PROBLEM SET 5

CLAUDIA GENTILE

Exercise 5.1

a) • For a randomly chosen individual x_i:

$$E(\ln w_i^U - \ln w_i^N \mid x_i) = E(x_i'\beta^U - x_i'\beta^N + u_i^U - u_i^N \mid x_i) = x_i'\beta^U - x_i'\beta^N$$

• For a union worker with x_i:

$$E[\ln w_i^U - \ln w_i^N \mid x_i, U_i^* > 0] = \cancel{E[\ln w_i^U - \ln w_i^N}$$

$$U_i^* = \delta_0 + \delta_1(x_i'\beta^U - x_i'\beta^N + u_i^U - u_i^N) + x_i'\delta_2 + z_i'\delta_3 - v_i > 0$$

$$\underbrace{\delta_1(u_i^U - u_i^N) - v_i}_{=\varepsilon} > \underbrace{-\delta_0 - x_i'\delta_2 - z_i'\delta_3 - \delta_1(x_i'\beta^U - x_i'\beta^N)}_{=-c}$$

define:

$$P(U_i^* > 0 \mid x_i, z_i) = \Phi(c) \quad \text{(normalizing } \sigma_\varepsilon^2 = 1\text{)}$$

$$E[\ln w_i^U - \ln w_i^N \mid x_i, U_i^* > 0] = x_i'\beta^U - x_i'\beta^N + E[u_i^U - u_i^N \mid x_i, \varepsilon > c]$$

b) No it is not possible.

$$E[\ln w_i^u \mid U_i^* > 0] = \cancel{E[\ln} \; x_i'\beta^u + E[u_i^u \mid \varepsilon > c] =$$

$$= x_i'\beta^u + E[u_i^u \mid \cancel{\delta_0 +} \delta_1(u_i^u + u_i^N) + v_i > c] = *$$

I can further simplify the expression making a regression of

$$u_i^u = a_1(\delta_1(u_i^u + u_i^N) + v_i) + \underbrace{V}_{\text{orthogonal component}}$$

$$* = x_i'\beta^u + E[a_1\varepsilon + V \mid \varepsilon > c] =$$

$$= x_i'\beta^u + E[a_1\varepsilon \mid \varepsilon > c] + E[V \mid \varepsilon > c] =$$

$$\text{" } E[V] = 0$$

$$= x_i'\beta^u + a_1 E[\varepsilon \mid \varepsilon > c]$$

In order to ~~make~~ ~~make an~~ estimate β^u by using only data on x_i and w_i for union members we would need to include the selection term in the regression. Indeed since c contains x the selection term is correlated with our regressors and excluding it from the regression would bias ~~our~~ the results.

c) if we assume that u_i^u, u_i^N and v_i have a trivariate normal distribution with mean 0 and covariance matrix $\Omega = \begin{bmatrix} \sigma_u^2 & 0 & 0 \\ 0 & \sigma_N^2 & 0 \\ 0 & 0 & \sigma_v^2 \end{bmatrix}$ and $U_i = 1$ if $U_i^* > 0$, $U_i = 0$ if $U_i^* \leq 0$, we can estimate

$$\ln L = \sum_i U_i \left[\ln F(d) + \ln\left(\frac{f(u_i^u/\sigma_u)}{\sigma_u}\right)\right] + (1-U_i)\left[\ln(1-F(d)) + \ln\left(\frac{f(u_i^N/\sigma_N)}{\sigma_N}\right)\right]$$

d) Heckman - two steps approach

Following from point b:

$$E[\ln w_i^u \mid U_i^* > 0] = x_i'\beta^u + a_1\sigma_\varepsilon E\left[\frac{\varepsilon}{\sigma_\varepsilon} \,\middle|\, \frac{\varepsilon}{\sigma_\varepsilon} > \frac{c}{\sigma_\varepsilon}\right] =$$

$$= x_i'\beta^u + a_1\sigma_\varepsilon\lambda\left(\frac{c}{\sigma_\varepsilon}\right)$$ where λ is the inverse mill ratio.

Note $\varepsilon = \delta_1(u_i^u + u_i^N) + v_i$ is a normal random variable since it's a linear combination of normal random variables.

STEP 1:

From point b) $Pr(U_i^* > 0 \mid x_i, z_i) = \Phi(a_0 + x_i'a_1 + z_i'a_2)$ can be estimated by probit.

The estimated $\hat{a}_0 + x_i' \hat{a}_1 + z_i' \hat{a}_2 = \widehat{\left(\frac{c}{\sigma_\varepsilon}\right)}$ (to see why look back at definition of c which is given under point b)

STEP 2: We can use $\widehat{\left(\frac{c}{\sigma_\varepsilon}\right)}$ to compute $\lambda\left(\widehat{\frac{c}{\sigma_\varepsilon}}\right)$.

Then we can perform an OLS regression of $\ln w_i^U$ on x_i' and $\lambda\left(\frac{\hat{c}}{\sigma_\varepsilon}\right)$. The estimate coefficient on x_i' is $\hat{\beta}_U$, consistent estimator of β_U.

Same procedure to estimate $\hat{\beta}_N$

e) They can be recovered from the second step.

$\hat{a}_2 = \hat{\delta}_3$

$\hat{a}_0 = \hat{\delta}_0$

$\hat{a}_1 = \delta_1(\beta^U - \beta^N)$ $\quad \hat{\delta}_1 = \dfrac{\hat{a}_1}{\hat{\beta}_U - \hat{\beta}^N}$

f) The ED coefficients are all higher in magnitude in the non-union wage equation. In both ~~equal~~ union and non-union workers coefficients on ED_1 and ED_2 are negative = people that have a low degree of education of level 1 and 2 have a lower wage relative to people that have no education ceteris paribus. Coefficients on ED3 - ED5 show that the relative rate of return to education is higher in the non-unionized sector.

g) Union = $U_i^* = \gamma_0 + \gamma_1 X + \gamma_2 Z + \varepsilon^* > 0 \Rightarrow -\varepsilon^* < \gamma_0 + \gamma_1 X + \gamma_2 Z$

selectivity variable $= -0.168 = E\left[-\varepsilon^* \mid -\varepsilon^* < \gamma_0 + \gamma_1 X + \gamma_2 Z\right]$ $\quad \hat{\varepsilon}_U^* = 0.168$

Non-union = $U_i^* = \gamma_0 + \gamma_1 X + \gamma_2 Z + \varepsilon^* < 0 \Rightarrow -\varepsilon^* > \gamma_0 + \gamma_1 X + \gamma_2 Z$

selectivity var $= 0.136 = E\left[-\varepsilon^* \mid -\varepsilon^* > \gamma_0 + \gamma_1 X + \gamma_2 Z\right]$

$\hat{\varepsilon}_N^* = -0.136$

$\Rightarrow$ workers select in the sector where they ~~expect~~ have an higher wage

The fact that $\hat{\varepsilon}_U^* > 0$ means unionized workers expect to benefit from joining the union, while $\hat{\varepsilon}_N^* < 0$ means non-unionized workers expect to lose from joining the union. ~~Therefore we observe unionized~~ wage only ~~for~~ ~~people~~ workers ~~that expect~~ have $U_i^* > 0$.

h) The wage differential seems to have quite a big impact on the choice of the sector following what is explained in point g-

Also Table 6 in the paper supports this claim.

a) see ~~tab~~ regression tables at the end

Exercise 5.2

(y_1 = inlf y_2 = kidslt6)

b) The bivariate probit model estimates jointly

$Pr(y_1 = 1 \mid x) = \Phi(x'\beta_1)$ $y_1 = \Phi(x'\beta_1) + \varepsilon_1$ (*)

$Pr(y_2 = 1 \mid x) = \Phi(x'\beta_2)$ $y_2 = \Phi(x'\beta_2) + \varepsilon_2$ (**)

and it estimates $cov(\varepsilon_1, \varepsilon_2 \mid x) = \rho$. (Assuming = $E(\varepsilon_i \mid x) = 0 \quad i = 1,2$)

The advantage of the bivariate probit model is that we can ~~use~~ test for $\rho = 0$. LR test $\Rightarrow$ reject $\rho = 0$. This provides evidence that there are some variables that are in ε_2 (determine selection into motherhood) that also are in ε_1 (determine selection into labor force participation).

c) Include ~~Pr~~ kidslt6 in first biprobit equation.

$E(y_1|x) = Pr(y_1 = 1 \mid x) = \Phi(x'\beta_1 + y_2\beta_3)$ $y_1 = \Phi(x'\beta_1 + y_2\beta_3) + \varepsilon_1$

$E(y_2|x) = Pr(y_2 = 1 \mid x) = \Phi(x'\beta_2)$ $y_2 = \Phi(x'\beta_2) + \varepsilon_2$

Assuming $E(\varepsilon_1 \mid x, y_2) = 0$

$E(\varepsilon_2 \mid x) = 0$

We know from model (**) that y_2 and x are correlated, so if we do not include y_2 in equation 1 $E(\varepsilon_1 \mid x) \neq 0$, which would bias our results. Anyhow by not including y_1 into the second equation we are assuming ~~y~~ selection into ~~the~~ labor force does not affect selection into motherhood. If this was the case then $E(\varepsilon_2 \mid x) \neq 0$. Plus we are assuming there are no other variables correlated with x that influence selection in labor force.

$\hat{\rho} = 0.0952$ LR test of $\rho = 0$ has a p-value = 0.6.

We do not reject $\rho = 0$. We do not reject $cov(\varepsilon_1, \varepsilon_2 \mid x, y_1) = 0$.

Once x and y_2 are controlled for there are no other unobservables that determine both selection into motherhood and selection into working.

d) If we believe selection on ~~unob~~ observables is stronger ~~an upper bound for selection~~ than selection ~~for~~ on unobservables, we can use selection on observables to build an upper bound for selection on unobservables.

~~y1z~~ $cov(\varepsilon_1, \varepsilon_2 \mid x)$ represents a measure of the selection on unobservables. The greater is this covariance the more unobservable variables causing selection into motherhood

Scanned by CamScanner

also cause selection into labour force participation $\Rightarrow$ the greater is the selection on unobservables.

Selection on observables can be measured as = $\dfrac{cov(x'\beta_1, x'\beta_2)}{var(x'\beta_1)}$.

Then we can estimate the bivariate probit imposing $\rho = \dfrac{cov(x'\beta_1, x'\beta_2)}{var(x'\beta_1)}$ to get an estimate with an upper bound on selection. Imposing this ρ in the biprobit ~~We expect the impact of y_2 on y_1~~ model would give a lower bound effect ~~estimate~~ of ~~the~~ having more than one child on labor force participation.

e) Imposing a constraint on $\rho = -0.7$ the p-value on Kidsm1 becomes 0.07. We fail to reject the coefficient is equal to zero at the 5% significance level. In the tables coming below I report estimates for $\rho = -0.75$ for which the coefficient is not statistically significant (neither at the 10% significance level).

f) $\rho = \dfrac{cov(x'\beta_1, x'\beta_2)}{var(x'\beta_1)} = \dfrac{-0.000385}{0.0496} = -0.0078$

$\Rightarrow$ value in point e is not plausible.

g) Reestimating the bivariate ~~mod~~ probit imposing $\rho = -0.0078$ I get as lower bound effect of having more than one ~~one more~~ child on labor force participation = having more than one child decreases the probability to participate in the labor force by 10 percentage points; (obtained running the margins command) ~~after~~

Exercise 5.2 - TABLES

POINT A

Probit regression

=1 if in lab frce, 1975	(faminc - wage*hours)/1000	-0.011	
		(2.44)*	
	years of schooling	0.105	
		(4.38)**	
	actual labor mkt exper	0.125	
		(6.89)**	
	exper^2	-0.002	
		(3.46)**	
	woman's age in yrs	-0.029	
		(4.19)**	
kidslt6d	(faminc - wage*hours)/1000		0.006
			(1.01)
	years of schooling		0.027
			(0.91)
	actual labor mkt exper		-0.056
			(1.98)*
	exper^2		0.001
			(0.97)
	woman's age in yrs		-0.106
			(9.91)**
	Constant	-0.619	3.252
		(1.49)	(5.74)**
N		753	753

* $p<0.05$; ** $p<0.01$

POINT B

Biprobit regression

=1 if in lab frce, 1975	(faminc - wage*hours)/1000	-0.012
		(2.51)*
	years of schooling	0.108
		(4.50)**
	actual labor mkt exper	0.127
		(6.98)**
	exper^2	-0.002
		(3.52)**
	woman's age in yrs	-0.029
		(4.23)**
	Constant	-0.633
		(1.52)
kidslt6d	(faminc - wage*hours)/1000	0.004
		(0.81)
	years of schooling	0.029
		(0.98)
	actual labor mkt exper	-0.057
		(2.12)*
	exper^2	0.001

		(1.05)
	woman's age in yrs	-0.101
		(9.74)**
	Constant	3.074
		(5.52)**
athrho	Constant	-0.559
		(6.39)**
N		753

* *p*<0.05; ** *p*<0.01

POINT C

Biprobit regression (2)

=1 if in lab frce, 1975	# kids < 6 years	-0.995
		(4.05)**
	(faminc - wage*hours)/1000	-0.012
		(2.44)*
	years of schooling	0.130
		(5.16)**
	actual labor mkt exper	0.120
		(6.25)**
	exper^2	-0.002
		(2.98)**
	woman's age in yrs	-0.058
		(6.08)**
	Constant	0.607
		(1.15)
kidslt6d	(faminc - wage*hours)/1000	0.006
		(1.04)
	years of schooling	0.027
		(0.89)
	actual labor mkt exper	-0.056
		(1.98)*
	exper^2	0.001
		(0.98)
	woman's age in yrs	-0.106
		(9.92)**
	Constant	3.263
		(5.76)**
athrho	Constant	0.096
		(0.52)
N		753

* *p*<0.05; ** *p*<0.01

POINT E

Biprobit regression (unconstrained)

=1 if in lab frce, 1975	kids1	-1.243
		(2.45)*
	kidsm1	-2.161
		(4.19)**

	(faminc - wage*hours)/1000	-0.012 (2.38)*
	years of schooling	0.130 (5.21)**
	actual labor mkt exper	0.114 (5.07)**
	exper^2	-0.002 (2.43)*
	woman's age in yrs	-0.063 (5.09)**
	Constant	0.873 (1.22)
kidslt6d	(faminc - wage*hours)/1000	0.006 (1.08)
	years of schooling	0.026 (0.88)
	actual labor mkt exper	-0.056 (1.97)*
	exper^2	0.001 (0.98)
	woman's age in yrs	-0.106 (9.96)**
	Constant	3.269 (5.79)**
athrho	Constant	0.237 (0.73)
N		753

* p<0.05; ** p<0.01

Biprobit regression (constrained rho = -0.75)

=1 if in lab frce, 1975	kids1	0.547 (4.54)**
	kidsm1	-0.318 (1.35)
	(faminc - wage*hours)/1000	-0.012 (2.59)**
	years of schooling	0.104 (4.40)**
	actual labor mkt exper	0.130 (7.32)**
	exper^2	-0.002 (3.82)**
	woman's age in yrs	-0.019 (2.61)**
	Constant	-1.096 (2.58)**
kidslt6d	(faminc - wage*hours)/1000	0.004 (0.69)
	years of schooling	0.030 (1.05)

	actual labor mkt exper	-0.055
		(2.14)*
	exper^2	0.001
		(1.04)
	woman's age in yrs	-0.096
		(9.77)**
	Constant	2.882
		(5.38)**
athrho	Constant	-0.973
N		753

* *p*<0.05; ** *p*<0.01

POINT G

Biprobit regression (constrained rho = -0.0078)

=1 if in lab frce, 1975	kids1	-0.852
		(5.56)**
	kidsm1	-1.805
		(6.00)**
	(faminc - wage*hours)/1000	-0.012
		(2.51)*
	years of schooling	0.129
		(5.14)**
	actual labor mkt exper	0.123
		(6.56)**
	exper^2	-0.002
		(3.17)**
	woman's age in yrs	-0.055
		(6.85)**
	Constant	0.435
		(0.95)
kidslt6d	(faminc - wage*hours)/1000	0.006
		(1.01)
	years of schooling	0.027
		(0.91)
	actual labor mkt exper	-0.056
		(1.98)*
	exper^2	0.001
		(0.97)
	woman's age in yrs	-0.106
		(9.91)**
	Constant	3.251
		(5.74)**
athrho	Constant	-0.008
N		753

* *p*<0.05; ** *p*<0.01