

1 Identification

1.1 Heuristic Identification

1. “We don’t have enough sample size to identify the causal effects of the problem.”: As long as the identification set is singleton, we can identify the causal effects of the problem regardless of the sample size. However, smaller sample size might indeed make it harder for researchers to conclude “strong” identification, so researchers might end up with a less conclusive identification.
2. “We don’t have a good identification strategy so I need to use a structural model.”: It is not the case that “the more things you estimate, the better your identification is.” A better strategy might be finding more useful observables for your identification process.
3. “Because I have a structural model, I don’t need to think about identification.”: Even if you have a structural model, you are still trying to figure out the causality. In this case, identification does matter so much.
4. “Because I can use the maximum likelihood estimator, I can identify that.”: Generally, MLE is about “estimating” but not “identifying”. Estimations from MLE do not guarantee identification.

1.2 Identification of OLS

Proof. We prove that β in this simple OLS model is identified by contradiction. So first, suppose that there exists $\beta^* \neq \beta$ belonging to the same identification set of β , i.e.

$$\beta^* x_i + \epsilon_i = \beta x_i + \epsilon_i$$

which implies $\beta^* = \beta (\Rightarrow \Leftarrow)$

□

1.3 Identification of a Factor Model

1. The only observable is $\{y_{it}\}$, so we start from it.

$$\begin{aligned} Cov(y_{it}, y_{it-1}) &= Cov(v_{it} + \epsilon_{it}, v_{it-1} + \epsilon_{it-1}) \\ &= Cov(\rho v_{it-1} + \zeta_{it} + \epsilon_{it}, v_{it-1} + \epsilon_{it-1}) & (1) \\ &= \rho Cov(v_{it-1}, v_{it-1}) & (2) \\ &= \rho \sigma_v^2 & (3) \end{aligned}$$

Note that from (1) to (2) we use the fact that ϵ_{it} and ζ_{it} are both i.i.d. across individuals, over time, and of each other.

Similiarly,

$$\begin{aligned}
 Cov(y_{it}, y_{it-2}) &= Cov(v_{it} + \epsilon_{it}, v_{it-2} + \epsilon_{it-2}) \\
 &= Cov(\rho(v_{it-2} + \zeta_{it-1}) + \zeta_{it} + \epsilon_{it}, v_{it-2} + \epsilon_{it-2}) \\
 &= Cov(\rho^2 v_{it-2} + \rho \zeta_{it-1} + \zeta_{it} + \epsilon_{it}, v_{it-2} + \epsilon_{it-2}) \quad (4) \\
 &= \rho^2 Cov(v_{it-2}, v_{it-2}) \quad (5) \\
 &= \rho^2 \sigma_v^2
 \end{aligned}$$

Again from (4) to (5) we use the fact that ϵ_{it} and ζ_{it} are both i.i.d. across individuals, over time, and of each other. Therefore, ρ is identified since

$$\rho = \frac{\rho^2 \sigma_v^2}{\rho \sigma_v^2} = \frac{Cov(y_{it}, y_{it-2})}{Cov(y_{it}, y_{it-1})}$$

□

2. Now we have already identified ρ , so we can think of using it to identify other parameters. First, notice that

$$Var(y_{it}) = Var(v_{it} + \epsilon_{it}) = \sigma_v^2 + \sigma_\epsilon^2$$

so we can think of identifying σ_ϵ^2 by first identifying σ_v^2 . By equation (3) we can derive that

$$\sigma_v^2 = \frac{Cov(y_{it}, y_{it-1})}{\rho}$$

so σ_v^2 is identified. Then, σ_ϵ^2 can be identified by

$$\sigma_\epsilon^2 = Var(y_{it}) - \sigma_v^2$$

□

3. Now we have already identified ρ , σ_v^2 , and σ_ϵ^2 , so we can think of using them to identify other parameters. First, notice that

$$\begin{aligned}
 Var(v_{it}) &= Var(\rho v_{it-1} + \zeta_{it}) = \rho^2 \sigma_v^2 + \sigma_\zeta^2 \\
 \Rightarrow \sigma_v^2 &= \rho^2 \sigma_v^2 + \sigma_\zeta^2 \\
 \Rightarrow (1 - \rho^2) \sigma_v^2 &= \sigma_\zeta^2 \\
 \Rightarrow \sigma_v^2 &= \frac{\sigma_\zeta^2}{1 - \rho^2}
 \end{aligned}$$

σ_ζ^2 is thus identified.

□

4. We can calculate $Var(y_{it})$, $Cov(y_{it}, y_{it-1})$, and $Cov(y_{it}, y_{it-2})$. Then follow the precedures presented above, we can estimate these parameters.

1.4 Simulation of MLE

1. By the additive property of normal distribution,

$$y_i \sim N(0, \sigma_1^2 + \sigma_2^2)$$

so the probability density function (PDF) of y is

$$f(y) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2 + \sigma_2^2}} e^{-\frac{1}{2}\left(\frac{y}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)^2}$$

Therefore, the likelihood function is

$$\mathcal{L}(\sigma_1^2 + \sigma_2^2) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2 + \sigma_2^2}} e^{-\frac{1}{2}\left(\frac{y_i}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)^2}$$

Moreover, the log-likelihood function is

$$\begin{aligned} \ell(\sigma_1^2 + \sigma_2^2) &= \sum_{i=1}^N \ln \left[\frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2 + \sigma_2^2}} e^{-\frac{1}{2}\left(\frac{y_i}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)^2} \right] \\ &= \sum_{i=1}^N \ln \left[\frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2 + \sigma_2^2}} \right] + \sum_{i=1}^N \left[-\frac{1}{2} \left(\frac{y_i}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right)^2 \right] \\ &= -\frac{N}{2} \ln(\sigma_1^2 + \sigma_2^2) - \frac{N}{2} \ln 2\pi - \frac{1}{\sigma_1^2 + \sigma_2^2} \sum_{i=1}^N y_i^2 \end{aligned}$$

2 Potential Outcome Framework

1. With exact potential outcome notations in class, we can rewrite the Roy model into

$$w_i = D_i \cdot w_{1i} + (1 - D_i) \cdot w_{0i}$$

where w_i is often denoted by Y_i , w_{1i} is often denoted by $Y_i(1)$, and w_{0i} is often denoted by $Y_i(0)$.

2. As explained above,

- $Y_i(0)$ is w_{0i} , which means the potential outcome (wage) without treatment (not to immigrate).
- $Y_i(1)$ is w_{1i} , which means the potential outcome (wage) with treatment (immigrate).

3. D_i means the state of being treated (immigrate) or not.

More specifically, D_i is a dummy that takes value of 1 when $w_{1i} > w_{0i}$, i.e. $Y_i(1) > Y_i(0)$ and takes value of 0 when $w_{1i} < w_{0i}$, i.e. $Y_i(1) < Y_i(0)$.

3 Control for Observables

3.1 Rosenbaum and Rubin

3.2 Propensity Score

- 1.