Impact of social demographics on household earnings in the US

Name: Imisi Raphael Aiyetan Course: Econometrics 512

Note: Check the appendix for the detailed results and codes Final Examination

The dataset contains the following variables for a sample of individuals within the age of 37-45:

- Treat-1 if Year of Schooling is 15, 0 otherwise- This is the **treatment variable.** It reflects the number of years the individual spent in the school. The year falls between 12 and 19 inclusive.
- 1 if female, 0 if male. **Covariate**
- Ethnicity- This also distinguishes between black and Hispanic race. 1 if black, 0 otherwise and 1 If Hispanic, 0 otherwise. **Covariate**
- Total out-of-school work experience. Covariate
- Married-1 if married, 0 otherwise. **Covariate**
- Ssf-1 if father schooling is 15, 0 otherwise. **Instrument** variables
- Ssm-1 if mother schooling 15, 0 otherwise. **Instrument** variables
- Sib-1 if siblings are 3, 0 otherwise. **Instrument** variables
- Earnings- the wage of the individual. This is the main **dependent variable** in this analysis Data source: Earnings function of US for 2000.

Link to dataset: https://www.macmillanihe.com/companion/gujarati-econometrics-by-example-2e/learning-resources/Data-sets/

Question 1:

 $\ln earnings_i = \beta_0 + \beta_1 treatment_i + \beta_2 w \exp_i + \beta_3 Female_i + \beta_4 ethblack_i + \beta_5 ethhisp_i + \beta_6 married_i + u_i$ Where:

Lnearnings: the wage of the individual.

Treatment: Treat-1 if Year of Schooling is 15, 0 otherwise- This is the **treatment variable**. It reflects the number of years the individual spent in the school.

Wexp: Total out-of-school work experience. Covariate

Female: 1 if female, 0 if male. Covariate Ethblack: 1 if black, 0 otherwise. Covariate Ethhisp: 1 if Hispanic, 0 otherwise. Covariate Married: 1 if married, 0 otherwise. Covariate

SSf: 1 if father schooling is 15, 0 otherwise. **Instrument** variable **SSm**: if mother schooling 15, 0 otherwise. **Instrument** variable

Sib: 1 if siblings are 3, 0 otherwise. **Instrument** variable

Code:

clear all

use "/Users/imisiaiyetan/Downloads/Table19_4_data.dta"

* To generate the treatment variable

gen treatment = 0

replace treatment = 1 if s \geq =15

* To generate the instrumental variables

gen ssf = 0

replace ssf = 1 if sf > = 15

gen ssm = 0

replace ssm = 1 if sm > = 15

gen sib = 1

replace sib = 0 if siblings > 3

Question 2:

This study considers Instrumental Variables (IV) analysis to estimate the effect of the treatment defined in question 1 on the outcome. Before proceeding to IV regression, the study starts by estimating the effect of treatment on outcome using Ordinary Least Square (OLS).

Table 1: OLS without robust

- . * Question 2
- . * Befor jumping into iv-regression, we estimate the effect of treatment
- . *and other covariates on earnings
- . reg earnings treatment wexp female ethblack ethhisp married

Source	SS	df	MS		Number of obs	= 540
					F(6, 533)	= 30.09
Model	30490.8343	6 50	81.80572		Prob > F	= 0.0000
Residual	90027.5936	533 16	8.907305		R-squared	= 0.2530
					Adj R-squared	= 0.2446
Total	120518.428	539 22	3.596341		Root MSE	= 12.996
earnings	Coef.	Std. Err	. t	P> t	[95% Conf.	Interval]
						
treatment	13.23761	1.26289	10.48	0.000	10.75676	15.71846
wexp	.4819368	.1269443	3.80	0.000	.2325642	.7313094
female	-6.968988	1.149743	-6.06	0.000	-9.227572	-4.710404
ethblack	-3.88289	1.841426	-2.11	0.035	-7.500233	2655468
ethhisp	-4.003112	2.553102	-1.57	0.117	-9.018489	1.012265
married	.6896808	1.238592	0.56	0.578	-1.74344	3.122801
_cons	11.19942	2.67259	4.19	0.000	5.949315	16.44952
-						

Table 2: OLS with robust

- . *Robust ols
- . reg earnings treatment wexp female ethblack ethhisp married, robust

Linear regression	Number of obs =	540
,	F(6, 533) =	17.60
	Prob > F =	0.0000
	R-squared =	0.2530
	Root MSE =	12.996

earnings	Coef.	Robust Std. Err.	t	P> t	[95% Conf.	Interval]
treatment	13.23761	1.493037	8.87	0.000	10.30465	16.17057
wexp	.4819368	.1163086	4.14	0.000	.2534573	.7104162
female	-6.968988	1.107371	-6.29	0.000	-9.144335	-4.793641
ethblack	-3.88289	1.374433	-2.83	0.005	-6.582861	-1.182919
ethhisp	-4.003112	1.375958	-2.91	0.004	-6.706078	-1.300145
married	.6896808	1.213343	0.57	0.570	-1.69384	3.073202
_cons	11.19942	2.231041	5.02	0.000	6.816705	15.58213

It is evident from the two results that there isn't different between OLS with and OLS without robust. However, if we assumed that the treated variable is not randomly assigned, then the treatment and the error term are correlated (i.e. there is unobservables that correlate with the treatment). Therefore, to check if this is true, the study proceeds with IV regression analysis.

Question 3:

The study starts by estimating the IV-regression using Two Stages Least Square (TSLS) and alternatively estimate IV-regression using Generalized Method of Moment (GMM).

Table 3: IV-Regression (2SLS)

- . ivregress 2sls earnings (treatment= ssf ssm sib) wexp female ethblack ethhisp $\ensuremath{///}$
- > married

Instrumental variables (2SLS) regression

Number of obs = 540
Wald chi2(6) = 109.32
Prob > chi2 = 0.0000
R-squared = 0.1512
Root MSE = 13.763

earnings	Coef.	Std. Err.	Z	P> z	[95% Conf	. Interval]
treatment	23.99881	3.525716	6.81	0.000	17.08853	30.90908
wexp	.7779901	.1616396	4.81	0.000	.4611824	1.094798
female	-7.041232	1.217784	-5.78	0.000	-9.428045	-4.654419
ethblack	-1.630569	2.066161	-0.79	0.430	-5.680171	2.419032
ethhisp	-2.533138	2.740232	-0.92	0.355	-7.903893	2.837617
married	309908	1.346226	-0.23	0.818	-2.948462	2.328646
_cons	3.147162	3.737522	0.84	0.400	-4.178246	10.47257

Instrumented: treatment

Instruments: wexp female ethblack ethhisp married ssf ssm sib

Table 4: IV-Regression (GMM)

. ivregress gmm earnings (treatment= ssf ssm sib) wexp female $% \left(1\right) =0$ ethblack ethhisp /// > married

Instrumental variables (GMM) regression

Number of obs = 540 Wald chi2(6) = 77.19 Prob > chi2 = 0.0000 R-squared = 0.1541 Root MSE = 13.74

GMM weight matrix: Robust

earnings	Coef.	Robust Std. Err.	z	P> z	[95% Conf.	Interval]
treatment wexp female ethblack ethhisp married _cons	23.83851 .745146 -7.004812 -1.401401 -2.50877 0138035 3.439198	4.730701 .171683 1.137456 1.688912 1.886951 1.341536 3.749987	5.04 4.34 -6.16 -0.83 -1.33 -0.01 0.92	0.000 0.000 0.000 0.407 0.184 0.992	14.5665 .4086536 -9.234184 -4.711607 -6.207126 -2.643166 -3.910641	33.11051 1.081638 -4.77544 1.908806 1.189586 2.615559 10.78904

Instrumented: treatment

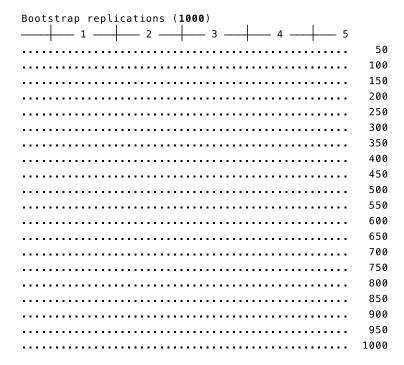
Instruments: wexp female ethblack ethhisp married ssf ssm sib

To estimate the standard errors, I considered bootstrap method as the most appropriate method given the IV regression. First of all, the study estimates bootstrap IV-regression before it proceeds to estimate the standard errors.

Table 5: Bootstrap IV-regression

- . ivregress 2sls earnings (treatment= ssf ssm sib) wexp female ethblack ethhisp ///
- > married, vce(bootstrap, reps(1000) seed(423))

(running ivregress on estimation sample)



Instrumental variables (2SLS) regression

Number of obs = 540 Wald chi2(6) = 70.60 Prob > chi2 = 0.0000 R-squared = 0.1512 Root MSE = 13.763

	Observed	Bootstrap			Normal-based			
earnings	Coef.	Std. Err.	z	P> z	[95% Conf.	. Interval]		
treatment	23.99881	4.864031	4.93	0.000	14.46548	33.53213		
wexp	.7779901	.1792944	4.34	0.000	.4265796	1.129401		
female	-7.041232	1.236768	-5.69	0.000	-9.465252	-4.617212		
ethblack	-1.630569	1.771078	-0.92	0.357	-5.101819	1.84068		
ethhisp	-2.533138	2.086988	-1.21	0.225	-6.62356	1.557284		
married	309908	1.46668	-0.21	0.833	-3.184547	2.564731		
_cons	3.147162	3.885418	0.81	0.418	-4.468119	10.76244		

Instrumented: treatment

Instruments: wexp female ethblack ethhisp married ssf ssm sib

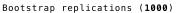
Table 6: Standard Error Estimation

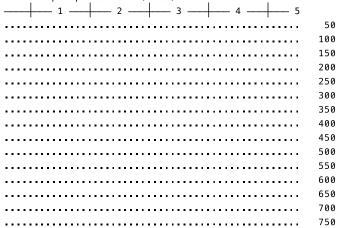
```
. bootstrap <code>_b[treatment] _se[treatment]</code>, <code>reps(1000) seed(423) ///</code>
```

> (treatment= ssf ssm sib) wexp female ethblack ethhisp married, robust

(running ivregress on estimation sample)

(note: file bootstrap_data.dta not found)





Bootstrap results

Number of obs = 540 Replications = 1000

_bs_1: _b[treatment]
_bs_2: _se[treatment]

	Observed Coef.	Bootstrap Std. Err.	Z	P> z		.—based Interval]
_bs_1	23.99881	4.864031	4.93	0.000	14.46548	33.53213
_bs_2	4.761793	.8396382	5.67	0.000	3.116132	6.407453

. use "/Users/imisiaiyetan/Downloads/bootstrap_data.dta", clear (bootstrap: ivregress)

. gen temp=(_bs_1+ 23.99881)^2

. egen temp2=mean(temp)

. gen se=sqrt(temp2)

. sum se

Variable	0bs	Mean	Std. Dev.	Min	Max
se	1000	48.06244	0	48.06244	48.06244

> saving(bootstrap_data, replace): ivregress 2sls earnings ///

The reasons why the standard errors generated by bootstraps approach are the correct standard errors

Answer:

Inference using conventional standard errors in 2SLS is based on an estimate of a moment that in finite samples often does not exist (as the coefficient has no finite variance when exactly identified). The bootstrap solved this problem by using resampling technique to estimate the percentiles of distributions, which always exist, and does much better. In that case, while asymptotic theory favors the resampling of the t-statistic, then avoiding finite sample 2SLS standard estimate altogether and focusing on the bootstrap resampling of the coefficient distribution alone provides the best performance, with tail rejection probabilities on IV coefficients that are very close to nominal size in iid, non-iid, low and high leverage settings.

Interpretation of the treatment effect estimate Answer:

In this study, the treatment effect, which is the average causal effect of individuals receiving 15 years free education on their earnings. The results as shown in Table 3 or Table 4 indicates that the treatment when account for endogeneity (e.g. siblings, father schooling and mother schooling) has a significant positive on earnings. Intuitively, the result shows that when an individual is incentivized to stay in school for 15 year or more than and he/she is from an educated background with not more than three siblings, individual earnings increase by 24 percent. In addition, the covariates (i.e. total work experience and female) are highly significant to explain earning behaviors. In terms of Average Treatment Estimation (ATE), the difference between the mean of the outcome for controls and the mean of the outcome for the treatment gives 30.3 (note: this value is calculated as the 24.7-0.7+7). In this case, it is plausible since I assumed the treatment is homogeneous and the whole population is affected by the instrumental variable.

Table 7 IV-regression bias

. estat firststage

First-stage regression summary statistics

Variable	R−sq.	Adjusted R−sq.	Partial R-sq.	F(3,531)	Prob > F
treatment	0.2335	0.2220	0.1439	29.7495	0.0000

Minimum eigenvalue statistic = 29.7495

Critical Values Ho: Instruments are weak	<pre># of endogenous regressors: # of excluded instruments:</pre>				
2SLS relative bias	5% 13.91	10% 9.08	20% 6.46	30% 5.39	

10% 15% 20% 25% 25LS Size of nominal 5% Wald test 22.30 12.83 9.54 7.80 LIML Size of nominal 5% Wald test 6.46 4.36 3.69 3.32