Name: Imisi Raphael Aiyetan

Course: Econometrics 512

Problem Set 4

Problem 1: Verify the measurement error result.

$$Y_i = b_0 + b_1 T_i + u_i \tag{1}$$

Where $u_i \sim N(0, s_u^{\,2})$

$$t_{1i} = T_i + x_i \tag{2}$$

Where $x_i \sim N(0, s_i^{\,2})$

$$t_{2i} = T_i + h_i \tag{3}$$

Where $h_i \sim N(0, s_x^{\,2})$

Table 1: Measurement Error Regression Estimation

$b_1(t_{1i})$=-0.03064272				$b_1(t_{2i})$=0.011488510		
	T_i		t_{1i}		t_{2i}	
Dependent: Y	Coefficient	P>|t|	Coefficient	P>|t|	Coefficient	P>|t|
b_1	-0.061285	0.566000	0.011489	0.706000	0.0346435	0.246000
Constant	9.998374	0.000000	9.962787	0.000000	9.9496940	0.000000
	N=1000 R-squared =0.0003 F-Stat=0.33(0.5660)		N=1000 R-squared=0.0001 F-Stat=0.14(0.7063)		N=1000 R-squared=0.0014 F-Stat=1.35(0.2455)	

Question (a): If you run a regression of Y_i on T_i you get something close to b_1

Interpretation:

The regression analysis shows that when Y_i is regressed on T_i, the treatment has an insignificant negative impact on Y_i. The implication of this result is that when the unobserved T_i is assumed to follow a uniform distribution, it is insignificant to the model as shown in Table 1. **Appendix 1** contains the **STATA output** for this question and also the **STATA codes** are shown in **Appendix 8**.

Question (b): If you run a regression of Y_i and t_{1i} you get something close to $b_1 \frac{Var(T_i)}{Var(T_i)+s_x^2}$

Interpretation:

The regression analysis showed that when Y_i is regressed on t_{1i} , t_{1i} has an insignificant positive effect on Y_i. Also, the value of the estimated parameter (0.011489) of this model is equal to $b_1 \frac{Var(T_i)}{Var(T_i)+s_x^2}$ = 0.01148851. The implication is that when the noisy version of T_i is observed the estimated parameters are equal, indicating the impact of i.i.d measurement error with variance s_x^2 on the model. **Appendix 2** contains the **STATA output** for this question and also the **STATA codes** are shown in **Appendix 8**.

Question (c): Doing IV using t_{2i} as an instrument for t_{1i} gives an estimate close to b_1

Interpretation:

Table 1 also shows the result, which answer this question. The result shows that when t_{2i} is used an instrument in this model, it has an insignificant positive impact on Y_i. The estimated value here (i.e.0.0346435) is close to $b_1(t_{1i})$=-0.03064272 in the absolute term. The implication is that when t_{2i} is correlated with t_{1i} but uncorrelated with error term the esitimated parameter is closed to b_1. **Appendix 3** contains the **STATA output** for this question and also the **STATA codes** are shown in **Appendix 8**.

Problem 2: How bootstrapping affect standard error estimation in an IV setting.

Table 2: Linear Regression and IV Regression Estimations with and without Bootstrapping

	OLS(N=711)		OLS-Bootstrap (Replications=1000)			IVR(N=2588)		IVR-Bootstrap (Replication=1000)	
	R-squared=0.0465		R-squared=0.0465			R-squared=0.0174		R-squared=0.0174	
	F-Stat=4.06(0.0002)		Waldchi2=20.21(0.0051)			Waldchi2=49.49(0.000)		Waldchi2=9.50(0.2187)	
Variables	Coefficients	Robust Std. Err.	Coefficients	Bootstrap Std.Err.	Variables	Coefficients	Robust Std. Err.	Coefficients	Bootstrap Std.Err.
Pct_insclnxtyr	-0.051748 (0.001)	0.015011	-0.051748 (0.004)	0.017789	ctuition17	0.825821 (0.000)	0.220123	0.825821 (0.099)	0.500323
Mhighgrad	1.705758 (0.135)	1.138679	1.705758 (0.207)	1.350351	mhighgrad	1.714111 (0.006)	0.620411	1.714111 (0.227)	1.418963
msomcol	1.654315 (0.185)	1.249697	1.654315 (0.257)	1.460228	msomcol	1.641105 (0.011)	0.647933	1.641105 (0.254)	1.439741
fhighgrad	2.047867 (0.082)	1.175058	2.047867 (0.135)	1.371592	fhighgrad	0.346595 (0.580)	0.626745	0.346595 (0.808)	1.429186
fsomcol	-0.694729 (0.586)	1.274121	-0.694729 (0.657)	1.565226	fsomcol	-1.267924 (0.064)	0.685348	-1.267924 (0.418)	1.564162
parincome	0.015601 (0.074)	0.008707	0.015601 (0.145)	0.010704	parincome	0.008082 (0.015)	0.003311	0.008082 (0.294)	0.007051
afqt	0.025462 (0.115)	0.016147	0.025462 (0.163)	0.018246	afqt	-0.02129 (0.0104)	0.008217	-0.021294 (0.246)	0.018361
Constant	4.478025 (0.001)	1.398370	4.478025 (0.007)	1.661462	constant	2.573612 (0.000)	0.602760	2.573612 (0.054)	1.333937

Table3: Mean Estimation

		N=1000		
Variables	Mean	Std.Err.	(95% Conf. Interval)	
Pct_insclnxtyr	-0.0524788	0.0005625	-0.0535827	-0.0513749
ctuition17	0.8145842	0.0158216	0.7835368	0.8456315

Question(a): Estimate the regression via OLS with robust standard errors (based on the usual asymptotic approximation accounting for generic heteroskedasticity). Report the 95% confidence interval for the variable "pctinsclnxtyr".

Interpretation:

Table 2 presents the result to this section. In assessing the relationship between the exogeneous variables in this model and the dependent variable, the result shows that Pct_insclnxtyr, fhighgrad, and parincome have different significant impact on dayssmklm17. In addition, the robust standard error result as shown in table also support our findings. However, other variables in this model are insignificant to impact dayssmklm17. Furthermore, as shown in Appendix, we are 95% confident that the estimated value of Pct_insclnxtyr fall between -0.0812201 and -0.022751. **Appendix 4** contains the **STATA output** for this question and also the **STATA codes** are shown in **Appendix 9**.

Question(b): Now estimate the standard errors by using the non-parametric bootstrap. This involves re-sampling observations with replacement (until you have N observations, where N is the numb er of observations in the regression), estimating the model, saving the coefficient, and then doing this many times over (you should use 1,000 replications in each of these exercises). You can then compute standard errors or do hypothesis testing based on the distribution of the coefficients. This is sometimes called the "bootstrap-c", where "c" stands for coefficient. If you use Stata, look at the "bootstrap" command. Remember to set a seed for replication purposes (since bootstrapping is based on a random process, you won't get the same answer every time unless you set the seed). Report the 95% confidence interval for the variable "pctinsclnxtyr".

Interpretation:

Table 2 also presents the result of the bootstrapping OLS. In assessing the relationship between the exogeneous variables in this model and the dependent variable, the result shows that while others variables are insignificant to the model, Pct_insclnxtyr, significantly impact on dayssmklm17. In addition, the robust standard error result as shown in table also support our findings that Pct_insclnxtyr is very important to the model. Furthermore, as shown in Appendix, we are 95% confident that the estimated value of Pct_insclnxtyr fall between -0.0866 and -0.01688. **Appendix 5** contains the **STATA output** for this question and also the **STATA codes** are shown in **Appendix 9**.

Question (c)

Now repeat parts (a) and (b) but with IV estimation, using "ctuition17" as an instrument for "pctinsclnxtyr".

Interpretation:

For the IV estimation with bootstrapping, Table 2 also presents the result. The result shows that ctuition17, mhighgrad, msomcol, fsomcol, parincome, and afqt significantly have impact on dayssmklm17. The implication of this result is that IV estimation is very robust for these data. Furthermore, as shown in Appendix, we are 95% confident that the estimated value of ctuition17, fall between -0.3944 and 1.2573. Similarly, Table 2 also presents the result of the bootstrapping IV estimation, the result shows that while other variables are insignificant to the model, ctuition17, significantly impact on dayssmklm17. In addition, the robust standard error result as shown in table also support our findings that ctuition17 is very important to the model. Furthermore, as shown in Appendix, we are 95% confident that the estimated value of ctuition17 fall between -0.1548 and 1.8064. **Appendix 6 and 7** contains the **STATA output** for this question and also the **STATA codes** are shown in **Appendix 9**.

Question (d)

A new and important paper by Alwyn Young (presented in class) suggests that bootstrapping may provide more accurate inference in IV settings than traditional asymptotic inference (even when adjusting for heteroskedasticity or correlated errors). Why is IV especially susceptible to these problems in finite samples (relative to OLS)?

Interpretations

The following explain the reasons

1. Non-iid error processes adversely affect the size and power of IV estimates, increasing the bias of IV relative to OLS.
2. 2SLS estimation is based a moment which does not exist in finite sample as the coefficient when identified does not have finite variance.

In conclusion, bootstrapping method provides solution to this problem.