HW3
Labor Economics

Chia-wei, Chen*

October 14, 2022

1 Normalization of the Selection Equation

1.1 Distribution of $F_U(U)$

We consider the CDF of the CDF, that is,

> How much of the U's will the CDF of U be less than y?

Apparently this is exactly the definition of a CDF, hence this number is trivially y, and therefore we know

$$\Pr(F_U(U) \leq y) = y$$

, which is a uniform distribution between $[0, 1]$

1.2 $\mathscr{P}(z) = F_U(v(z))$ [1]

The original definition of selection is

$$D = \mathbb{1}(U \leq v(Z)) = \mathbb{1}(\tilde{U} \leq F_U(v(Z)))$$

where $\tilde{U} \equiv F_U(U)$. This holds because CDF is a increasing function.

Therefore inside the propensity score we substitute $D = 1$ with the condition

$$F_U(U) \leq F_U(v(Z))$$

*R10323045

[1] I denote the propensity score as $\mathscr{P}$.

Hence

$$\mathscr{P}(z) = \Pr(D=1 \mid Z=z) = \Pr(F_U(U) \leq F_U(v(Z)) \mid Z=z) = \Pr(F_U(U) \leq F_U(v(z)))$$

Because $F_U(U)$ is uniformly distributed, we conclude that

$$\mathscr{P}(z) = F_U(v(z))$$

1.3 Show that $D = \mathbb{1}\{\tilde{U} \leq \mathscr{P}(Z)\}$

A direct substitution should do the work.

$$D = \mathbb{1}(U \leq v(Z)) = \mathbb{1}(\tilde{U} \leq F_U(v(Z))) = \mathbb{1}(\tilde{U} \leq p(Z))$$

2 Derivation of the Weights for LATE

$$\text{LATE}_{z'}^{z} = \frac{E[Y \mid Z=z] - E\left[Y \mid z=z'\right]}{E[D \mid Z=z] - E\left[D \mid z=z'\right]} = \int_0^1 \text{MTE}(u) \times \frac{1\left\{u \in \left[p\left(z'\right), p(z)\right]\right\}}{p(z) - p\left(z'\right)} du$$

2.1

Proof. First, note that $\mathscr{P}(z) = \Pr(D=1 \mid Z=z)$, which is the proportion of observations that gets the treatment given the state of instrument z, which will be $\mathbb{E}(D \mid Z=z)$.

Accordingly, $\mathbb{E}(D \mid Z=z) - \mathbb{E}(D \mid Z=z') = \mathscr{P}(z) - \mathscr{P}(z')$, thus corresponds to the denominator of the weight. □

2.2

Show that $\mathbb{E}[Y \mid Z=z] = \mathbb{E}[Y_1 \mid U \leq \mathscr{P}(Z)]\mathscr{P}(z) + \mathbb{E}[Y_0 \mid U > \mathscr{P}(Z)](1 - \mathscr{P}(z))$

Proof. We start by expression y in the potential outcome framework

$$\begin{aligned}\mathbb{E}[Y \mid Z=z] &= \mathbb{E}[Y_1 D + Y_0(1-D) \mid Z=z] \\ &= \mathbb{E}[Y_1 D \mid Z=z] + \mathbb{E}[Y_0(1-D) \mid Z=z]\end{aligned}$$

The probability of $D=1$ given z is the propensity score $\mathscr{P}(z) = \Pr(D=1 \mid Z=z)$, and $D=1$ implies $\tilde{U} \leq \mathscr{P}(Z)$. Together we get

$$= \mathbb{E}[Y_1 \mid \tilde{U} \leq \mathscr{P}(z)]\mathscr{P}(z) + \mathbb{E}[Y_0 \mid \tilde{U} > \mathscr{P}(z)](1 - \mathscr{P}(z)) \tag{1}$$

Where $1 - \mathscr{P}(z)$ is the probability of not choosing $D = 1$, which is the probability of choosing $D = 0$ □

2.3

Show that

$$\mathbb{E}[Y \mid Z = z] = \int_0^{\mathscr{P}(z)} \mathbb{E}[Y_1 \mid \tilde{U} = u]du + \int_{\mathscr{P}(z)}^1 \mathbb{E}[Y_0 \mid \tilde{U} = u]du$$

Proof. Let us start with $\mathbb{E}[Y_1 \mid \tilde{U} \leq \mathscr{P}(z)]$ in Eq. (1). The condition is a range, hence we expand it into

$$\mathbb{E}[Y_1 \mid \tilde{U} \leq \mathscr{P}(z)] = \int_0^1 \frac{\mathbb{E}[Y_1 \mid \tilde{U} = u]}{\Pr(\tilde{U} \leq \mathscr{P}(z))} \mathbb{1}\{u \leq \mathscr{P}(z)\} dF_{\tilde{U}}(u)$$

Since $\tilde{U} \sim \text{Unif}[0, 1]$, $\Pr(\tilde{U} \leq \mathscr{P}(z)) = \mathscr{P}(z)$, and $dF_{\tilde{U}}(u) = 1du$. The above term simplifies to

$$\frac{1}{\mathscr{P}(z)} \int_0^{\mathscr{P}(z)} \mathbb{E}[Y_1 \mid \tilde{U} = u]du$$

Similarly,

$$\mathbb{E}[Y_0 \mid \tilde{U} > \mathscr{P}(z)] = \frac{1}{1 - \mathscr{P}(z)} \int_{\mathscr{P}(z)}^1 \mathbb{E}[Y_0 \mid \tilde{U} = u]du$$

Substitute into Eq. (1), we get

$$\mathbb{E}[Y \mid Z = z] = \int_0^{\mathscr{P}(z)} \mathbb{E}[Y_1 \mid \tilde{U} = u]du + \int_{\mathscr{P}(z)}^1 \mathbb{E}[Y_0 \mid \tilde{U} = u]du$$

□

2.4

Proof.

$$\int_0^{\mathscr{P}(z)} \mathbb{E}[Y_1 \mid \tilde{U} = u]du + \int_{\mathscr{P}(z)}^1 \mathbb{E}[Y_0 \mid \tilde{U} = u]du - \int_0^{\mathscr{P}(z')} \mathbb{E}[Y_1 \mid \tilde{U} = u]du - \int_{\mathscr{P}(z')}^1 \mathbb{E}[Y_0 \mid \tilde{U} = u]du$$
$$= \int_{p(z')}^{p(z)} \mathbb{E}[Y_1 \mid \tilde{U} = u]du - \int_{p(z')}^{p(z)} \mathbb{E}[Y_0 \mid \tilde{U} = u]du$$
$$= \int_{p(z')}^{p(z)} \mathbb{E}[Y_1 - Y_0 \mid \tilde{U} = u]du$$
$$= \int_{p(z')}^{p(z)} \mathrm{MTE}(u)du$$

□

3 Policy Relevance Treatment Effect

3.1 ATE

The ATE measures

$$\mathbb{E}[Y_1 - Y_0]$$

It is the expected difference of future average earning on whether one attends college or not.

3.2 ATT

The ATT measures

$$\mathbb{E}[Y_1 - Y_0 \mid D = 1]$$

It is the expected difference of future average earning for people that attended college. This is a what if question: *What will be a bachelors' future earning if he doesn't go to college?*

3.3 PRTE

Although the treatment effect mentioned above tells us some aspects of potential outcome, it is not really useful for policy makers. For example, given a ATE, a consultant tell the policy makers *This is the average different of earning, but I can't tell you whether changing the tuition makes any different.* Similarly, the ATT tell the policy maker *I don't know whether this policy make people tend to attend college more or not, but if they already attend college, this will be how much they earn.*

The policy relevant treatment effect, on the other hand, considers together the difference in the outcome we are interested in (Y) as well as the change in the decision making process according to a policy driven channel ($Z*$ changing D).

With PRTE, we can estimate the effect of conducting a policy.

3.4 Relationship with LATE

The local average treatment effect is defined as

$$\text{LATE} = \mathbb{E}[Y_1 - Y - 0 \mid D_z = z, D_{z*} = z*]$$

and its estimation

$$\widehat{LATE} = \frac{\mathbb{E}[Y \mid Z = z*] - \mathbb{E}[Y \mid Z = z]}{\mathbb{E}[D \mid Z = z*] - \mathbb{E}[D \mid Z = z]}$$

Notice that the only different with PRTE is the condition. Requiring $\mathbb{E}(Y*) = \mathbb{E}[Y \mid Z = z*]$ and vice versa is the key for the two to match. Intuitively, the PRTE will be equivalent to LATE if the samples that we consider are the ones that given a change in policy, it will definitely react to it.

4 Arellano-Bond

$$Y_{it} = \rho Y_{it-1} + \delta_i + \epsilon_{it}$$

with $Cov(\epsilon_{it}, Y_{is}) = 0 \quad \forall s \leq t-1$

4.1 Show the inconsistency of FE estimator

We first demean the entire model

$$Y_{it} - \bar{Y}_i = \rho(Y_{it-1} - \bar{Y}_{i,-1}) + (\epsilon_{it} - \bar{\epsilon}_i)$$

This is an OLS with no constant term. Ideally we can extract ρ by regressing $Y_{it} - \bar{Y}_i$ on $(Y_{it-1} - \bar{Y}_{i,-1})$, but there exist correlation between the error term and the regressor, therefore it is doomed to be inconsistent (if T is not big enough).

Proof.

$$\begin{aligned} &Cov((Y_{it-1} - \bar{Y}_{i,-1}), (\epsilon_{it} - \bar{\epsilon_i})) \\ &= Cov(Y_{it-1}, \epsilon_{it}) - Cov(Y_{it-1}, \bar{\epsilon_i}) - Cov(\bar{Y}_{i,-1}, \epsilon_{it}) + Cov(\bar{Y}_{i,-1}, \bar{\epsilon_i}) \end{aligned} \tag{2}$$

Note that

$$\begin{aligned} Cov(Y_{it-1}, \bar{\epsilon_i}) &= Cov(Y_{it-1}, \frac{1}{T}\sum_{0}^{T} \epsilon_{it}) \\ &= \frac{1}{T} Cov(Y_{it-1}, \epsilon_{it-1}) \neq 0 \end{aligned}$$

The last equation comes from the assumption that $Cov(\epsilon_{it}, Y_{is}) = 0 \quad \forall s \leq t-1$

Other terms in Eq. (2) are zero according to the assumption as well.

Therefore we prove that as long as T is finite, the fixed effect estimation for ρ is inconsistent.

□

.

4.2 Taking first difference

$$\begin{array}{rcl} Y_{it} & = & \rho Y_{it-1} + \delta_i + \epsilon_{it} \\ -)Y_{it-1} & = & \rho Y_{it-2} + \delta_i + \epsilon_{it-1} \\ \hline \\ Y_{it} - Y_{it-1} & = & \rho(Y_{it-1} - Y_{it-2}) + (\epsilon_{it} - \epsilon_{it-1}) \end{array}$$

4.3 OLS on first difference

The correlation between the error term and the regressor still remains.

Proof.

$$\begin{aligned} &Cov((Y_{it-1} - Y_{it-2}), (\epsilon_{it} - \epsilon_{it-1})) = ... \\ &= Cov(Y_{it-1}, \epsilon_{it-1}) \neq 0 \end{aligned}$$

□

We can, however, choose an instrumental variable such that it correlates with y_{it-1} but not y_{it}. The second lag term will Y_{it-2} be a good choice.