Problem Set 1

Carlos Enrique Lezama Jacinto

Instituto Tecnológico Autónomo de México

The Model

Assume that the economic model generating the data for potential outcomes is of the form:

$$
\begin{aligned}
Y_1 &= \alpha + \varphi + U_1, \\
Y_0 &= \alpha + U_0,
\end{aligned}
$$

where U_1 and U_0 represent the unobservables in the potential outcome equations and φ represents the benefit associated with the treatment ($D = 1$). Individuals decide whether or not to receive the treatment ($D = 1$ or $D = 0$) based on a latent variable I:

$$
I = Z\gamma + V,
$$

where Z and V represent observables and unobservables, respectively. Thus, we can define a binary variable D indicating treatment status,

$$
D = \mathbb{1}\left[I \geq 0\right].
$$

Finally, we assume that the error terms in the model are not independent even conditioning on the observables, i.e. $U_1 \not\perp\!\!\!\perp U_0 \not\perp\!\!\!\perp V \mid Z$, but $(U_1, U_0, V) \perp\!\!\!\perp Z$.

1. Treatment Parameters

Let $\beta = Y_1 - Y_0 = \varphi + U_1 - U_0$ such that, in regression notation, $Y = \alpha + \beta D + \varepsilon$ where $\varepsilon = U_0$.

Average Treatment Effect.

$$
\begin{aligned}
ATE &= E\left[\beta\right] \\
&= E\left[\varphi + U_1 - U_0\right] \\
&= \varphi
\end{aligned}
$$

Treatment on the Treated.

$$
\begin{aligned}
TT &= E\left[\beta \mid D = 1\right] \\
&= E\left[\varphi + U_1 - U_0 \mid Z\gamma + V \geq 0\right] \\
&= \varphi + E\left[U_1 - U_0 \mid V \geq -Z\gamma\right]
\end{aligned}
$$

Treatment on the Untreated.

$$
\begin{aligned}
TUT &= E\left[\beta \mid D = 0\right] \\
&= E\left[\varphi + U_1 - U_0 \mid Z\gamma + V < 0\right] \\
&= \varphi + E\left[U_1 - U_0 \mid V < -Z\gamma\right]
\end{aligned}
$$

Marginal Treatment Effect.

$$
\begin{aligned}
MTE &= E\left[\beta \mid I = 0,\ V = v\right] \\
&= E\left[\varphi + U_1 - U_0 \mid I = 0,\ V = v\right] \\
&= \varphi + E\left[U_1 - U_0 \mid Z\gamma = -V,\ V = v\right]
\end{aligned}
$$

Instrumental Variables.

$$
\hat{\beta}_{\mathrm{IV}}(J(Z)) = \frac{\operatorname{Cov}(J(Z), Y)}{\operatorname{Cov}(J(Z), D)} \xrightarrow{p} \beta
$$

My solutions to the first problem set in Advanced Microeconometrics (ECO – 20513).

Ordinary Least Squares.

$$\hat{\beta}_{\text{OLS}} = \frac{\text{Cov}(Y, D)}{\text{Var}(D)} \implies \hat{\beta}_{\text{OLS}} = \left(D^T D\right)^{-1} D^T Y$$

Local Average Treatment Effect.

$$\begin{aligned} LATE &= E\left[\beta \mid D(z) = 0, D(z') = 1\right] \\ &= E\left[\beta \mid z\gamma < -V \leq z'\gamma\right] \\ &= \varphi + E\left[U_1 - U_0 \mid -z'\gamma \leq V < -z\gamma\right] \end{aligned}$$

2. Some closed form expressions

Suppose that the error terms in the model have the following structure:

$$\begin{aligned} U_1 &= \sigma_1 \epsilon, \\ U_0 &= \sigma_0 \epsilon, \\ V &= \sigma_V^* \epsilon, \\ \epsilon &\sim \mathcal{N}(0, 1). \end{aligned}$$

So, we can say that

$$\begin{aligned} TT &= \varphi + E\left[U_1 - U_0 \mid V \geq -Z\gamma\right] \\ &= \varphi + E\left[\epsilon(\sigma_1 - \sigma_0) \mid \epsilon \geq \frac{-Z\gamma}{\sigma_V^*}\right] \\ &= \varphi + (\sigma_1 - \sigma_0) E\left[\epsilon \mid \epsilon \geq \frac{-Z\gamma}{\sigma_V^*}\right] \\ &= \varphi + (\sigma_1 - \sigma_0) \frac{\phi\left(-Z\gamma/\sigma_V^*\right)}{1 - \Phi\left(-Z\gamma/\sigma_V^*\right)}, \end{aligned}$$

and

$$\begin{aligned} TUT &= \varphi + E\left[U_1 - U_0 \mid V < -Z\gamma\right] \\ &= \varphi + E\left[\epsilon(\sigma_1 - \sigma_0) \mid \epsilon < \frac{-Z\gamma}{\sigma_V^*}\right] \\ &= \varphi + (\sigma_1 - \sigma_0) E\left[\epsilon \mid \epsilon < \frac{-Z\gamma}{\sigma_V^*}\right] \\ &= \varphi - (\sigma_1 - \sigma_0) \frac{\phi\left(-Z\gamma/\sigma_V^*\right)}{\Phi\left(-Z\gamma/\sigma_V^*\right)} \\ &= \varphi + (\sigma_0 - \sigma_1) \frac{\phi\left(-Z\gamma/\sigma_V^*\right)}{\Phi\left(-Z\gamma/\sigma_V^*\right)}. \end{aligned}$$

3. Parametrization 1

Now, to add more structure to the problem, suppose that $Z = (1, Z_1, Z_2)$, and $\gamma = (\gamma_0, \gamma_1, \gamma_2)$. Also, suppose that

$$\begin{aligned} \gamma_0 &= 0.2, & \gamma_1 &= 0.3, & \gamma_2 &= 0.1, \\ \sigma_1 &= 0.012, & \sigma_0 &= 0.05, & \sigma_V^* &= 1, \\ \alpha &= 0.02, & \varphi &= 0.2. \end{aligned}$$

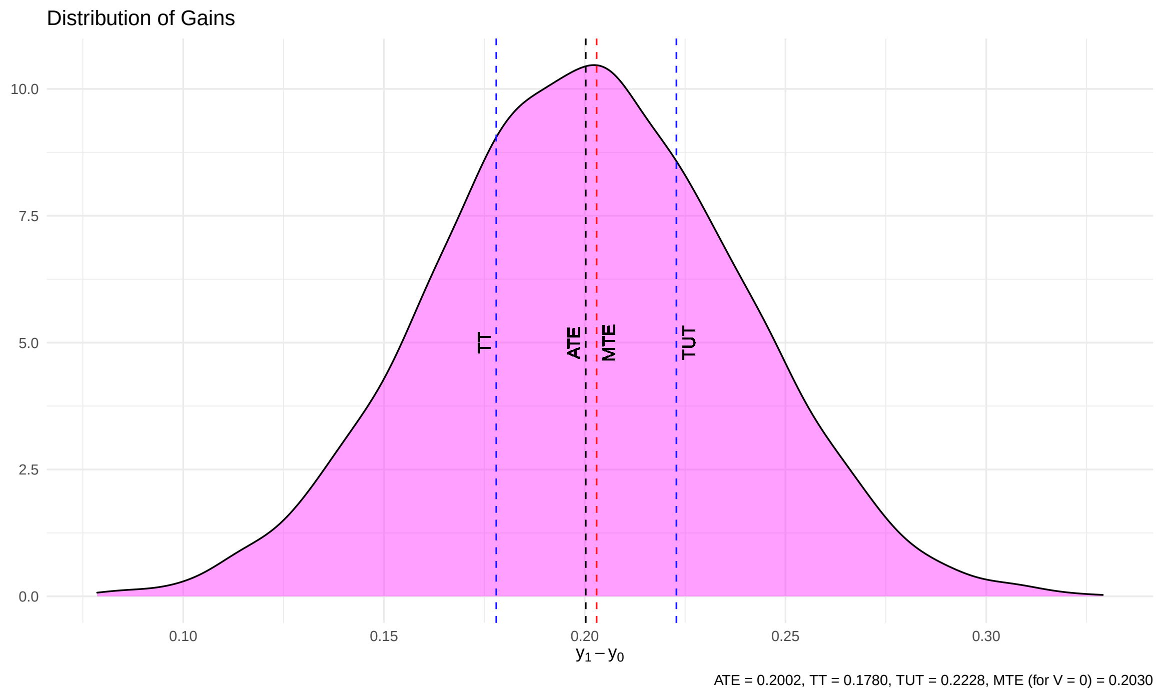
Finally,

$$\begin{aligned} Z_1 &\sim \mathcal{N}(-1, 9), \\ Z_2 &\sim \mathcal{N}(1, 9), \end{aligned}$$

where $Z_1 \perp\!\!\!\perp Z_2$. Since we only observe either Y_1 or Y_0, the econometrician observes the outcome described as follows:

$$Y = DY_1 + (1 - D)Y_0.$$

```r
set.seed(1234)
TOL <- 0.01

N <- 5000 # random sample observations

alpha <- 0.02
phi <- 0.2

g <- matrix(c(0.2, 0.3, 0.1))
sigma1 <- 0.012
sigma0 <- 0.05
sigmaV <- 1

eps <- rnorm(N)

U1 <- sigma1 * eps
U0 <- sigma0 * eps
V <- sigmaV * eps

Z <- cbind(rep(1, N),
           rnorm(N, -1, 3),
           rnorm(N, 1, 3))

I <- (Z %*% g) + V
D <- ifelse(I >= 0, 1, 0)

Y1 <- alpha + phi + U1
Y0 <- alpha + U0

Y <- (D * Y1) + ((1 - D) * Y0)

df <- data.frame('beta' = Y1 - Y0,
                 'decision' = D,
                 'y1' = Y1,
                 'y0' = Y0,
                 'net utility' = I)

ATE <- df %>%
    pull(beta) %>%
    mean()

TT <- df %>%
    filter(decision == 1) %>%
    pull(beta) %>%
    mean()

TUT <- df %>%
    filter(decision == 0) %>%
    pull(beta) %>%
    mean()

MTE <- df %>%
    filter(abs(net.utility - V) <= TOL) %>%
    pull(beta) %>%
    mean()
```



4. Parametrization 2

```
LATE <- function(param, old, new) {
    Z.old <- Z
    Z.new <- Z

    Z.old[, param] <- old
    Z.new[, param] <- new

    I.old <- (Z.old %*% g) + V
    D.old <- ifelse(I.old >= 0, 1, 0)

    I.new <- (Z.new %*% g) + V
    D.new <- ifelse(I.new >= 0, 1, 0)

    test <- ifelse((D.old == 0) & (D.new == 1), 1, 0)

    temp.df <- data.frame(
        'beta' = Y1 - Y0,
        'test' = test
    )

    r <- temp.df %>%
        filter((test == 1)) %>%
        pull(beta) %>%
        mean()

    return(r)
}

regress <- function(beta) {
    intercept <- mean(Y) - (beta * mean(D))
    intercept + (beta * D)
}
```

A. $LATE(z_1 = -2, z_1' = 1)$.

```
LATE.1 <- LATE(2, -2, 1)
```

$LATE(z_1 = -2, z_1' = 1) = 0.2051.$

B. $LATE(z_2 = 0, z_2' = 2)$**.**

```
LATE.2 <- LATE(3, 0, 2)
```

$LATE(z_2 = 0, z_2' = 2) = 0.1982.$

C. $IV(Z_1)$**.**

```
beta.IV.Z1 <- (cov(Z[, 2], Y) / cov(Z[, 2], D))[1, 1]
Y.IV.Z1 <- regress(beta.IV.Z1)
```

$\hat{\beta}_{\text{IV}}(Z_1) = 0.2045.$

D. $IV(Z_2)$**.**

```
beta.IV.Z2 <- (cov(Z[, 3], Y) / cov(Z[, 3], D))[1, 1]
Y.IV.Z2 <- regress(beta.IV.Z2)
```

$\hat{\beta}_{\text{IV}}(Z_2) = 0.206.$

E. OLS**.**

```
beta.OLS <- (cov(D, Y) / var(D))[1, 1]
Y.OLS <- regress(beta.OLS)
```

$\hat{\beta}_{\text{OLS}} = 0.237.$

Given our regression outcomes, the OLS regression shows less variance than the other estimators, but it is more biased. These results exhibit the conflict in trying to simultaneously minimize these two sources of error. Similar estimations, also imply that there is not single best optimization algorithm as it is stated in the so called *no free lunch theorem.*

5. GMM

Using $Z = (Z_1, Z_2)$ as instruments and $\hat{V}_0 = \dfrac{1}{n}\sum_{i=1}^{n} z_i' z_i$ as the weighting matrix, we can estimate our model by the generalized method of moments.

Hence, we can obtain

$$\hat{\beta}_{\text{GMM}} = \left(D'ZV_0^{-1}Z'D\right)^{-1} D'ZV_0^{-1}Z'Y.$$

Note that we can rewrite $\hat{V}_0 = \dfrac{1}{n}Z'Z$.

Thus,

```
subset.Z <- Z[, -1]

V0 <- (1 / N) * (t(subset.Z) %*% subset.Z)

beta.GMM <-
    (
        solve(t(D) %*% subset.Z %*% solve(V0) %*% t(subset.Z) %*% D) %*%
            t(D) %*% subset.Z %*% solve(V0) %*% t(subset.Z) %*% Y
    )[1, 1]

Y.GMM <- regress(beta.GMM)
```

Estimator	Estimate	Standard Error
$\hat{\beta}_{\text{GMM}}$	0.2087	0.0321
$\hat{\beta}_{\text{IV}}(Z_1)$	0.2045	0.0331
$\hat{\beta}_{\text{IV}}(Z_2)$	0.206	0.0327
$\hat{\beta}_{\text{OLS}}$	0.237	0.0288

6. Likelihood Function

We know that

$$\begin{aligned} Pr\left(D=1 \mid Z\right) &= Pr\left(\gamma Z + V \geq 0\right) \\ &= Pr\left(\frac{-Z\gamma}{\sigma_V^*} \leq \epsilon\right) \\ &= 1 - Pr\left(\epsilon \leq \frac{-Z\gamma}{\sigma_V^*}\right) \\ &= 1 - \Phi\left(\frac{-Z\gamma}{\sigma_V^*}\right), \end{aligned}$$

and

$$\begin{aligned} Pr\left(D=0 \mid Z\right) &= Pr\left(\gamma Z + V < 0\right) \\ &= Pr\left(\epsilon \leq \frac{-Z\gamma}{\sigma_V^*}\right) \\ &= \Phi\left(\frac{-Z\gamma}{\sigma_V^*}\right). \end{aligned}$$

Hence,

$$\begin{aligned} \mathcal{L} &= \prod_{d_i=0}\left[\Phi\left(\frac{-Z\gamma}{\sigma_v^*}\right)\right]\prod_{d_i=1}\left[1-\Phi\left(\frac{-Z\gamma}{\sigma_v^*}\right)\right] \\ &= \prod_i\left[\Phi\left(\frac{-Z\gamma}{\sigma_v^*}\right)\right]^{1-d_i}\left[1-\Phi\left(\frac{-Z\gamma}{\sigma_v^*}\right)\right]^{d_i}, \end{aligned}$$

and

$$\begin{aligned} \ell &= \log\mathcal{L} \\ &= \sum_i(1-d_i)\log\left(\Phi\left(\frac{-Z\gamma}{\sigma_V^*}\right)\right) \\ &+ \sum_i d_i\log\left(1-\Phi\left(\frac{-Z\gamma}{\sigma_V^*}\right)\right) \\ &= \sum_i(1-d_i)\log\left(\Phi\left(\frac{-\gamma_0 - z_1\gamma_1 - z_2\gamma_2}{\sigma_V^*}\right)\right) \\ &+ \sum_i d_i\log\left(1-\Phi\left(\frac{-\gamma_0 - z_1\gamma_1 - z_2\gamma_2}{\sigma_V^*}\right)\right). \end{aligned}$$

7. Discussion

The variance on V (σ_V^*) is not identified and so we normalize it to 1.