

ECON5521: Homework 2 Solution

Ting-Chih Hung*
r12323020@ntu.edu.tw

March 23, 2024

Notation.

- $X \perp Y$: X is orthogonal to Y ; i.e., $E[XY] = 0$.
- $X \perp\!\!\!\perp Y$: X is independent of Y .
- $X \perp\!\!\!\perp Y \mid Z$: X is independent of Y conditional on Z .

1 IDENTIFICATION

1.1 *Heuristic Identification*

Recall the identification terminology and evaluate the following statements.

1. “We don’t have enough sample size to identify the causal effects of the problem.”
2. “We don’t have a good identification strategy so I need to use a structural model.”
3. “Because I have a structural model, I don’t need to think about identification.”
4. “Because I can use the maximum likelihood estimator, I can identify that.”

Solution. Identification problems are irrelevant to the sample size. It is about whether parameters can be identified from the data. Even with a structural model, identification is still a concern. The maximum likelihood estimator is about estimation, not identification.

1.2 *Identification of OLS*

Consider the model:

$$y_i = \beta x_i + \varepsilon_i,$$
$$\varepsilon_i \perp\!\!\!\perp X_i,$$

*All errors are my own. If you find any mistakes, please feel free to contact me.

$$\varepsilon_i \sim N(0, 1).$$

Show that β is identified with the definition we used during the lecture. (Hint: Think about OLS.)

Solution. Suppose the distribution of the observed data is described by the model $P \in \mathcal{P}$. We say β is identified if the identified set is a singleton for all $P \in \mathcal{P}$. We can show that β is identified by the contradiction. Suppose β is not identified. Then there exists $\beta_1 \neq \beta_2$ such that

$$\beta_1 x_i + \varepsilon_i = \beta_2 x_i + \varepsilon_i,$$

which is impossible.

1.3 Identification of a Factor Model

Consider the following model:

$$\begin{aligned} y_{it} &= \nu_{it} + \varepsilon_{it}, \\ \nu_{it} &= \rho \nu_{it-1} + \zeta_{it}. \end{aligned}$$

Think of y_{it} as income for individual i at time t . ε_{it} is the idiosyncratic income shock, e.g., lottery income. ν_{it} denotes the permanent component that follows an AR(1) process. For example, ν_{it} could be the unobserved productivity. The only observed data is $\{y_{it}\}$. We assume that $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ and $\zeta_{it} \sim N(0, \sigma_\zeta^2)$. Both are assumed to be iid across individuals, over time, and of each other. We assume $0 < |\rho| < 1$ so that the process is stationary.

1. Show that ρ is identified.

Solution. Since each ν is orthogonal to ε and ζ , and each ε is orthogonal to ζ ,

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

Similarly, $\text{Cov}(y_t, y_{t-2}) = \rho^2 \sigma_\nu^2$. Hence, ρ is identified by

$$\rho = \frac{\text{Cov}(y_t, y_{t-2})}{\text{Cov}(y_t, y_{t-1})}.$$

2. Show that σ_ε^2 is identified.

Solution. Note that

$$\text{Var}(y_t) = \text{Var}(\nu_{it} + \varepsilon_{it}) = \sigma_\nu^2 + \sigma_\varepsilon^2. \quad (2)$$

Since ρ is identified, σ_ν^2 is identified by Equation (1). Hence, σ_ε^2 is identified by Equation (2).

3. Show that σ_ζ^2 is identified.

Solution. Note that

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

Since ρ and σ_ν^2 are identified, σ_ζ^2 is identified.

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

Solution. By the analogy and plug-in principle, we can construct the following estimators:

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

1.4 Simulation of MLE

$$\begin{aligned} y_i &= \varepsilon_i^1 + \varepsilon_i^2, \\ \varepsilon_i^1 &\sim N(0, \sigma_1^2), \\ \varepsilon_i^2 &\sim N(0, \sigma_2^2), \\ \varepsilon_i^1 &\perp \varepsilon_i^2. \end{aligned}$$

The data is $\{y_i\}_{i=1}^N$. Assume y_i to be iid across i .

1. Write down the likelihood function.

Solution. The normal density function is

$$f(y_i | \sigma_1^2, \sigma_2^2) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left(-\frac{y_i^2}{2(\sigma_1^2 + \sigma_2^2)}\right).$$

Hence, the likelihood function is

$$\begin{aligned} L(\sigma_1^2, \sigma_2^2) &= \prod_{i=1}^N f(y_i | \sigma_1^2, \sigma_2^2) \\ &= \left[2\pi(\sigma_1^2 + \sigma_2^2)\right]^{-N/2} \exp\left(-\frac{1}{2(\sigma_1^2 + \sigma_2^2)} \sum_{i=1}^N y_i^2\right), \end{aligned}$$

and the log-likelihood function is

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

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

Solution. The R code is as follows:

```
# 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.

Solution. It is not identified. We know the observed data is described by the model $P \in \mathcal{P}$, the family of normal distributions. Given any $P \in \mathcal{P}$, we cannot find a unique σ_1^2 and σ_2^2 that generates the observed data.

4. How about $\sigma_1^2 + \sigma_2^2$? Show that it is identified or it is not identified.

Solution. It is identified. The identification is established by the fact that given any $P \in \mathcal{P}$, we can find a unique $\sigma_1^2 + \sigma_2^2$ that generates the observed data; that is, $P_{\sigma_1^2 + \sigma_2^2} = P$.

5. Does the procedure in question 2 make sense?

Solution. No, because σ_1^2 and σ_2^2 are not separately identified. The procedure in question 2 is not meaningful.

2 POTENTIAL OUTCOME FRAMEWORK

Recall our Roy model.

$$w_0 = \mu_0 + \varepsilon_0,$$

$$w_1 = \mu_1 + \varepsilon_1.$$

1. Use the exact potential outcome notations in class to write out the model.

Solution. For each individual i , we have an observed outcome Y_i , a binary treatment indicator D_i , and the potential wage $Y_i(d)$, which is the wage that would have been observed if D_i were set to d . The relationship between the observed wage and the potential wages is

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

2. What is $Y_i(0)$? What is $Y_i(1)$?

Solution. In this case, $Y_i(0)$ is the wage of individual i that would have been observed if she stayed in her home country, and $Y_i(1)$ is the wage of individual i that would have been observed if she migrated to the host country.

3. What is D_i ?

Solution. D_i is the binary treatment indicator, which is equal to 1 if individual i migrates to the host country, and 0 otherwise. In Roy model, the treatment assignment mechanism is explicitly modeled as

$$D_i = 1\{w_1 > w_0\}.$$

3 CONTROL FOR OBSERVABLES

3.1 Rosenbaum and Rubin

Write down propensity score proof (Rosenbaum and Rubin). Remark at each equality, write down the reason why that is true. Convince yourself.

Solution. The propensity score theorem is as follows.

THEOREM 3.1 (Propensity Score Theorem).

$$\{Y(0), Y(1)\} \perp\!\!\!\perp D \mid X \implies \{Y(0), Y(1)\} \perp\!\!\!\perp D \mid p(X),$$

where $p(X) := P(D = 1 \mid X)$ is the propensity score.

Proof. Here we use $Y(d)$ to denote $\{Y(0), Y(1)\}$.

$$\begin{aligned} & P(D = d \mid Y(d), e(X)) \\ &= \int_{\mathcal{X}} P(D = d \mid Y(d), e(X), X = x) P(X = x \mid Y(d), e(X)) dx \\ &= \int_{\mathcal{X}} P(D = d \mid Y(d), X = x) P(X = x \mid Y(d), e(X)) dx \\ &= \int_{\mathcal{X}} P(D = d \mid X = x) P(X = x \mid Y(d), e(X)) dx \\ &= p(X) 1\{D = 1\} + (1 - p(X)) 1\{D = 0\}, \end{aligned}$$

where the first equality is established by the law of iterated expectation, the third equality holds by $Y(d) \perp\!\!\!\perp D \mid X$, and the last equality is obtained from the definition of the propensity score. \square

In words, if the unconfoundedness assumption holds, the treatment assignment mechanism is ignorable conditional on the propensity score.

3.2 Propensity Score

$$\begin{aligned} w_0 &= \mu_0 + \beta_1 X_1 + \varepsilon_0, \\ w_1 &= \mu_1 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_1. \end{aligned}$$

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

Solution. For simplicity, I assume

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

and X_1 and X_2 are exogenous, following the standard normal distribution.

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

Solution. 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 β_1 . The identification of $\mu_1 - \mu_0$ and β_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. Consider the following examples to illustrate the identification problem.

EXAMPLE 1. Suppose the selection mechanism is

$$\begin{aligned} D &= 1\{w_1 > w_0\} \\ &= 1\{\mu_1 - \mu_0 + \beta_2 X_2 + \varepsilon_1 - \varepsilon_0 > 0\}, \end{aligned}$$

and X_1 and X_2 are exogenous;

$$X_1 = \varepsilon_{X_1}, \quad X_2 = \varepsilon_{X_2}, \quad \varepsilon_{X_1} \perp\!\!\!\perp \varepsilon_{X_2},$$

where errors are mutually independent. Then, D is correlated with X_2 , ε_0 and ε_1 . In this case, $\mu_1 - \mu_0$ are not identified, but β_1 is identified because it is orthogonal to regression errors.

EXAMPLE 2. Suppose we have an perfect experiment, where D is randomly assigned. Then, $\mu_1 - \mu_0$ and β_1 are identified. One can fit a regression model of the form

$$w = \theta_0 + \theta_1 D + \theta_2 X_1 + \theta_3 I(D \times X_2) + \varepsilon,$$

where $I(D \times X_2)$ is the interaction term between D and X_2 . The coefficient θ_1 is $\mu_1 - \mu_0$, θ_2 is β_1 , and θ_3 is β_2 . Note that in this case, the average treatment effect is not $\mu_1 - \mu_0$, but

$$\mathbb{E}[w_1 - w_0] = \mathbb{E}[\mu_1 - \mu_0 + \beta_2 X_2] = \mu_1 - \mu_0 + \beta_2 \mathbb{E}[X_2].$$

The reason why $\mu_1 - \mu_0$, β_1 , and β_2 are identified is that we have three joint interventions (i.e., D , X_1 , X_2 are exogenous). In general, most of the time, we need joint interventions to identify the causal interaction ([Hernán and Robins, 2020](#)).

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

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

$$p(X) = P(D = 1 \mid X_1, X_2).$$

4. Derive the propensity score analytically.

Solution. 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;

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

Let $\nu := \mu_1 - \mu_0$. Then, the propensity score is

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

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.

Solution. The R code is as follows:

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

6. Use logit to estimate the propensity score.

Solution. The R code is as follows:

```
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?

Solution. The correlation coefficient is about 0.99.

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

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

Solution. The R code is as follows:

```
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 D_i . Can you recover the parameters?

Solution. 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?

Solution. 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.

REFERENCES

Hernán, M. A. and Robins, J. M. (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.