PSet 8

Spencer Papay - ssp2170

4/18/2017

Question 1

Assumptions: $E(\mu|X) = 0, Z\epsilon[0,1], cov(Z,X) \neq 0, cov(Z,\mu) = 0$

Now:

$\overline{Y_1} = \frac{1}{n_1}\sum_{i=Z_i=1} Y_i$

$\overline{Y_0} = \frac{1}{n_0} = \sum_{i=Z_i=0} Y_i$

$\overline{X_i} = \frac{1}{n_1}\sum_{i=Z_i=1} X_i$

$\overline{X_0} = \frac{1}{n_0}\sum_{i=Z_i=0} X_i$

$\hat{\beta}_1 = \frac{\sum\left(Z-\overline{Z}\right)\left(Y_i-\overline{Y}\right)}{\sum\left(Z-\overline{Z}\right)\left(X_i-\overline{X}\right)} = \frac{\sum Z_i\left(Y_i-\overline{Y}\right)}{\sum Z_i\left(X_i-\overline{X}\right)}$

$\sum z_i\left(Y_i-\overline{Y}\right) = \sum z_iY_i - \sum Z_i\overline{Y} = n_1Y_i - n_i\overline{Y} = n_1\left(\overline{Y_i}-\overline{Y}\right)$

$\overline{Y} = \frac{1}{n}\sum Y_i = \frac{1}{n}\left(\sum_{i=z_i=1} Y_i \ + \sum_{i=z_i=0} Y_i\right) = \frac{1}{n}\left(n_1\overline{Y_i} + n_0\overline{Y_0}\right) = \frac{1}{n}\left(n_1\overline{Y_i} + \left(n-n_1\right)Y_0\right)$

$n_1\left(\overline{Y_i}-\overline{X}\right) = n_1\left(\frac{n-n_1}{n}Y_i \ - Y_0\right)$

$\sum Z_i\left(X_i-\overline{X}\right) = n_1\left(\overline{X_i}-\overline{X}\right) = n_1\left(\frac{n-n_1}{n}\ \overline{X_1} - X_0\right)$

$\hat{\beta}_1 = \frac{n_1\left(\frac{n-n_1}{n}\right)\left(\overline{Y_1}-\overline{Y_0}\right)}{n_1\left(\frac{n-n_1}{n}\right)\left(\overline{X_1}-\overline{X_0}\right)}$

$\boxed{= \frac{\overline{Y_1}-Y_0}{\overline{X_1}-X_0}}$

Question 2

a) $Y = \beta_0 + \beta_1X_1^* + \beta_2'X_2 + \mu$ X_1 is observed X_1^* $X_1 = X_1^* + \epsilon_1$, meaning uncorrelated with X_1^* and X_2. OLS regression of Y on a constant term X_1 and X_2 gives inconsistent esitmator with errors-in-variables bias.

$Y_i = \beta_0 + \beta_1X_1 + [\beta\ _1\left(X_1^*-X_1\right) + \mu_1'] + \beta_2'X_2 + \mu$

We run the regression: $Y_i = \beta_0 + \beta_1X_1 + \mu' + \beta_2'X_2 + \mu$

$\mu' = \beta_1\left(X_1^*-X_1\right) + \mu$

With measurement error, X_1 is correlated with μ' $\therefore$ β_1 is biased.

$cov\left(X_1, \mu\ '\right) = cov\left(X_1, \beta_1\left(X_1^*-X_1\right) + \mu\right)$

$= \beta_1 cov\left(X_1, X_1^*-X_1\right) + cov\left(X_1, \mu_i\right)$

$= \beta_1 cov\left(X_1, X_1^*-X_1\right)$ if $cov(X_1, \mu_i) = 0$

Here, $X_1 = X_1^* + \epsilon_1$, which indicates zero random noise when $corr(X_1^*, \epsilon_1) = 0$ and $corr(\mu_i, \epsilon_1) = 0$

$var(X_1^*) = \sigma_X^2 + \sigma_\epsilon^2$ $cov(X_1, X_1^*-X_1) = cov(X + \epsilon_\mu', -\epsilon_i) = -\sigma_\epsilon^2$

$\therefore cov(X_1, \mu') = -\beta_1 \sigma_v^2$

$$\hat{\beta}_1 \rightarrow^p \beta_1 - \beta_1 \frac{\sigma_v^2}{\sigma_{x_1}^2} = \left(1 - \frac{\sigma_v^2}{\sigma_{x_1}^2}\right)\beta_1 \;\; = \boxed{\left(\frac{\sigma_x^2}{\sigma_x^2 + \sigma_v^2}\right)\beta_1}$$

b) $\hat{\beta}_1$ is biased towards zero. An important assumption is $corr(X_1^*, \epsilon_1) = 0$.

c) $Z_1 = X_1^* + \epsilon_2$, remember that ϵ_2 is uncorrelated with $X_1^*, X_2, \mu, \epsilon_1$

As in the best guess model, $Z_1 = E(X_1^*|W_i)$ where $E(U|W_i) = 0$. $\therefore cov(Z_i, \mu') = cov(Z_1, \beta_1(X_1^* - Z_1) + \mu_1) = \beta_1 cov(Z_1, X_1^* - Z_1) + cov(Z_1, \mu_i) = 0$ because $Z_i = E(X_i^*|W_i)$ and $E(U|W_i) = 0$ $\boxed{\therefore cov(Z_i, \mu') = 0 \therefore \hat{\beta}_1 \text{is unbiased}}$

Question 3

a) β_1 is the parameter of interest, which indicates how much your lifetime earnings are if you're a veteran in the Vietnam War.

b) Being a veteran has a cosntant positive or negative effect on the log of lifetime earnings. If one is a veteran, since the variable is binary, the model predict you will earn a set amount difference than those who aren't.

c) $\hat{\beta}_2 = \frac{cov(wage, \beta_1)}{cov(veteran, \beta_2)}$

Parameter of interest is by how much more or less one who was a veteran in Vietnam makes on the log of lifetime earnings, β_1. This is different than (a) because population and sample differs, and therefore a different estimator. In sum, the terms have different meanings and interpretations.

d) It would not be a good estimator because it is then correlated with the error as preferences are involved. It is also no longer randomly assigned, and is thus biased and inconsistent.

e) If an individual was eligible for the draft, there was a likelihood he would participate in the Vietnam War. It is a good IV because it was a randomized process and thus even more random than study of veterans who volunteered in the army. We ignore cases of those who dodged the draft for medical reasons or fled to Canada. This is a good IV as well because $corr(Z_1, V_i) \neq 0$ and $corr(Z_i, \mu) = 0$.

f) No, because this removes randomization which is replaced by education decision in the sample's past, which is correlated with the error, $\therefore corr(Z_1, \mu) \neq 0$.

Question 4

a)

```
m1 <- lm(qty~price, data = fishdata)
m2 <- lm(qty~price + day1 +day2 + day3 + day4 + cold +
rainy, data = fishdata)
m3 <- ivreg(qty~price|stormy, data=fishdata)
m4 <- ivreg(qty~ price + day1 + day2 + day3 +day4 + cold
+rainy|stormy + day1 +day2 + day3 + day4 + cold + rainy,
data=fishdata)
stargazer(m1,m2,m3,m4, type = "latex" , title="OLS and IV
Eetimates of Demand andFunctions with Stormy Weather as IV", digits=5, float.env = "sidewaystable")
```

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: Tue, Apr 18, 2017 - 01:37:57 % Requires LaTeX packages: rotating

Table 1: OLS and IV Eetimates of Demand andFunctions with Stormy Weather as IV

	Dependent variable:			
	qty			
	OLS		*instrumental variable*	
	(1)	(2)	(3)	(4)
price	−0.54087***	−0.54455***	−1.08241**	−1.22280**
	(0.17864)	(0.17520)	(0.46572)	(0.53200)
day1		0.03162		−0.03329
		(0.20661)		(0.22620)
day2		−0.49348**		−0.53278**
		(0.20353)		(0.21973)
day3		−0.53924**		−0.57558**
		(0.20603)		(0.22212)
day4		0.09477		0.11788
		(0.20114)		(0.21594)
cold		−0.06160		0.06805
		(0.13448)		(0.17255)
rainy		0.06658		0.07203
		(0.17747)		(0.18998)
Constant	8.41867***	8.61689***	8.31379***	8.44175***
	(0.07622)	(0.16159)	(0.11462)	(0.21549)
Observations	111	111	111	111
R^2	0.07758	0.22288	−0.00019	0.10982
Adjusted R^2	0.06912	0.17007	−0.00937	0.04932
Residual Std. Error	0.71558 (df = 109)	0.67567 (df = 103)	0.74514 (df = 109)	0.72315 (df = 103)
F Statistic	9.16732*** (df = 1; 109)	4.22016*** (df = 7; 103)		
Note:				*p<0.1; **p<0.05; ***p<0.01

b) The standard errors in the paper do not appear to be robust, so they are not the same as in our first two regressions. Our m4 also differs from the paper's. Our m3 standard errors appear to be the usual HC1 errors.

c) In a storm, it was more difficult to fish offshore, so fewer fish would be caught. However, this doesn't imply that demand for fish would change as well.

d) If we run a regression of prices on stormy, see below. We find that stormy's coefficient is statistically significant. Stormy is not correlated with the error term in the explanatory equation, but we cannot feasibly check if it is correlated with the endogenous explanatory variables.

```
 m7 <- lm(price~stormy, data = fishdata)
cov7<- vcov(m7, type="HC1")
robust.se7 <- sqrt(diag(cov7))
coeftest(m7, vcov. = vcovHC(m7, type="HC1"))
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept) -0.290333   0.039557 -7.3396 4.068e-11 ***
## stormy       0.335262   0.073755  4.5456 1.425e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

e) The standard errors seem quite similar on most variables with the exception of price. However, the regression appears to fit the data better when including windspeed^2.

```
m5 <- ivreg(qty~price|stormy + windspd +windspd2,
data=fishdata)
cov5 <- vcovHC(m5, type="HC1")
robust.se5 <- sqrt(diag(cov5))
m6 <- ivreg(qty~ price + day1 + day2 + day3 +day4 + cold
+rainy|stormy + day1 +day2 + day3 + day4 + cold + rainy +
windspd +windspd2, data=fishdata)
cov6 <- vcovHC(m6, type="HC1")
robust.se6 <- sqrt(diag(cov6))
stargazer(m1,m2,m3,m4,m5,m6, type="latex" , title="OLS and
IV Estimates of Demand and Functions with Stormy Weather as IV, including Windspeed", digits=5, float.en
```

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: Tue, Apr 18, 2017 - 01:37:59 % Requires LaTeX packages: rotating

f) IV estimation increased the implied elasticity of demand over 100%, making obvious the specification errors in the OLS estimator that were not apparent before.

g) We test the hypothesis that the elasticity of demand is equal to -1. Hypotheses are: $H_0 = \beta_2 = -1$ vs. $\beta_2 \neq -1$. The t-test statistic is $\frac{\beta_2 - 1}{se(\beta_2)}$. With 10 degrees of freedom and at the 5% significance level, the rejection region is $|t| > 2.228$. $t_{m3} = -0.17$, $t_{m4} = -0.42$, $t_{m5} = -0.24$, $t_{m6} = -0.24$.

None of the IV cases reject the null that demand is unit elastic. This means a unit increase produces the same amount of revenue, implying a local maximum.

h) m5 and m6 are small and weak, and should not be used.

```
m4First1 <- lm(price~stormy + day1 + day2 + day3 + day4 +
cold + rainy, data = fishdata)
m4First2 <- lm(price~1 + day1 + day2 + day3 + day4 + cold +
rainy, data = fishdata)
```

Table 2: OLS and IV Estimates of Demand and Functions with Stormy Weather as IV, including Windspeed

	Dependent variable:					
	qty					
	OLS			*instrumental variable*		
	(1)	(2)	(3)	(4)	(5)	(6)
price	−0.54087***	−0.54455***	−1.08241**	−1.22280**	−1.09530**	−1.11337**
	(0.17864)	(0.17520)	(0.46572)	(0.53200)	(0.42194)	(0.45216)
day1		0.03162		−0.03329		−0.02282
		(0.20661)		(0.22620)		(0.22049)
day2		−0.49348**		−0.53278**		−0.52644**
		(0.20353)		(0.21973)		(0.21502)
day3		−0.53924**		−0.57558**		−0.56971**
		(0.20603)		(0.22212)		(0.21744)
day4		0.09477		0.11788		0.11415
		(0.20114)		(0.21594)		(0.21165)
cold		−0.06160		0.06805		0.04714
		(0.13448)		(0.17255)		(0.16177)
rainy		0.06658		0.07203		0.07115
		(0.17747)		(0.18998)		(0.18635)
Constant	8.41867***	8.61689***	8.31379***	8.44175***	8.31129***	8.47000***
	(0.07622)	(0.16159)	(0.11462)	(0.21549)	(0.10816)	(0.20040)
Observations	111	111	111	111	111	111
R^2	0.07758	0.22288	−0.00019	0.10982	−0.00394	0.14336
Adjusted R^2	0.06912	0.17007	−0.00937	0.04932	−0.01315	0.08514
Residual Std. Error	0.71558 (df = 109)	0.67567 (df = 103)	0.74514 (df = 109)	0.72315 (df = 103)	0.74653 (df = 109)	0.70940 (df = 103
F Statistic	9.16732*** (df = 1; 109)	4.22016*** (df = 7; 103)				
Note:						*p<0.1; **p<0.05; ***p<0.0

```
 m4ftest <- waldtest(m4First1, m4First2,
 vcov=vcovHC(m4First1, type='HC1'))$F[2]
 m5First1 <- lm(price~stormy + windspd + windspd2, data =
fishdata)
m5First2 <- lm(price~1 + windspd + windspd2, data =
fishdata)
m5ftest <- waldtest(m5First1, m5First2,
vcov=vcovHC(m5First1, type='HC1'))$F[2]
m6First1 <- lm(price~stormy + day1 + day2 + day3 + day4 +
cold + rainy + windspd + windspd2, data = fishdata)
m6First2 <- lm(price~1 + day1 + day2 + day3 + day4 + cold +
rainy + windspd + windspd2, data = fishdata)
m6ftest <- waldtest(m6First1, m6First2,
vcov=vcovHC(m6First1, type='HC1'))$F[2]
m4ftest
```

```
## [1] 16.27249
```

```
m5ftest
```

```
## [1] 4.369539
```

```
m6ftest
```

```
## [1] 3.065217
```