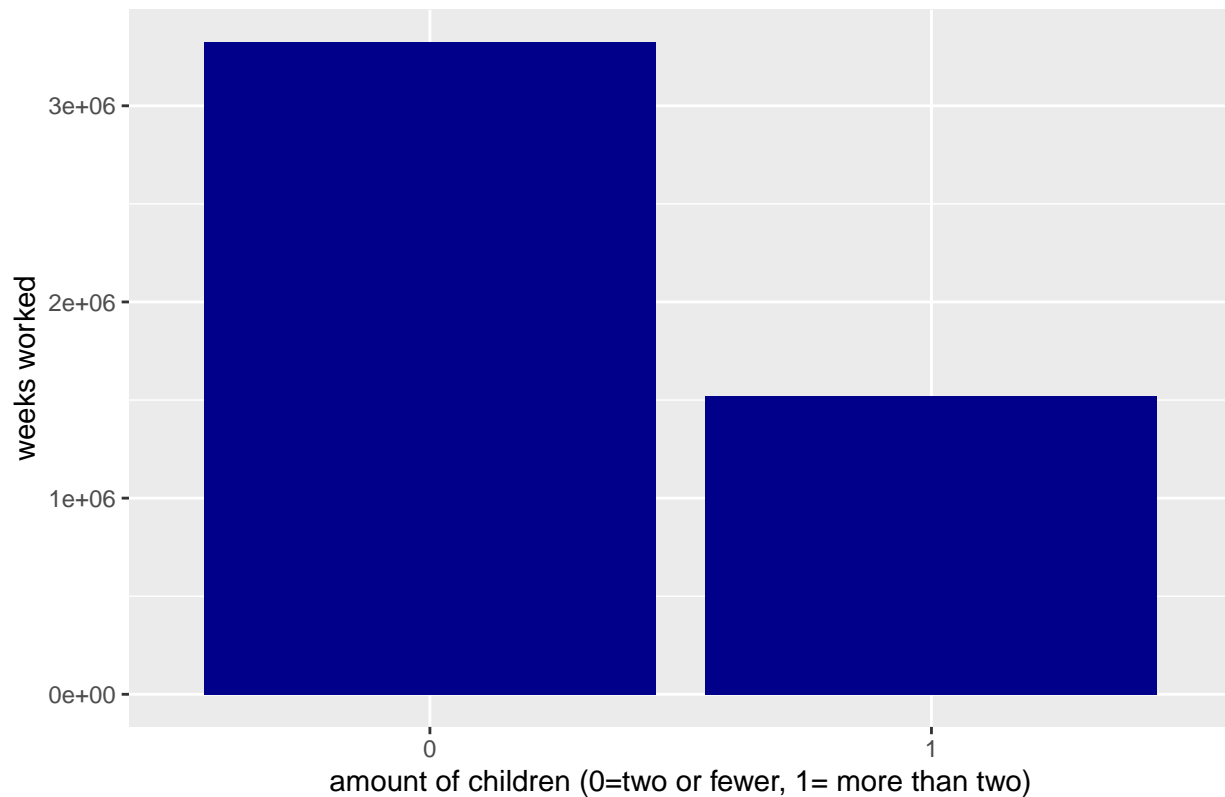
```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



(b)

```
ggplot(children, aes(y=weeksm1, x=as.factor(morekids))) + geom_bar(stat="identity", fill='darkblue') +
  labs(x='amount of children (0=two or fewer, 1= more than two)',
       y='weeks worked', title='Weeks worked by mothers with a certain number of kids')
```

Weeks worked by mothers with a certain number of kids



(3)

```
reg3b1 <- lm(weeksm1 ~ morekids, data=children)
summ(reg3b1, confint=TRUE)
```

```
## MODEL INFO:
## Observations: 254654
## Dependent Variable: weeksm1
## Type: OLS linear regression
##
## MODEL FIT:
## F(1,254652) = 3696.02, p = 0.00
## R2 = 0.01
## Adj. R2 = 0.01
##
## Standard errors: OLS
## -----
##               Est.    2.5%   97.5%   t val.    p
## -----
## (Intercept)    21.07   20.96   21.18   385.42   0.00
## morekids       -5.39   -5.56   -5.21  -60.79   0.00
## -----
```

These regression results tell us that having more than two kids would reduce the weeks a mother works by more than 5 on average.

(4)

The regression above is inappropriate because it fails to tell the whole story. This regression assumes that the one and only reason that women reduce the number of weeks they work is that they have more than two kids. Since there are obviously more reasons to reduce your work hours, it is clear that we are missing certain factors that also have an effect on this outcome. The effects we are missing here are so-called “omitted variables”. We need to find a way to address these omitted variables to get a more accurate estimate of this effect.

(5)

```
reg3b2 <- lm(morekids ~ samesex, data=children)
summ(reg3b2, confint=TRUE)
```

```
## MODEL INFO:
## Observations: 254654
## Dependent Variable: morekids
## Type: OLS linear regression
##
## MODEL FIT:
## F(1,254652) = 1237.22, p = 0.00
## R2 = 0.00
## Adj. R2 = 0.00
##
## Standard errors: OLS
## -----
##               Est.    2.5%   97.5%   t val.    p
## -----
## (Intercept)    0.35    0.34    0.35   253.79    0.00
## samesex        0.07    0.06    0.07    35.17    0.00
## -----
```

The regression results show a significant effect, yet the effect itself is very small. So parents whose first two kids were the same sex are more likely to have a 3rd (or more), but only by a small margin.

(6)

For a variable to be valid to serve as an instrumental variable for a certain measurement, it needs to fulfill both the independence assumption and the exclusion restriction. The independence assumption requires that the IV and the dependent variable do not have a common cause. In this case, the sex of the first two children and the amount of weeks a mother decides to work are not influenced by the same factor. The exclusion restriction requires that the IV does not have an effect on the dependent variable other than through the independent variable. In this case, the sex of the first two children should not have an effect on the hours worked by the mother, other than through the number of children she has.

(7)

```
reg3b3 <- ivreg(weeksm1 ~ morekids | samesex, data=children)
summary(reg3b3, confint=TRUE)
```

```
##
## Call:
## ivreg(formula = weeksm1 ~ morekids | same-sex, data = children)
##
## 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 ***
## ---
## 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
```

In this regression we see the same direction of the effect as seen earlier, but the size of the effect is considerably stronger. Thus, having more children still reduces the amount of labour the mother supplies, but even more than assumed before.

(8)

Never-takers and always-takers are difficult to theorize in this relationship. Always-takers are people who are assigned to the control group, but take the treatment. In this case that would mean they underreported the number of children they have. Inversely we have the never-takers, who should take the treatment (have more than 2 children) but don't, so they overreported the number of children they have. Both of these instances seem weird, but could have happened due to simple mistakes. It's unlikely that people actually forgot how many children they have, so it seems unlikely to me that never-takers or always-takers would significantly change the results.

(9)

```
reg3b4 <- ivreg(weeksm1 ~ morekids + agem1 + black + hispan + othrace |
               agem1 + black + hispan + othrace + same-sex, data=children)
summary(reg3b4, confint=TRUE)

##
## Call:
## ivreg(formula = weeksm1 ~ morekids + agem1 + black + hispan +
##      othrace | agem1 + black + hispan + othrace + same-sex, data = children)
##
## 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 ***
## ---
## 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
```

Here we see that the included covariates and IV reduce the effect a bit from the previous regression we did. It is now close to the very first estimate. We also see that the age of the mother, the mother being black and the mother being non-black, non-hispanic or non-white are significant factors. Apparently these factors alleviate the effect of having more kids on the mother's working hours somewhat.