National Taiwan University **Xiang Jyun, Jhang**
ECON 5211 Labor Economics I **Homework 1**
Prof. Kuan Ming, Chen **March 24, 2024**

1 Identification

1.1 Heuristic Identification

1. *"We don't have enough sample size to identify the causal effects of the problem."*

   The statement is false in most cases. Since the identification is to find the parameters underlying the data generating process, sample size has nothing to do with it. However, if the sample size is very small, then the model might not be able to sucessfully identify the casual effect even if it is identifiable.

2. *"We don't have a good identification strategy so I need to use a structural model."*

   Not having a good identification strategy can be interpreted as "Have identification set but lack method to estimate it". In this case, turning to the structural model can be a good idea.

3. *"Because I have a structural model, I don't need to think about identification."*

   The statement is false. When using structural model, identification is as crucial as reduced-form model since it helps us to understand the exact parameters underlying the data generating process.

4. *"Because I can use the maximum likelihood estimator, I can identify that."*

   The statement is false and it fails to distinguish the order between identification and estimation. Only after have a correct identification can we start to discuss estimation.

1.2 Identification of OLS

Given the OLS regression model

$$y_i = \beta x_i + \epsilon_i$$

suppose the identification set is not a singleton, that is

$$\beta x_i = \beta^* x_i, \quad \text{for all } x_i$$

hence $\beta = \beta^*$ and we have a contradiction. So the identification set is a singleton and thus β is identified.

1.3 Identification of a Factor Model

1. **Show that ρ is identified**

   We first consider the covariance between y_{it} and y_{it-1}. Given that ν, ε and ζ are mutually independent, we have

$$\begin{aligned}
\mathrm{Cov}(y_{it}, y_{it-1}) &= \mathrm{Cov}(\nu_{it} + \varepsilon_{it}, \nu_{it-1} + \varepsilon_{it-1}) \\
&= \mathrm{Cov}(\rho\nu_{it-1} + \zeta_{it} + \varepsilon_{it}, \nu_{it-1} + \varepsilon_{it-1}) \\
&= \rho\sigma_\nu^2 \qquad (1)
\end{aligned}$$

Following the same style, now we turn to consider covariance between y_{it} and y_{it-2}

$$\begin{aligned}
\mathrm{Cov}(y_{it}, y_{it-2}) &= \mathrm{Cov}(\nu_{it} + \varepsilon_{it}, \nu_{it-2} + \varepsilon_{it-2}) \\
&= \mathrm{Cov}(\rho\nu_{it-1} + \zeta_{it} + \varepsilon_{it}, \nu_{it-2} + \varepsilon_{it-2}) \\
&= \mathrm{Cov}(\rho^2\nu_{it-2} + \rho\zeta_{it-1} + \zeta_{it} + \varepsilon_{it}, \nu_{it-2} + \varepsilon_{it-2}) \\
&= \rho^2\sigma_\nu^2
\end{aligned}$$

As a consequence, ρ is identified by

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

2. **Show that σ_ϵ^2 is identified**

   Note that

   $$\mathrm{Var}(y_{it}) = \mathrm{Var}(\nu_{it} + \varepsilon_{it}) = \sigma_\nu^2 + \sigma_\epsilon^2$$

   Given we have identified ρ and equation 1, we can identify σ_ϵ^2 as well.

3. **Show that σ_ζ^2 is identified**

   Note that

   $$\mathrm{Var}(\nu_{it}) = \mathrm{Var}(\rho\nu_{it-1} + \zeta_{it}) = \rho^2\sigma_\nu^2 + \sigma_\zeta^2$$

   Given we have shown both ρ and σ_ν^2 are identified, we can identify σ_ζ^2 as well.

4. **How do we estimate these parameters? Write down an estimator**

   By analogy principle, we can construct the following estimators respectively

   $$\begin{aligned}
   \hat{\rho} &= \frac{\sum_{i=1}^N (y_{it} - \bar{y}_t)(y_{it-2} - \bar{y}_{t-2})}{\sum_{i=1}^N (y_{it} - \bar{y}_t)(y_{it-1} - \bar{y}_{t-1})} \\
   \hat{\sigma}_\nu^2 &= \frac{1}{\hat{\rho}N}\sum_{i=1}^N (y_{it} - \bar{y}_t)(y_{it-1} - \bar{y}_{t-1}) \\
   \hat{\sigma}_\varepsilon^2 &= \frac{1}{N}\sum_{i=1}^N (y_{it} - \bar{y}_t)^2 - \hat{\sigma}_\nu^2 \\
   \hat{\sigma}_\zeta^2 &= (1 - \hat{\rho}^2)\hat{\sigma}_\nu^2
   \end{aligned}$$

1.4 Simulation of MLE

1. **Write down the likelihood function**

   Note that $y_i = \epsilon_i^1 + \epsilon_i^2$ is the addition of two normal distributions, which also follows the normal distribution with mean 0 and variance $\sigma_1^2 + \sigma_2^2$, therefore we can list the pdf of y_i

   $$f(y_i; \sigma_1^2, \sigma_2^2) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp(-\frac{y_i^2}{2(\sigma_1^2 + \sigma_2^2)}), \quad \text{where } y_i \in \mathcal{R}$$

   Given the identical and independent samples, we can form the likelihood function as

   $$\mathcal{L}(\sigma_1^2, \sigma_2^2; y_1, y_2, \ldots, y_N) = \prod_{i=1}^{N} f(y_i; \sigma_1^2, \sigma_2^2)$$

   To simplify computation, we can take monotonic transformation on the likelihood function

   $$\ell(\sigma_1^2, \sigma_2^2) = \sum_{i=1}^{N} \log f(y_i; \sigma_1^2, \sigma_2^2) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma_1^2 + \sigma_2^2) - \sum_{i=1}^{N} \frac{y_i^2}{2(\sigma_1^2 + \sigma_2^2)}$$

2. **Let's draw just $N = 2$, use `optim` function to get maximum likelihood estimates**

```r
# Define the negative log-likelihood function
negative_log_lik <- function(parameter) {
    sigma_1 <- parameter[1]
    sigma_2 <- parameter[2]
    e1 <- rnorm(2, 0, sd = sigma_1)
    e2 <- rnorm(2, 0, sd = sigma_2)
    y <- e1 + e2
    log(sigma_1^2 + sigma_2^2) + log(2 * pi)
    + (1 / (2 * (sigma_1^2 + sigma_2^2))) * sum(y^2)
}
# Optimize the negative log-likelihood function
result <- optim(c(2, 2), negative_log_lik, method = "L-BFGS-B")
result$par

```

3. **Now stare at the model, can one separately identify σ_1^2 from σ_2^2? Show that it is identified or it is not identified**

   No. Since $\epsilon_i^1 \perp \epsilon_i^2$ and we can only observe $\epsilon_i^1 + \epsilon_i^2$, it's impossible to identify σ_1^2 or σ_2^2 individually given no further assumption is imposed.

4. **How about σ_{1+2}? Show that it is identified or it is not identified.**

   We can identify $\sigma_1^2 + \sigma_2^2$, which is the variance of Y_i, by using sample variance as our estimator.

5. **Does the procedure in question 2 make sense?**

   No since the σ_1^2 and σ_2^2 can not be separately identified, it didn't make sense to find MLE estimates.

2 Potential Outcome Framework

2.1 Use the exact potential outcome notation in class to write out the model

$$w_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

2.2 What is $Y_i(0)$? What is $Y_i(1)$

$Y_i(1)$ represents the wage if individual i is observed to choose to migrate; while $Y_i(0)$ represents the outcome if individual i is observed to choose not to migrate

$$\begin{cases} Y_i(1) = w_{i,1} = \mu_{i,1} + \epsilon_{i,1} \\ Y_i(0) = w_{i,0} = \mu_{i,0} + \epsilon_{i,0} \end{cases}$$

2.3 What is D_i?

$D_i \in \{0, 1\}$ reflects on whether individual i choose to migrate, and in Roy model, the criterion for individual to choose migration or not is

$$D_i = \mathbb{I}\{w_1 > w_0\}$$

3 Control for Observables

3.1 Rosenbaum and Rubin

$$\begin{aligned} P[D = 1|Y_0, Y_1, P] &= E[P[D = 1|Y_0, Y_1, P, X]|Y_0, Y_1, P] \\ &= E[P[D = 1|Y_0, Y_1, X]|Y_0, Y_1, P], && (P \text{ is a function of } X) \\ &= E[P[D = 1|X]|Y_0, Y_1, P] && (\text{CIA}) \\ &= E[P(X)|Y_0, Y_1, P] && (\text{Definition}) \\ &= E[P|Y_0, Y_1, P] \\ &= P \end{aligned}$$

3.2 Propensity Score

1. **Pick you favorite $\beta_1, \beta_2 \neq 0$. Go back to your simulation in the first homework and re-simulate the model.**

   We assume that

   $$\begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

   and X_1, X_2 also follow standard normal distribution independently

```r
library(tidyverse)
N <- 1e6
mu0 <- 0.8
mu1 <- 1
b1 <- 1
b2 <- 2
e0 <- rnorm(N, 0, 1)
```

```r
e1 <- rnorm(N, 0, 1)
x1 <- rnorm(N, 0, 1)
x2 <- rnorm(N, 0, 1)
df_exp1 <- tibble(
    e0 = e0,
    e1 = e1,
    x1 = x1,
    x2 = x2,
    w0 = mu0 + b1 * x1 + e0,
    w1 = mu1 + b1 * x1 + b2 * x2 + e1,
    d = if_else(w1 > w0, 1, 0),
    w = d * w1 + (1 - d) * w0
    )

```

2. **What is an example of X_1? Is β_1 identified?**

   X_1 can be some observable characteristics of the individual, such as gender, age, education, and so on. For example, if you are a college graduate, you are more likely to have a higher wage no matter you migrate to the host country or not. Note that the relationship between the observed outcome and the potential outcomes is

$$\begin{aligned} w &= w_0 + D(w_1 - w_0) \\ &= \mu_0 + \beta_1 X_1 + \varepsilon_0 + D(\mu_1 - \mu_0 + \beta_2 X_2 + \varepsilon_1 - \varepsilon_0) \\ &= \mu_0 + D(\mu_1 - \mu_0) + \beta_1 X_1 + \{\varepsilon_0 + D(\beta_2 X_2 + \varepsilon_1 - \varepsilon_0)\}, \end{aligned}$$

   where the first term μ_0 is similar to the intercept in the linear regression model, the treatment effect $(\mu_1 - \mu_0)$ is similar to the coefficient of the treatment indicator D, β_1 is similar to the coefficient of the covariate X_1, and $\{\varepsilon_0 + D(\beta_2 X_2 + \varepsilon_1 - \varepsilon_0)\}$, which represents the heterogeneity of the baseline untreated outcome and of the causal effect, is similar to the error term in the linear regression model. In general, since $\{\varepsilon_0 + D(\beta_2 X_2 + \varepsilon_1 - \varepsilon_0)\}$ is not necessarily orthogonal to D and X_1, a regression of Y on D and X_1 may not identify $(\mu_1 - \mu_0)$ and μ_1. The identification of $(\mu_1 - \mu_0)$ and μ_1 depends on the specific data generating process. The standard regression strategy is to include additional variables in a regression model to break the correlation between the error term and the covariates

3. **Define the propensity score using the notation set up here.**

   The propensity score is the selection probability conditional on observables; that is

$$p(X) = P(D = 1 | X_1, X_2)$$

4. **Derive the propensity score analytically**

   Suppose that X_1 and X_2 are exogenous, and people choose to migrate to the host country if and only if their potential wage in the host country is higher than that in their home country. Suppose ε_0 and ε_1 are jointly normally distributed

$$\begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01}^2 \\ \sigma_{01}^2 & \sigma_1^2 \end{pmatrix} \right)$$

Define $\nu \triangleq \mu_1 - \mu_0$, then the propensity score is

$$\begin{aligned} p(X) &= P(D = 1|X_1, X_2) \\ &= P(\mu_1 - \mu_0 + \beta_2 X_2 + \varepsilon_1 - \varepsilon_0 > 0|X_1, X_2) \\ &= P(\nu > \mu_0 - \mu_1 - \beta_2 X_2|X_1, X_2) \\ &= 1 - P\left(\nu \leq \frac{\mu_0 - \mu_1 - \beta_2 X_2}{\sigma_\nu}|X_1, X_2\right) \\ &= 1 - \Phi\left(\frac{\mu_0 - \mu_1 - \beta_2 X_2}{\sigma_\nu}\right) \end{aligned}$$

This fact uses the fact that X_1 and X_2 are exogenous. Without this assumption, the analytical derivation of the propensity score will be very complicated.

5. **Create a column in your simulated data for the estimated propensity score using the derived formula above.**

```
df_exp1 <- df_exp1 %>%
    mutate(p1 = 1 - pnorm((mu0 - mu1 - b2 * x2) / sqrt(2)))

```

6. **Use logit to estimate the propensity score.**

```
model_logit <- glm(d ~ x1 + x2, data = df_exp1, family = binomial)
df_exp1 <- df_exp1 %>%
    mutate(p2 = predict(model_logit, type = "response"))

```

7. **What is the correlation coefficient of the above two types of propensity scores?**

```
cor(df_exp1$p1, df_exp1$p2)
[1] 0.9997235

```

8. **Use both types of propensity score to conduct IPW estimates.**

```
df_exp1 <- df_exp1 %>%
    mutate(
        ipw1 = ifelse(d == 1, 1 / p1, 1 / (1 - p1)),
        ipw2 = ifelse(d == 1, 1 / p2, 1 / (1 - p2)))
lm(w ~ d, data = df_exp1, weights = df_exp1$ipw1)
lm(w ~ d, data = df_exp1, weights = df_exp1$ipw2)

```

9. **People regress w_i on X_i . Can you recover the parameters?**

   Suppose the treatment assignment mechanism is not ignorable conditional on X_1 and X_2, as in Example 1. I cannot recover the parameters by regression.

10. **Now estimate by adding "control variables." Does that work?**

    It does not work. Since the treatment assignment mechanism is not ignorable conditional on the control variables, the standard regression strategy does not work. If we can observe ε_0 and ε_1, then we can recover the parameters