Introduction to Estimating Treatment Effects

Charlie Gibbons
Political Science 236

September 30, 2008

Outline

1. Ordinary Least Squares
   - Assumptions
   - Coefficients
   - Omitted variables bias

2. Treatment effects
   - Potential outcomes framework
   - Under treatment independence
   - Under conditional independence

OLS assumptions

Assumptions of ordinary least squares:

1. $Y_i = x_i'\beta + \epsilon_i$
2. X is non-stochastic
3. X is non-singular
4. $\mathbb{E}(\epsilon_i) = 0 \Rightarrow \mathbb{E}(\epsilon_i | X_i) = 0 \Rightarrow \mathbb{E}(X_i \epsilon_i) = 0$
5. $\mathbb{E}(\epsilon_i) = \sigma^2$

OLS assumptions

Assumptions of ordinary least squares:

1. $Y_i = x_i'\beta + \epsilon_i$
2. X is non-stochastic
   $X'X$ converges to a non-stochastic matrix
3. X is non-singular
   Acquire more observations
4. $\mathbb{E}(\epsilon_i) = 0 \Rightarrow \mathbb{E}(\epsilon_i | X_i) = 0 \Rightarrow \mathbb{E}(X_i \epsilon_i) = 0$
   Instrumental variables, regression discontinuity
5. $\mathbb{E}(\epsilon_i) = \sigma^2$
   Generalized least squares, clustering

OLS assumptions

Assumptions of ordinary least squares:

1. $Y_i = x_i' \beta + \epsilon_i$
2. X is non-stochastic
   $X'X$ converges to a non-stochastic matrix
3. X is non-singular
   Acquire more observations
4. $\mathbb{E}(\epsilon_i) = 0 \Rightarrow \mathbb{E}(\epsilon_i | X_i) = 0 \Rightarrow \mathbb{E}(X_i \epsilon_i) = 0$
   Instrumental variables, regression discontinuity
5. $\mathbb{E}(\epsilon_i) = \sigma^2$
   Generalized least squares, clustering

The linear model assumption is often the hardest to conquer.

Solving for OLS coefficients

Minimizing squared error

Matrix notation	Summation notation
$Y_i = x_i'\beta + \epsilon_i$	$Y_i = \sum_{k=1}^{K} x_{ik}\beta_k + \epsilon_i$
$\min_b (Y - Xb)^2$	$\min_b \sum_{i=1}^{N} \left(Y_i - \sum_{k=1}^{K} x_{ik}b_k\right)^2$
$-2X'(Y - Xb) = 0$	$-2\sum_{i=1}^{N} \left(Y_i - \sum_{k=1}^{K} x_{ik}b_k\right)x_{ij} = 0$
	$\forall\ j \ \in \{1, \ldots, K\}$

OLS in matrix form

From the minimization result:

$$\begin{aligned} X'(Y - Xb) &= 0 \\ X'Y &= X'Xb \\ b &= \left(X'X\right)^{-1} X'Y \end{aligned}$$

b is BLUE: Best Linear Unbiased Estimator

Omitted variables bias

Suppose that the true linear model is

$$Y = X\beta + Z\gamma + \epsilon$$

but we do not include Z in our regression. Hence, we estimate

$$\begin{aligned} Y &= X\beta + \tilde{\epsilon} \\ \tilde{\epsilon} &= Z\gamma + \epsilon \end{aligned}$$

yielding

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y$$

Omitted variables bias

This gives

$$
\begin{aligned}
\hat{\beta} &= \left(X'X\right)^{-1} X'Y \\
&= \left(X'X\right)^{-1} X' \left(X\beta + Z\gamma + \epsilon\right) \\
&= \left(X'X\right)^{-1} X'X\beta + \left(X'X\right)^{-1} X'Z\gamma \\
&= \beta + \underbrace{\left(X'X\right)^{-1} X'Z\gamma}_{\text{Bias}}
\end{aligned}
$$

Omitted variables bias

$\text{Bias} = (X'X)^{-1} X'Z\gamma$

When is the bias 0?

1. $\gamma = 0$: No omitted variables
2. $X'Z = 0$: No correlation between X and the omitted variables Z

Omitted variables bias

$\text{Bias} = (X'X)^{-1} X'Z\gamma$

When is the bias 0?

1. $\gamma = 0$: No omitted variables
2. $X'Z = 0$: No correlation between X and the omitted variables Z

Omitted variables bias

$$\text{Bias} = (X'X)^{-1} X'Z\gamma$$

When is the bias 0?

1. $\gamma = 0$: No omitted variables
2. $X'Z = 0$: No correlation between X and the omitted variables Z

Potential outcomes framework

There are two possible outcomes: $Y_i(1)$ if individual i undergoes treatment T and $Y_i(0)$ if he does not. T_i is an indicator of treatment status. The treatment effect for i is $\tau_i = Y_i(1) - Y_i(0)$. Hence,

$$Y_i = (1 - T_i)Y_i(0) + T_i Y_i(1) = Y_i(0) + T_i \tau_i$$

Notice that we are not assuming any error in Y_i. But there is a distribution of τ_i.

Treatment effects

The two main effects of interest are the *average treatment effect (ATE)* and the *average treatment effect for the treated (ATT).*

Treatment effects

The two main effects of interest are the *average treatment effect (ATE)* and the *average treatment effect for the treated (ATT).*

$$
\begin{aligned}
\text{ATE} &= \mathbb{E}(\tau_i) \\
\text{ATT} &= \mathbb{E}(\tau_i | \underbrace{T = 1}_{\text{Distribution of treated}}) \\
&= \mathbb{E}_1(\tau_i)
\end{aligned}
$$

Treatment effects

The two main effects of interest are the *average treatment effect (ATE)* and the *average treatment effect for the treated (ATT)*.

$$\begin{aligned} \text{ATE} &= \mathbb{E}(\tau_i) \\ \text{ATT} &= \mathbb{E}(\tau_i | \underbrace{T = 1}_{\text{Distribution of treated}}) \\ &= \mathbb{E}_1(\tau_i) \end{aligned}$$

These effects are often conditioned on X:

$$\begin{aligned} \text{ATE}(x) &= \mathbb{E}(\tau_i | X = x) \\ \text{ATT}(x) &= \mathbb{E}_1(\tau_i | X = x) \end{aligned}$$

Under treatment independence

Of course, τ_i is not observable since both outcomes cannot be realized for each individual.

If $(Y_i(0), Y_i(1)) \perp T_i$, then the distributions of outcomes in both the entire population and among the treated are the same. Additionally, covariates do not matter. Hence, the ATE and ATT will be equal. This result stems from random treatment assignment.

Under treatment independence

By the law of iterated expectations and independence,

$$\begin{aligned}
\text{ATT} &= \mathbb{E}_1\left[Y_i(1) - Y_i(0)\right] \\
&= \mathbb{E}\left[Y_i(1) - Y_i(0)\right] = \text{ATE} \\
\mathbb{E}_0(Y) &= \mathbb{E}_0\left[(1 - T_i)Y_i(0) + T_iY_i(1)\right] \\
&= \mathbb{E}_0\left[Y_i(0)\right] = \mathbb{E}\left[Y_i(0)\right] \\
\mathbb{E}_1[Y] &= \mathbb{E}_1\left[(1 - T_i)Y_i(0) + T_iY_i(1)\right] \\
&= \mathbb{E}_1\left[Y_i(1)\right] = \mathbb{E}\left[Y_i(1)\right]
\end{aligned}$$

Under treatment independence

Hence,

$$\begin{aligned}\text{ATE} &= \mathbb{E}\left[Y_i(1) - Y_i(0)\right] = \mathbb{E}\left[Y_i(1)\right] - \mathbb{E}\left[Y_i(0)\right] \\ &= \mathbb{E}_1\left[Y_i(1)\right] - \mathbb{E}_0\left[Y_i(0)\right]\end{aligned}$$

Under treatment independence

Therefore,

$$\text{ATE} = \text{ATT} = \mathbb{E}_1\left[Y_i(1)\right] - \mathbb{E}_0\left[Y_i(0)\right]$$

Under treatment independence

Therefore,

$$\text{ATE} = \text{ATT} = \mathbb{E}_1 \left[Y_i(1) \right] - \mathbb{E}_0 \left[Y_i(0) \right]$$

These parameters can be estimated by

$$\frac{1}{\#\{i\ :\ T_i = 1\}} \sum_{\{i\,:\,T_i=1\}} Y_i - \frac{1}{\#\{i\ :\ T_i = 0\}} \sum_{\{i\,:\,T_i=0\}} Y_i$$

This is just a simple difference in means between the two groups.

Under conditional independence

How can this problem be solved in the absence of random treatment assignment?

Under conditional independence

How can this problem be solved in the absence of random treatment assignment?

We need the assumption of *unconfoundedness*:

$$(Y_i(0), Y_i(1)) \perp T_i | X_i$$

Under conditional independence

This assumption intuitively states that, if we observe X, we are able to determine all the ways in which the treatment group differs from the control; the X covariates explain treatment assignment completely. For this reason, this assumption is also known as *selection on observables.*

Under conditional independence

Note that all the results shown in the previous set of slides, namely that the ATT equals the ATE, remain after conditioning on X. Specifically,

$$\text{ATE}(x) = \text{ATT}(x) = \mathbb{E}_1\left[Y|X = x\right] - \mathbb{E}_0\left[Y|X = x\right]$$

Hence, to find $ATT(x)$, find the average value of Y for those members of the treated group with $X = x$ and subtract it from the average Y for the control population with $X = x$.

Under conditional independence

If $\text{ATT}(x) = \text{ATT}(x') \; \forall \; x, x'$, then there is a *homogeneous treatment effect*. Then the $\text{ATT}(x) = \text{ATT}$.

Under conditional independence

If $\text{ATT}(x) = \text{ATT}(x') \; \forall \; x, x'$, then there is a *homogeneous treatment effect*. Then the $\text{ATT}(x) = \text{ATT}$.

If $\exists \; x \neq x' \; : \; \text{ATT}(x) \neq \text{ATT}(x')$, then there is a *heterogeneous treatment effect*. In this case, the *unconditional ATT* is an average of the $\text{ATT}(x)$ over all values of $x_i \; : \; T_i = 1$, namely the distribution of values of x_i among the treated population.

Under conditional independence

If $\text{ATT}(x) = \text{ATT}(x') \; \forall \; x, x'$, then there is a *homogeneous treatment effect*. Then the $\text{ATT}(x) = \text{ATT}$.

If $\exists \; x \neq x' \; : \; \text{ATT}(x) \neq \text{ATT}(x')$, then there is a *heterogeneous treatment effect*. In this case, the *unconditional ATT* is an average of the $\text{ATT}(x)$ over all values of $x_i \; : \; T_i = 1$, namely the distribution of values of x_i among the treated population.

In principle, we could find the ATE by taking the expectation of the ATE over all possible values of X. But, we would really be relying on our selection on observables assumption.