Lec 8: Selection

Eric Hsienchen Chu*

Spring, 2024

(⊛) Suggested readings: Hansen (2022), Ch27.

1 Sample Selection Model

Motivation. Suppose Y_i is wage (log), X_i is characteristics, and u_i is unobserved skills. Consider $Y_i = m(X_i;\beta) + u_i$, where $\mathbb{E}[u_i] = 0,\ X_i \perp\!\!\!\perp u_i$ (so $\mathbb{E}[u_i|X_i] = 0$) and m is known up to β.

Problem. We only observe Y_i's that are *employed*! Let employed be $D_i = 1$, then

$$D_i = \mathbb{1}\{X_i'\gamma + v_i \geq 0\}, \tag{1.1}$$

where v_i is utility from work. We observe $Y_i \iff D_i = 1 \iff X_i'\gamma + v_i \geq 0$.

Assume $X_i \perp\!\!\!\perp v_i \implies X_i \perp\!\!\!\perp (u_i, v_i)$. What we really "observe" is

$$\begin{aligned}
\mathbb{E}[Y_i|X_i, \underbrace{X_i'\gamma + v_i \geq 0}_{\text{the "selection"}}] &= \mathbb{E}[m(X_i;\beta) + u_i|X_i, X_i'\gamma + v_i \geq 0] & (1.2)\\
&= m(X_i;\beta) + \mathbb{E}[u_i|X_i, X_i'\gamma + v_i \geq 0] & (1.3)\\
&= m(X_i;\beta) + \underbrace{\mathbb{E}[u_i|X_i'\gamma + v_i \geq 0]}_{\text{Selection bias } (\bigstar)} \longleftarrow \text{since } X_i \perp\!\!\!\perp v_i & (1.4)
\end{aligned}$$

Instead of our CEF of interest:

$$\mathbb{E}[Y_i|X_i] = m(X_i;\beta) \tag{1.5}$$

Remark. The selection bias $= 0$ if $u_i \perp\!\!\!\perp (X_i, v_i)$. But $u_i \perp\!\!\!\perp v_i$ is likely to be violated! Why? Skills (u_i) and utility from work (v_i) can be correlated positively.

*Department of Economics, University of Wisconsin-Madison. hchu38@wisc.edu. This is lecture notes from the second half of ECON710: Economic Statistics and Econometrics II. Instructor: Prof. Harold Chiang. Materials and sources: Harold's handwritten notes.

Question. Can we identify the Selection bias $\mathbb{E}[u_i|X_i'\gamma + v_i \geq 0]$?
Answer. We need more assumptions.

Theorem 1.1. Under the above setup, if in addition, we assume joint normality of (u_i, v_i):

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \right), \tag{1.6}$$

where σ_v^2 is normalized to 1. Then,

$$\mathbb{E}\left[u_i|X_i'\gamma + v_i \geq 0\right] = \sigma_{uv}\frac{\phi(X_i'\gamma)}{\Phi(X_i'\gamma)}, \tag{1.7}$$

where $\phi(\cdot)$ is the PDF and $\Phi(\cdot)$ is the CDF of $\mathcal{N}(0,1)$.

Remark. We denote $\boldsymbol{\lambda}(X_i'\gamma) \equiv \frac{\phi(X_i'\gamma)}{\Phi(X_i'\gamma)}$, the "inverse Mills Ratio." Note that $\phi'(t) = -t\phi(t)$.

Proof. By Property of bivariate Normal distribution, we have

$$u_i = \frac{\sigma_{uv}}{\sigma_v^2} \cdot v_i + \varepsilon_i = \sigma_{uv} v_i + \varepsilon_i, \text{ for some } \varepsilon_i \perp\!\!\!\perp v_i (\varepsilon_i \perp\!\!\!\perp (X_i, v_i)) \tag{1.8}$$

Then,

$$\begin{aligned}
\mathbb{E}\left[u_i|X_i'\gamma + v_i \geq 0\right] &= \mathbb{E}\left[\sigma_{uv} v_i + \varepsilon_i|X_i'\gamma + v_i \geq 0\right] && (1.9)\\
&= \mathbb{E}\left[\sigma_{uv} v_i|X_i'\gamma + v_i \geq 0\right] + \mathbb{E}\left[\varepsilon_i|X_i'\gamma + v_i \geq 0\right] && (1.10)\\
&= \sigma_{uv}\mathbb{E}\left[v_i|X_i'\gamma + v_i \geq 0\right] + 0 && (1.11)\\
&= \sigma_{uv}\mathbb{E}\left[v_i|v_i \geq -X_i'\gamma\right] && (1.12)\\
&= \sigma_{uv}\left[\frac{1}{1-\Phi(-X_i'\gamma)}\int_{-X_i'\gamma}^{\infty} t\phi(t)dt\right] && (1.13)\\
&= \sigma_{uv}\left[\frac{1}{\Phi(X_i'\gamma)}\int_{-X_i'\gamma}^{\infty} \left[\frac{-d(\phi(t))}{dt}\right]dt\right] \longleftarrow \text{by symmetry of } \Phi && (1.14)\\
&= \sigma_{uv}\left[\frac{1}{\Phi(X_i'\gamma)}\int_{-\infty}^{X_i'\gamma} \left[\frac{d(\phi(t))}{dt}\right]dt\right] && (1.15)\\
&= \sigma_{uv}\frac{1}{\Phi(X_i'\gamma)}\,\phi(t)\Big|_{-\infty}^{X_i'\gamma} = \sigma_{uv}\frac{\phi(X_i'\gamma)}{\Phi(X_i'\gamma)} && (1.16)
\end{aligned}$$

□

2 Heckit Estimator with 2-step Estimation

Summary. From Equation (1.7), we obtain a nice expression for Selection bias:

$$\mathbb{E}\left[u_i | X_i'\gamma + v_i \geq 0\right] = \sigma_{uv} \frac{\phi(X_i'\gamma)}{\Phi(X_i'\gamma)}. \tag{2.1}$$

Let's denote $\delta \equiv \sigma_{uv}$ and define parameters $(\beta', \gamma', \delta')'$. Construct $D_i = \mathbb{1}\{X_i'\gamma + v_i \geq 0\}$ $\implies$ CCP with joint normality: $\mathbb{P}(D_i = 1|X_i) = \Phi(X_i'\gamma)$, where $\Phi(\cdot)$ is the CDF of Probit. Since we do not know "γ", we use the following **2-step estimation**:

① Estimate $\hat{\gamma}$ by Probit (MLE): $D_i = \mathbb{1}\{X_i'\gamma + v_i \geq 0\}$, get $\hat{\gamma}$
$\implies$ construct $\hat{\lambda}_i(X_i'\hat{\gamma})$

② Regress Y_i on $(X_i, \hat{\lambda}_i(X_i'\hat{\gamma})) \implies$ recover estimates $(\hat{\beta}', \hat{\delta}')'$:

$$\begin{aligned}
\mathbb{E}[Y_i | X_i'\gamma + v_i \geq 0] &= m(X_i; \beta) + \mathbb{E}[u_i | X_i'\gamma + v_i \geq 0] && (2.2)\\
&= m(X_i; \beta) + \delta\hat{\lambda}(X_i'\hat{\gamma}) && (2.3)
\end{aligned}$$

3 General 2-step Estimation

Motivation. The Sample Selection Model is a special case of 2-step Estimation with a *finite* dimensional first-stage nuisance parameter.

Definition 3.1 (2-step Estimation)**.** Suppose we have two unknown parameters (θ_0, γ_0) with the following system of moment equations m & g:

$$\begin{cases} \mathbb{E}\left[m(W_i; \gamma_0)\right] = 0, \text{ where } \gamma_0 \text{ is 'nuisance' parameter} \\ \mathbb{E}\left[g(W_i; \theta_0, \gamma_0)\right] = 0, \text{ where } \theta_0 \text{ is 'parameter of interest'} \end{cases} \tag{3.1}$$

⊛ Note that g has same dim as θ, and m has same dim as γ.
We can estimate $\hat{\theta}$ by 2-step Estimation:

① Estimate $\hat{\gamma}$ by solving sample analog $\frac{1}{n}\sum_{i=1}^{n} m(W_i; \hat{\gamma}) = 0$

② Use this $\hat{\gamma}$ to estimate $\hat{\theta}$ by solving sample analog $\frac{1}{n}\sum_{i=1}^{n} g(W_i; \hat{\theta}, \hat{\gamma}) = 0$

Example 3.1 (2SLS)**.** If # of endogeneous variables matches # of exogeneous variables, then 2SLS is well-behaved.

Example 3.2 (Sample Selection Model)**.** We can apply the 2-step Estimation for Sample selection model:

① Get $\hat{\gamma}$: by Probit of $\boldsymbol{D}$ on $\boldsymbol{X}$, we have:

$$m(Z;\gamma) = \left[\boldsymbol{D}\frac{\phi(\boldsymbol{X}'\gamma)}{\Phi(\boldsymbol{X}'\gamma)} - (1-\boldsymbol{D})\frac{\phi(\boldsymbol{X}'\gamma)}{1-\Phi(\boldsymbol{X}'\gamma)}\right]\boldsymbol{X} \tag{3.2}$$

$\Longrightarrow$ construct $\boldsymbol{\lambda}(\boldsymbol{X}'\hat{\gamma}) = \frac{\phi(\boldsymbol{X}'\hat{\gamma})}{\Phi(\boldsymbol{X}'\hat{\gamma})}$

② OLS of $\boldsymbol{Y}$ on $(\boldsymbol{X},\ \boldsymbol{\lambda}(\boldsymbol{X}'\hat{\gamma}))$ using only $\boldsymbol{D}=1$, we get:

$$g(Z;\theta,\gamma) = \boldsymbol{D}\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{\lambda}(\boldsymbol{X}'\gamma)\end{bmatrix}\left(\boldsymbol{Y} - \boldsymbol{X}'\beta - \delta\boldsymbol{\lambda}(\boldsymbol{X}'\gamma)\right),\ \text{where } \theta = \begin{pmatrix}\beta\\ \delta\end{pmatrix} \tag{3.3}$$

Question. How to do inference?

Answer. We can stack g & m to form $\tilde{g}(Z;\theta,\gamma) = \begin{bmatrix}m(Z;\gamma)\\ g(Z;\theta,\gamma)\end{bmatrix}$ Then, we view it as a GMM and solve:

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{g}(Z_i;\hat{\theta},\hat{\gamma}) = 0 \tag{3.4}$$

References

Hansen, B. E. (2022). Econometrics. Princeton University Press. https://users.ssc.wisc.edu/~bhansen/econometrics/