Econometrics part 2, PS 5

Mangiante Giacomo

April 18, 2019

Discussed with: Krishna Snirivasan, Miriam Venturini, Aleksandra Malova

1 The Roy model

1.1 a)

Expected union/non-union wage differential for a randomly chosen individual with characteristics x_i:

$$E[ln(w_i^u) - ln(w_i^n)|x_i] = x_i^{'}\beta^u + E[u^u|x_i] - x_i^{'}\beta^n - E[u^n|x_i] = x_i^{'}(\beta^u - \beta^n)$$

Expected wage differential for a union worker with characteristics x_i:

$$E[ln(w_i^u) - ln(w_i^n)|x_i, U^* > 0] = x_i^{'}(\beta^u - \beta^n) + E[u^u - u^n|\delta_0 + \delta_1(ln(w_i^u) - ln(w_i^n)) + x_i^{'}\delta_2 + z_i^{'}\delta_3 - v_i > 0]$$

We can define the error component in the second term as $D = v_i - \delta_1(u^u - u^n)$ and the observed component as $c = \delta_0 + \delta_1(x_i^{'}(\beta^u - \beta^n)) + x_i^{'}\delta_2 + z_i^{'}\delta_3$ obtaining:

$$E[u^u - u^n|D < c]$$

In order to solve for this term, we perform the following regression:

$$(u^u - u^n) = a_1 D + V$$

With V orthogonal to D. Since $u^u - u^n$ distributes as a normal with mean 0 and variance $\sigma_u^2 + \sigma_n^2$:

$$a_1 = \frac{Cov(u^u - u^n, D)}{Var(D)} = \frac{\delta_1(\sigma_u^2 + \sigma_n^2)}{\delta_1^2(\sigma_u^2 + \sigma_n^2) + \sigma_v^2}$$

Therefore we can rewrite:

$$E[u^u - u^n|D < c] = E[a_1 D + V|D < c] = a_1 E[D|D < c] = a_1 \sigma E[\frac{D}{\sigma}|\frac{D}{\sigma} < \frac{c}{\sigma}] = a_1 \sigma \lambda(-c)$$

Finally:

$$E[ln(w_i^u) - ln(w_i^n)|x_i, U^* > 0] = x_i^{'}(\beta^u - \beta^n) + a_1 \sigma \lambda(-c) = x_i^{'}(\beta^u - \beta^n) + \delta_1(\sigma_u^2 + \sigma_n^2)\lambda(-c)$$

1.2 b)

No, we can not retrieve the coefficients for β by simply running a two linear regressions because we have a sample selection, i.e. the workers are not randomly assigned to be part of a union or not. Therefore, if we run a OLS regression we would have biased coefficient due to omitted variable/endogeneity problem.

1.3 c)

Both equations could be estimated jointly by maximum likelihood using a biprobit model. Replacing the two equations for $ln(w_i^u)$ and $ln(w_i^n$ into the dummy variable U (equals 1 if $U^* > 0$ and o otherwise) we can compute:

$$Pr(U = 1|x_i) = Pr(\delta_0 + \delta_1(ln(w_i^u) - ln(w_i^n)) + x_i^{'}\delta_2 + z_i^{'}\delta_3 - v_i > 0)$$

Which can be further simplified by gathering the error term components (in point a) defined as D) on one side of the inequality;

$$Pr(D > -(\delta_0 + \delta_1(x_i^{'}(\beta^u - \beta^n)) + x_i^{'}\delta_2 + z_i^{'}\delta_3))$$

Dividing both sides by the standard error of D (defined in the previous point as σ):

$$Pr(U = 1|x_i) = \Phi(\frac{\delta_0 + \delta_1(x_i^{'}(\beta^u - \beta^n)) + x_i^{'}\delta_2 + z_i^{'}\delta_3)}{\sigma}) = \Phi(d)$$

Therefore we can solve for the Maximum Likelihood of a function with a Bernoulli distribution:

$$logL = \sum_{i=1}^{\infty} Ulog(\Phi(d)) + (1 - U)(1 - log(\Phi(d)))$$

1.4 d)

A two-steps method we could use is the Heckman model: in the first step we perform a probit on union dummy controlling for observable and normalizing the variance of the error component to 1 in order to compute the likelihood of the workers of being part of a union. This would allow to compute the estimated inverse Mill ratio $\lambda(-c)$. In the second step, we run a linear regression on $ln(w_i^u)$ (and similarly for $ln(w_i^n)$) including the controls and the estimated inverse Mill ratio. This would give us $\hat{\beta}^u$ and $\hat{\beta}^n$.

1.5 e)

In order to compute the structural parameters of the union status equation, i.e. δ, we could use the estimated $\hat{\beta}^u$ and $\hat{\beta}^n$ to computed the fitted values $\hat{ln}(w^u$ and $\hat{ln}(w^n$ and include them in the probit of U^*.

1.6 f)

From Table 1 and 2 we see that the wage differential is larger for non-unionized workers. Moreover, the market experience has a larger effect for unionized workers.

There is less difference in wage for female as well as for black workers for non-unionized. Finally, the same holds looking at the variable for health impediments.

1.7 g)

A positive value of the selectivity variable in Table 1 means that given certain characteristics the average wage of workers unionized is higher than for workers non unionized. The opposite for Table 2. Therefore, we see that the workers optimally select into union/non-union.

1.8 h)

Looking at Table 6, we can see that the role played by wage differential in explaining the probability of union membership is statistically significant.

1.9 i)

The estimates of the reduced form, unlike the ones from the structural model, reported in Table 7 give the net effects for the various factors on union status.

2 Application: female labor supply using Altonji-Elder-Taber-Bounds

2.1 a)

See Stata do file.

2.2 b)

The biprobit approach allows us to control for the potential dependency between the two decisions (enter the labor force and having a child) by introducing correlation among the error terms of the two probit regressions.

2.3 c)

Since we suspect that the selection into labor force is affected by the selection into motherhood as well we should include this variable in the main probit as well.

The implicit assumption we make is that the selection into labor force is affected by the selection into motherhood but not the other way around. Moreover, we further assume that the joint distribution of the error terms is a bivariate normal distribution.

The correlation ρ between the error terms is equal to 0.2976. However, the coefficient is not significant therefore once we control for the selection into motherhood the correlation between errors disappear.

2.4 d)

In their paper, Altonji, Elder and Taber quantify and bound the selection bias due to selection on unobservables by varying the correlation between the error terms of the two probits.

2.5 e)

In the Stata code for point e) I iterate the biprobit regression keeping changing the value for artrho until the z statistics relative to the coefficient of the dummy variable resulted lower than 1.28 (i.e. not statistically significant at 10% level).

The relative value of ρ that makes the effect of having more than one child on labor force participation not significant is approximately bounded between -0.47 and -0.63.

2.6 f)

The value of ρ we found seems reasonable: if we assume that the selection into motherhood has a negative effect on the selection into the labor force than if we do not take into account this effect we would expect a negative correlation between the error terms.

2.7 g)

Since in the bivariate case it holds that:

$$0 \leq \rho \leq \frac{Cov(X'\beta, X'\gamma)}{Var(X'\gamma)}$$

We can find the lower bound relative to the coefficient for the dummy variable by imposing $\rho = Cov(X'\beta, X'\gamma)/Var(X'\gamma)$ which in our case corresponds to $\rho = 0.476$. The estimated lower bound for the coefficient is then -1.747.