ECONOMETRICS II - PS 5

Miriam Venturini

a)

$$\mathbb{E}\left[\ln w_i^U - \ln w_i^N \mid X_i\right] = \mathbb{E}\left[x_i'\beta^U + u_i^U - x_i'\beta^N - u_i^N \mid X_i\right]$$

$$= \mathbb{E}\left[x_i'(\beta^U - \beta^N) \mid X_i\right] + \mathbb{E}\left[u_i^U - u_i^N \mid X_i\right] = x_i'(\beta^U - \beta^N)$$

$$\mathbb{E}\left[\ln w_i^U \mid \delta_0 + \delta_1(\ln w_i^U - \ln w_i^N) + x_i'\delta_2 + z_i'\delta_3 - v_i > 0, X_i\right] =$$

$$= \mathbb{E}\left[x_i'\beta^U + u_i^U \mid \delta_0 + \delta_1(x_i'(\beta^U - \beta^N) + u_i^U - u_i^N) + x_i'\delta_2 + z_i'\delta_3 - v_i > 0, X_i\right]$$

$$= x_i'\beta^U + \mathbb{E}\Big[u_i^U \mid u_i^U - u_i^N - \frac{v_i}{\delta_1} > \underbrace{-x_i'(\beta^U - \beta^N) - \frac{(\delta_0 - x_i'\delta_2 - z_i'\delta_3)}{\delta_1}}_{-C}, X_i\Big]$$

$$= x_i'\beta^U + \mathbb{E}\left[u_i^U \mid u_i^U - u_i^N - \frac{v_i}{\delta_1} > -C, X_i\right]$$

regress u_i^U on $u_i^U - u_i^N - \frac{v_i}{\delta_1}$

$$u_i^U = a_u\left(u_i^U - u_i^N - \frac{v_i}{\delta_1}\right) + \varepsilon_i$$

$$a_u = \frac{Cov\left(u_i^U, u_i^U - u_i^N - \frac{v_i}{\delta_1}\right)}{Var\left(u_i^U - u_i^N - \frac{v_i}{\delta_1}\right)} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_N^2 + \frac{\sigma_v^2}{\delta_1^2}}$$

$$= x_i'\beta^U + \mathbb{E}\left[a_u\left(u_i^U - u_i^N - \frac{v_i}{\delta_1}\right) + \varepsilon_i \mid u_i^U - u_i^N > -C\right]$$

$$= x_i'\beta^U + a_u\,\mathbb{E}\Big[\underbrace{u_i^U - u_i^N - \frac{v_i}{\delta_1}}_{D} \mid u_i^U - u_i^N > -C\Big] + \overbrace{\mathbb{E}\left[\varepsilon_i \mid u_i^U - u_i^N > -C\right]}^{=\mathbb{E}[\varepsilon_i] = 0}$$

$$= x_i'\beta^U + a_u\,\mathbb{E}\left[D \mid D > -C\right] \qquad d = -\frac{C}{\sigma}$$

$$= x_i'\beta^U + a_u\sigma\,\mathbb{E}\left[\frac{D}{\sigma} \mid \frac{D}{\sigma} > d\right] = x_i'\beta^U + a_u\sigma\lambda(d) = x_i'\beta^U + a_u\sigma\frac{\phi(d)}{1-\Phi(d)}$$

$$\mathbb{E}\left[\ln w_i^N \mid \delta_0 + \delta_1(\ln w_i^U - \ln w_i^N) + x_i'\delta_2 + z_i'\delta_3 - v_i > 0, X_i\right] =$$

$$= \mathbb{E}\left[x_i'\beta^N + u_i^N \mid \delta_0 + \delta_1(x_i'(\beta^U - \beta^N) + u_i^U - u_i^N) + x_i'\delta_2 + z_i'\delta_3 - v_i > 0, X_i\right]$$

$$= x_i'\beta^N + \mathbb{E}\left[u_i^N \mid \delta_0 + \delta_1(x_i'(\beta^U - \beta^N) + u_i^U - u_i^N) + x_i'\delta_2 + z_i'\delta_3 - v_i > 0, X_i\right]$$

$$= x_i'\beta^N + \mathbb{E}\left[u_i^N \mid u_i^U - u_i^N - \frac{v_i}{\delta_1} > -x_i'(\beta^U - \beta^N) - \frac{(\delta_0 - x_i'\delta_2 - z_i'\delta_3)}{\delta_1}, X_i\right]$$

$$= x_i'\beta^N + \mathbb{E}\left[u_i^N \mid u_i^U - u_i^N - \frac{v_i}{\delta_1} > -C, X_i\right] = x_i'\beta^N + \mathbb{E}\left[a_N\left(u_i^U - u_i^N - \frac{v_i}{\delta_1}\right) + \varepsilon_i \mid u_i^U - u_i^N > -C\right]$$

$$u_i^N = a_N\left(u_i^U - u_i^N - \frac{v_i}{\delta_1}\right) + \varepsilon_i$$

$$a_N = \frac{\sigma_N^2}{\sigma_u^2 + \sigma_N^2 + \frac{\sigma_v^2}{\delta_1^2}}$$

$$= x_i'\beta^N + a_N\,\mathbb{E}\left[D \mid D > -C\right] \qquad d = -\frac{C}{\sigma}$$

$$= x_i'\beta^N + a_N\,\mathbb{E}\left[\frac{D}{\sigma} \mid \frac{D}{\sigma} > d\right] = x_i'\beta^N + a_N\sigma\frac{\phi(d)}{1-\Phi(d)}$$

$$\mathbb{E}\left[\ln w_i^U - \ln w_i^N \mid U_i^* > 0\right] = x_i'(\beta^U - \beta^N) + (a_u - a_N)\sigma\left(\frac{\phi(d)}{1-\Phi(d)}\right)$$

$$a_u - a_N = \frac{\sigma_u^2 - \sigma_N^2}{\sigma_u^2 + \sigma_N^2 + \frac{\sigma_v^2}{\delta_1^2}}$$

b) No, it's not possible since the selection in the 2 groups depends on the difference in wage that each person would face

$$\mathbb{E}\left[u_i^U \mid U_i = 1\right] \neq 0 \qquad \mathbb{E}\left[u_i^N \mid U_i = 0\right] \neq 0$$

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 $\Omega = \begin{bmatrix} \sigma_u^2 & 0 & 0 \\ 0 & \sigma_N^2 & 0 \\ 0 & 0 & \sigma_v^2 \end{bmatrix}$

$U_i = 1$ if $U_i^* > 0$ $\quad U_i = 0$ if $U_i^* \leq 0$

we can estimate

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

d) Heckman 2-step method

First step: you run a probit regression on D and estimate $\hat{\gamma}$

- $\Pr(\ln w_1 > \ln w_2) = \Pr(D_* > d) = \Phi(d\gamma)$ where γ is the parameter vector and d is a vector of all the observables included in d (Z and X in this case)

You assume $\sigma = 1$ (probit assumption)

- use $\hat{\gamma}$ to estimate $\hat{d} \Rightarrow \lambda(\hat{d})$

Second step: you run 2 OLS regressions of $\ln w_U$ and $\ln w_N$ on X and $\lambda(\hat{d})$

$\hookrightarrow$ we get coefficients $\hat{\beta}_N$ and $\hat{\beta}_U$

e) You can run a probit with latent variable

$$U_i^* = \delta_0 + \delta_1(\ln w_i^U - \ln w_i^N) + x_i'\delta_2 + z_i'\delta_3 - v_i$$

and observed $U_i = 1$ if $U_i^* > 0$

You plug in for $\log w_i^U - \ln w_i^N$ values

$$\widehat{\log w_i^U} = x_i'\hat{\beta}^U$$
$$\widehat{\log w_i^N} = x_i'\hat{\beta}^N$$

$$U_i = \delta_0 + x_i'\left[(\hat{\beta}^U - \hat{\beta}^N) + \delta_2\right]\delta_1 + z_i'\delta_3 - v_i$$

$$= \gamma_0 + x_i\gamma_1 + z_i\gamma_2$$ just renaming

from here, after regression you have $\hat{\gamma}_1$; knowing $\hat{\beta}^U$ and $\hat{\beta}^N$ you can actually recover δ

f)

table 1

$ED_5 - ED_1 = 0.139 + 0.108 = 0.247$
$ED_5 - ED_4 = 0.139 - 0.111 = 0.028$
$ED_4 - ED_3 = 0.111 - 0.052 = 0.059$
$ED_3 - ED_2 = 0.052 + 0.033 = 0.085$
$ED_2 - ED_1 = -0.033 + 0.108 = 0.075$

table 2

$ED_5 - ED_1 = 0.282 + 0.049 = 0.331$
$ED_5 - ED_4 = 0.282 - 0.157 = 0.125$
$ED_4 - ED_3 = 0.157 - 0.087 = 0.07$
$ED_3 - ED_2 = 0.087 + 0.016 = 0.103$
$ED_2 - ED_1 = -0.016 + 0.049 = 0.033$

higher in NON-UNION sector

$ED_5 - ED_4$
$ED_4 - ED_3$
$ED_3 - ED_2$

$ED_5 - ED_1$ is higher in non-union than in union sector

for relatively high educated people the relative rate of return to education is higher in non-unionized sector for scarsely educated people the relative rate of return to education is higher in the unionized sector.

In the unionized sector market experience has a larger effect on wage increment

Being white increases your wage more in non-Union sector. The difference white - non white is smaller in UNION SECTOR

Being male increases your wage more in unionized sector. The difference male-female is bigger in UNION SECTOR

Having an health limitation reduces wages more in non-Union sector. The difference healthy - nonhealthy is bigger in non-Union sector

g) for the non-union sector the Mills ratio is defined $-\frac{f(\psi_i)}{F(\psi_i)} = -\frac{\phi(\gamma_0 + \gamma_1 x_i' + \gamma_2 z_i')}{\Phi(\gamma_0 + \gamma_1 x_i' + \gamma_2 z_i')}$

for the union sector it is defined $\frac{f(\psi_i)}{1-F(\psi_i)} = \frac{\phi(\gamma_0 + \gamma_1 x_i' + \gamma_2 z_i')}{1-\Phi(\gamma_0 + \gamma_1 x_i' + \gamma_2 z_i')}$

so the coefficient is positive in both cases $\rightarrow$ people self-select /are selected in the sector where their wage is actually "predicted" to be higher given observables (x_i and z_i)

h) The wage differential seems to have quite a big predicted effect on the choice of the sector, higher than most of the other variables taken into consideration

Table 6 supports this claim (in a structural form estimation)

The most powerful factor determining unionization status is the union - non union wage differential

i) The reduced form estimates are useful to identify the net effects of costs and benefits that each observable characteristic induces for the workers, it is basically estimating U replacing $\ln W_i$ as a function of x_i and z_i

For example, we observe that unionized firms select more highly educated workers, but the return on education in the non union sector will induce higher educated workers into the non union sector. So, while in table 6 we observe a high and significant effect of high education on the probability of being in union-sector, the effect is balanced out, in table 7, by the higher returns on education in the non-union sector. The net effect of high education is then negative but non significant.

Comments/answers to exercise 2 are on the dofile.