Sofia Guo PSET 6 Econ 142

Sofia Guo

3/8/2019

1(a). Estimate first stage of causal model:

We want to estimate the first stage model for the probability of working full-time in months 15, 20, 14, and 48 using treatment T_i as the instrumental variable. The model given is:

$$FT_i = \pi_0 + \pi_1 T_i + \eta_i$$

```r
#load packages
library(dplyr)
library(magrittr)
library(stargazer)
library(ggplot2)
library(AER)
```

```r
#load data
welfare <- read.csv('/Users/sofia/Box/Cal (sofiaguo@berkeley.edu)/2018-19/Spring 2019/Econ 142/PSETS/PSE

#estimate first stage models
ft_15_reg <- lm(ft15 ~ treatment, data = welfare)
ft_20_reg <- lm(ft20 ~ treatment, data = welfare)
ft_24_reg <- lm(ft24 ~ treatment, data = welfare)
ft_48_reg <- lm(ft48 ~ treatment, data = welfare)
```

```r
stargazer(ft_15_reg,
          ft_20_reg,
          ft_24_reg,
          ft_48_reg, type = "latex", title = "First stage least squares model estimates",
          header = F,
          font.size = "small",
          multicolumn = F,
          column.sep.width = '0.1pt',
          single.row = T,
          omit.stat = c("f", "ser"),
          table.placement = "H")
```

Table 1: First stage least squares model estimates

	Dependent variable:			
	ft15	ft20	ft24	ft48
	(1)	(2)	(3)	(4)
treatment	0.137^{***} (0.011)	0.115^{***} (0.011)	0.107^{***} (0.011)	0.051^{***} (0.013)
Constant	0.147^{***} (0.008)	0.161^{***} (0.008)	0.160^{***} (0.008)	0.234^{***} (0.009)
Observations	5,480	5,245	5,224	4,796
R^2	0.028	0.019	0.017	0.003
Adjusted R^2	0.028	0.019	0.017	0.003
Note:				*p<0.1; **p<0.05; ***p<0.01

1(b). Estimate reduced form models

The reduced form model is:

$$y_i = \gamma_0 + \gamma_1 T_i + v_i$$

```
#estimate reduced form
red_15_reg <- lm(welfare15 ~ treatment, data = welfare)
red_20_reg <- lm(welfare20 ~ treatment, data = welfare)
red_24_reg <- lm(welfare24 ~ treatment, data = welfare)
red_48_reg <- lm(welfare48 ~ treatment, data = welfare)
```

```
stargazer(red_15_reg,
          red_20_reg,
          red_24_reg,
          red_48_reg, type = "latex", title = "Reduced form model estimates",
          header = F,
          font.size = "small",
          multicolumn = F,
          column.sep.width = '0.1pt',
          single.row = T,
          omit.stat = c("f", "ser"),
          table.placement = "H")
```

Table 2: Reduced form model estimates

	Dependent variable:			
	welfare15	welfare20	welfare24	welfare48
	(1)	(2)	(3)	(4)
treatment	-0.141^{***} (0.012)	-0.116^{***} (0.012)	-0.104^{***} (0.012)	-0.033^{**} (0.013)
Constant	0.810^{***} (0.008)	0.769^{***} (0.009)	0.739^{***} (0.009)	0.572^{***} (0.010)
Observations	5,480	5,480	5,480	5,480
R^2	0.026	0.016	0.013	0.001
Adjusted R^2	0.026	0.016	0.012	0.001
Note:				*p<0.1; **p<0.05; ***p<0.01

1(c). Estimate causal model by OLS for each month

We now run an OLS regression with this model:

$$y_i = \beta_0 + \beta_1 FT_i + u_i$$

```r
#run OLS
ols_15 <- lm(welfare15 ~ ft15, data = welfare)
ols_20 <- lm(welfare20 ~ ft20, data = welfare)
ols_24 <- lm(welfare24 ~ ft24, data = welfare)
ols_48 <- lm(welfare48 ~ ft48, data = welfare)
```

```r
stargazer(ols_15,
          ols_20,
          ols_24,
          ols_48, type = "latex", title = "Causal model OLS estimates",
          header = F,
          font.size = "small",
          multicolumn = F,
          column.sep.width = '0.1pt',
          single.row = T,
          omit.stat = c("f", "ser"),
          table.placement = "H")
```

Table 3: Causal model OLS estimates

	Dependent variable:			
	welfare15	welfare20	welfare24	welfare48
	(1)	(2)	(3)	(4)
ft15	-0.567^{***} (0.012)			
ft20		-0.570^{***} (0.013)		
ft24			-0.537^{***} (0.014)	
ft48				-0.534^{***} (0.014)
Constant	0.861^{***} (0.006)	0.833^{***} (0.006)	0.801^{***} (0.006)	0.709^{***} (0.007)
Observations	5,480	5,245	5,224	4,796
R^2	0.283	0.268	0.225	0.224
Adjusted R^2	0.282	0.268	0.225	0.224
Note:				*p<0.1; **p<0.05; ***p<0.01

1(d). Estimate causal model (1) by IV in each month

```r
iv_15 <- ivreg(welfare15 ~ ft15 | treatment, data = welfare)
iv_20 <- ivreg(welfare20 ~ ft20 | treatment, data = welfare)
iv_24 <- ivreg(welfare24 ~ ft24 | treatment, data = welfare)
iv_48 <- ivreg(welfare48 ~ ft48 | treatment, data = welfare)
```

```r
stargazer(iv_15,
          iv_20,
          iv_24,
          iv_48, type = "latex", title = "Causal model IV estimates",
```

```
            header = F,
            font.size = "small",
            multicolumn = F,
            column.sep.width = '0.1pt',
            single.row = T,
            omit.stat = c("f", "ser"),
            table.placement = "H")
```

Table 4: Causal model IV estimates

	Dependent variable:			
	welfare15	welfare20	welfare24	welfare48
	(1)	(2)	(3)	(4)
ft15	-1.027^{***} (0.082)			
ft20		-1.055^{***} (0.105)		
ft24			-1.008^{***} (0.117)	
ft48				-0.830^{***} (0.260)
Constant	0.961^{***} (0.019)	0.939^{***} (0.024)	0.902^{***} (0.026)	0.786^{***} (0.068)
Observations	5,480	5,245	5,224	4,796
R^2	0.097	0.074	0.052	0.156
Adjusted R^2	0.097	0.074	0.052	0.155
Note:				*p<0.1; **p<0.05; ***p<0.01

1(e). Verify the beta

Verify that in each month, $\hat{\beta}_1^{IV} = \hat{\gamma}_1/\hat{\pi}_1$.

```
#month 15
red_15_reg$coefficients[2]/ft_15_reg$coefficients[2]
```

```
## treatment
## -1.026628
```

```
#month 20
red_20_reg$coefficients[2]/ft_20_reg$coefficients[2]
```

```
## treatment
## -1.009946
```

```
#month 24
red_24_reg$coefficients[2]/ft_24_reg$coefficients[2]
```

```
##  treatment
## -0.9695207
```

```
#month 48
red_48_reg$coefficients[2]/ft_48_reg$coefficients[2]
```

```
##  treatment
## -0.6597087
```

Althought the $\hat{\beta}_1^{IV}$ for 15 months corresponds to the IV estimate in d), the remaining three do not correspond to the coefficients. This is because there are missing values in the different categories involved when estimating the first and second stage separately:

```r
#check the dependent var for the first stage
sum(is.na(welfare$ft15)) #no missing values = why the 15 months corresponds to the IV estimate
```

```
## [1] 0
```

```r
sum(is.na(welfare$ft20))
```

```
## [1] 235
```

```r
sum(is.na(welfare$ft24))
```

```
## [1] 256
```

```r
sum(is.na(welfare$ft48))
```

```
## [1] 684
```

```r
#check the dependent var for the reduced from
sum(is.na(welfare$welfare15))
```

```
## [1] 0
```

```r
sum(is.na(welfare$welfare20))
```

```
## [1] 0
```

```r
sum(is.na(welfare$welfare24))
```

```
## [1] 0
```

```r
sum(is.na(welfare$welfare48))
```

```
## [1] 0
```

```r
#check the treatment variable
sum(is.na(welfare$treatment))
```

```
## [1] 0
```

Thus, there is a discrepancy between the $\hat{\gamma}_1$ and $\hat{\pi}_1$ that is caused by essentially different sample sizes between the first stage and reduced form regressions. Using the *ivreg* function allows for this discrepancy to be accounted for (so that there's no difference in the sample sizes) and also calculates the correct standard errors.

2. Conduct steps (a) - (e) for month 15 only with controls

We now include controls so that our causal model is:

$$\begin{aligned} y_i = \beta_0 + \beta_1 FT_i + \beta_2 imm_i + \beta 3 hsgrad_i + \beta_4 agelt25_i + \\ \beta_5 age35p_i + \beta_6 baseline_i + \beta_7 kid6_i + \beta_8 nvmarried_i + u_i \end{aligned}$$

2(a). Estimate the first stage for month 15

```r
first <- lm(ft15 ~ treatment + imm + hsgrad + agelt25 + age35p + working_at_baseline + anykidsu6 + neve
```

```r
stargazer(first,
          type = "latex", title = "First state
          w/controls estimates",
```

```r
          header = F,
          font.size = "small",
          multicolumn = F,
          column.sep.width = '0.1pt',
          single.row = T,
          table.placement = "H")
```

Table 5: First state w/controls estimates

	Dependent variable:
	ft15
treatment	0.138*** (0.010)
imm	−0.052*** (0.016)
hsgrad	0.095*** (0.011)
agelt25	0.016 (0.016)
age35p	−0.044*** (0.014)
working_at_baseline	0.264*** (0.014)
anykidsu6	−0.013 (0.013)
nevermarried	−0.014 (0.012)
Constant	0.086*** (0.014)
Observations	5,480
R^2	0.114
Adjusted R^2	0.113
Residual Std. Error	0.388 (df = 5471)
F Statistic	88.250*** (df = 8; 5471)
Note:	*p<0.1; **p<0.05; ***p<0.01

2(b). Estimate the reduced model for month 15

```r
reduced <- lm(welfare15 ~ treatment + imm + hsgrad + agelt25 + age35p + working_at_baseline + anykidsu6
```

```r
stargazer(reduced,
          type = "latex", title = "Reduced form
          w/controls estimates",
          header = F,
          font.size = "small",
          multicolumn = F,
          column.sep.width = '0.1pt',
          single.row = T,
          table.placement = "H")
```

Table 6: Reduced form w/controls estimates

	Dependent variable:
	welfare15
treatment	-0.141^{***} (0.011)
imm	0.028 (0.017)
hsgrad	-0.101^{***} (0.012)
agelt25	-0.033^{*} (0.017)
age35p	0.058^{***} (0.015)
working_at_baseline	-0.221^{***} (0.015)
anykidsu6	0.010 (0.014)
nevermarried	0.042^{***} (0.013)
Constant	0.855^{***} (0.015)
Observations	5,480
R^2	0.086
Adjusted R^2	0.085
Residual Std. Error	0.420 (df = 5471)
F Statistic	64.682^{***} (df = 8; 5471)
Note:	*p<0.1; **p<0.05; ***p<0.01

2(c). Estimate the causal model by OLS for month 15

```
causal <- lm(welfare15 ~ ft15 + imm + hsgrad + agelt25 + age35p + working_at_baseline + anykidsu6 + neve
```

```
stargazer(causal,
          type = "latex", title = "Causal model
          w/controls estimates",
          header = F,
          font.size = "small",
          multicolumn = F,
          column.sep.width = '0.1pt',
          single.row = T,
          table.placement = "H")
```

Table 7: Causal model w/controls estimates

	Dependent variable:
	welfare15
ft15	-0.538^{***} (0.013)
imm	0.0005 (0.015)
hsgrad	-0.051^{***} (0.010)
agelt25	-0.024 (0.015)
age35p	0.034^{***} (0.013)
working_at_baseline	-0.077^{***} (0.013)
anykidsu6	0.005 (0.012)
nevermarried	0.034^{***} (0.011)
Constant	0.867^{***} (0.013)
Observations	5,480
R^2	0.293
Adjusted R^2	0.292
Residual Std. Error	0.370 (df = 5471)
F Statistic	282.834^{***} (df = 8; 5471)
Note:	*p<0.1; **p<0.05; ***p<0.01

2(d). Estimate the causal model by IV for month 15

```
causal_iv <- ivreg(welfare15 ~ ft15 + imm + hsgrad + agelt25 + age35p + working_at_baseline + anykidsu6
```

```
stargazer(causal_iv,
          type = "latex", title = "Causal model by IV
          w/controls estimates",
          header = F,
          font.size = "small",
          multicolumn = F,
          column.sep.width = '0.1pt',
          single.row = T,
          table.placement = "H")
```

Table 8: Causal model by IV w/controls estimates

	Dependent variable:
	welfare15
ft15	-1.026^{***} (0.082)
imm	-0.025 (0.017)
hsgrad	-0.003 (0.014)
agelt25	-0.016 (0.017)
age35p	0.012 (0.015)
working_at_baseline	0.051^{*} (0.026)
anykidsu6	-0.003 (0.014)
nevermarried	0.028^{**} (0.013)
Constant	0.943^{***} (0.019)
Observations	5,480
R^2	0.101
Adjusted R^2	0.100
Residual Std. Error	0.417 (df = 5471)
Note:	*p<0.1; **p<0.05; ***p<0.01

2(e). Verify the beta for month 15 estimates

Verify that for month 15, $\hat{\beta}_1^{IV} = \hat{\gamma}_1/\hat{\pi}_1$.

```
#month 15
reduced$coefficients[2]/first$coefficients[2]
```

```
## treatment
## -1.025948
```

My estimates verify that $\hat{\beta}_1^{IV} = -1.026 = \hat{\gamma}_1/\hat{\pi}_1 = -1.0259$.

The addition of the controls does not seem to drastically affect my estimates for month 15 (decrease by approximately 0.001). This can tell us that the no-controls causal model was overestimating the effect of the treatment by 0.001 compared to the controlled causal model, but given such a small difference it seems that using the controls is not vital to obtaining similar results.