ECON220B Discussion Section 2

From Potential Outcome To Regression Arithmetic

Lapo Bini

PS1 Feedback

Roadmap

1. Introduction to Potential Outcomes

2. Randomized Control Trials

3. Some Unpleasant Linear Regression Arithmetic

Some Notation

- We want to express a causal statement, i.e. a comparison between two states of the world.

- An individual i could receive a treatment or not. We will denote with x_i the treatment status of the i^{th} unit: $x_i = \mathbb{1}\{treated\}$.

- Observable outcome for each individual: $y_i = x_1 y_i(1) + (1 - x_i) y_i(0)$. Objects of interest:

$$\tau_{ATE} \equiv E[y_i(1) - y_i(0)]$$
$$\tau_{ATT} \equiv E[y_i(1) - y_i(0) | x_i = 1]$$
$$\tau_{ATU} \equiv E[y_i(1) - y_i(0) | x_i = 0]$$

- **Key concept**: identification vs. estimation.

Road to Identification (1/2)

- Claim: $\tau_{ATE} = \pi\, \tau_{ATT} + (1-\pi)\tau_{ATU}$
  Proof

- We observe $E[y_i(1)|x_i = 1]$ and $E[y_i(0)|x_i = 0]$. Can we estimate the Average Treatment Effect as:

$$\tau_{ATE} = \Big(E[y_i(1)|x_i = 1] - E[y_i(0)|x_i = 0]\Big) \quad ? \quad \Longrightarrow \textbf{NO}$$

Road to Identification (2/2)

Let $E[y_i(1)|x_i=1] \equiv a$, $E[y_i(1)|x_i=0] \equiv b$, $E[y_i(0)|x_i=1] \equiv c$, $E[y_i(0)|x_i=0] \equiv d$, then:

$$\tau_{ATE} = \pi\,\tau_{ATT} + (1-\pi)\tau_{ATU}$$
$$\tau_{ATE} = \pi\,\big(a-c\big) + (1-\pi)\big(b-d\big)$$
$$\tau_{ATE} = \pi\,\big(a-c\big) + (1-\pi)\big(b-d\big) + \big(a-a\big) + \big(c-c\big) + \big(d-d\big)$$
$$\tau_{ATE} = \pi a + b - \pi b - \pi c - d + \pi d + (a-a) + (c-c) + (d-d)$$
$$\tau_{ATE} = (a-d) + \pi a + b - \pi b - \pi c - d + \pi d - a + c - c + d$$
$$\tau_{ATE} = (a-d) - (c-d) - a + \pi a + b - \pi b + c - \pi c - d + \pi d$$
$$\tau_{ATE} = (a-d) - (c-d) - (1-\pi)a + (1-\pi)b + (1-\pi)c - (1-\pi)d$$
$$\tau_{ATE} = (a-d) - (c-d) - (1-\pi)(a-c) + (1-\pi)(b-d)$$
$$\tau_{ATE} = (a-d) - (c-d) - (1-\pi)\big[(a-c) - (b-d)\big]$$
$$\tau_{ATE} = \Big(E[y_i(1)|x_i=1] - E[y_i(0)|x_i=0]\Big) \quad - \quad \underbrace{\Big(E[y_i(0)|x_i=1] - E[y_i(0)|x_i=0]\Big)}_{\text{Selection Bias}} \quad - \quad \underbrace{(1-\pi)\Big(\tau_{ATT} - \tau_{ATU}\Big)}_{\text{Heterogeneous Treatment Effect Bias}}$$

Randomized Control Trial

- We are going to assume that the treatment has been assigned to individuals independent of their potential outcome:

$$x_i \perp y_i(1), y_i(0)$$

- Implications:
  1. $E[y_i(0)|x_i = 1] - E[y_i(0)|x_i = 0] = 0$
  2. $\tau_{ATT} - \tau_{ATU} = 0$
  3. $\tau_{ATE} = E[y_i(1)|x_i = 1] - E[y_i(0)|x_i = 0]$

- Estimation: hat instead of expectation.

Regression Representation of Potential Outcome

- Remember: $\{y_1(1), \ldots, y_n(1)\}$ iid with $y_i(1) \sim (\mu_1, \sigma^2)$ we can always write a random variable into an expectation component and an error term: $y_i(1) = \mu_1 + u_i(1)$ with $u_i(1) \sim (0, \sigma^2)$.

$$
\begin{aligned}
y_i(1) &= \\
y_i(0) &= \\
y_i &=
\end{aligned}
$$

Implication Missing at Random

Given our assumption $x_i \perp y_i(1), y_i(0)$ we have:

1. $E[u_i|x_i] = 0$
2. $E[u_i x_i] = 0$
3. $E[u_i] = 0$

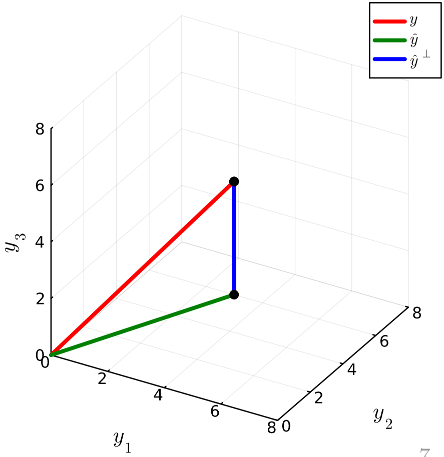
Orthogonal Projection

Orthogonal Projection $\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{y}}^{\perp}$
Given a vector space V and a vector subspace $\mathcal{M}$ there exists a unique $\hat{\mathbf{y}} \in \mathcal{M}$ such that:

$$\hat{\mathbf{y}} = \arg\min_{x \in \mathcal{M}} \|\mathbf{y} - \mathbf{x}\|^2_{L^2}$$

$\hat{\mathbf{y}}$ is the unique element characterized by $\langle \mathbf{y} - \hat{\mathbf{y}}, \mathbf{x} \rangle = 0$, which is known as **orthogonality property**.

$$\mathcal{M} \equiv \{(y_1, y_2, y_3) \in \mathbb{R}^3 : y_3 = 0\}$$



Orthogonal Projection vs Linear Regression

Orthogonal Projection $\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{y}}^{\perp}$ Given a vector space V and a vector subspace $\mathcal{M}$ there exists a unique $\hat{\mathbf{y}} \in \mathcal{M}$ such that:

$$\hat{\mathbf{y}} = \arg\min_{x \in \mathcal{M}} \|\mathbf{y} - \mathbf{x}\|^2_{L^2}$$

$\hat{\mathbf{y}}$ is the unique element characterized by $\langle \mathbf{y} - \hat{\mathbf{y}}, \mathbf{x} \rangle = 0$, which is known as **orthogonality property**.

Linear Regression $\mathbf{y} = X\beta + \mathbf{u}$ Given a $(nx1)$ vector $\mathbf{y}$ and a (nxd) matrix X there exists a unique $(dx1)$ vector of parameters β such that:

$$\beta = \arg\min_{b \in \mathbb{R}^d} E[(\mathbf{y} - Xb)^T(\mathbf{y} - Xb)]$$

β is the unique element characterized by $X'(\mathbf{y} - X\hat{\beta}) = 0$, which is known as **sample moment condition**.

OLS Matrix Form

Matrix calculus results: given $(nx1)$ vector $\mathbf{y}$, $(dx1)$ vector b, (nxn) symmetric matrix A, (nxd) matrix X, we have:

$$\frac{\partial \mathbf{y}' A \mathbf{y}}{\partial \mathbf{y}} = 2A\,\mathbf{y} \qquad \frac{\partial \mathbf{y}(b)'\mathbf{y}(b)}{\partial b} = \mathbf{y}'(b)\frac{\partial \mathbf{y}(b)}{\partial b} \qquad \frac{\partial \mathbf{Xb}}{\partial b} = X$$

$\therefore \frac{\partial}{\partial \beta} E[(\mathbf{y} - Xb)^T(\mathbf{y} - Xb)] =$

OLS Vector Form

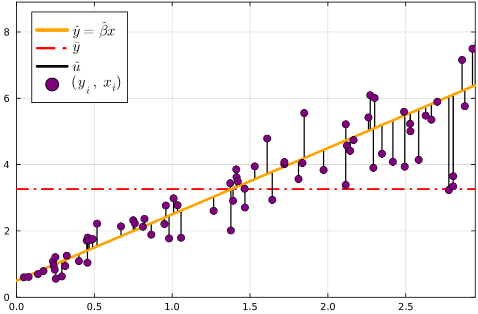
Regression: $y_i = \alpha + x_i^T \beta + u_i$ OLS: $(\alpha, \beta) = \arg\min_{(a,b)} E[(y_i - a - x_i^t b)^2]$

R^2 - Coefficient of Determination

The R^2 is a measure that indicates the proportion of the variance in the dependent variable that is explained by the independent variables in the model. Higher R^2 indicates that the regression model fits the data better.

Let's define:

- $\text{TSS} \equiv \sum_{i=1}^{n}(y_i - \bar{y})^2$
- $\text{ESS} \equiv \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$
- $\text{RSS} \equiv \sum_{i=1}^{n}(y_i - \hat{y})^2$



We want to show that:

(1) $\text{TSS} = \text{ESS} + \text{RSS}$

(2) and $R^2 \equiv 1 - \frac{\text{RSS}}{\text{TSS}}$

Projection Matrix

Given our vector space $\mathcal{V}$ and $\mathcal{M} \subseteq V$, there is an orthogonal projection $P_{\mathcal{M}} : \mathcal{V} \to \mathcal{M}$ with the property that:

$$P_{\mathcal{M}}(\mathbf{y}) = \hat{\mathbf{y}} = X\hat{\beta}$$

(1) Construct the projection matrix $P_{\mathcal{M}}(\cdot)$

(2) Show that $P_{\mathcal{M}}(\cdot)$ is symmetric, i.e. $P_{\mathcal{M}} = P'_{\mathcal{M}}$

(3) Show that $P_{\mathcal{M}}(\cdot)$ is idempotent, i.e. $P_{\mathcal{M}} \cdot P_{\mathcal{M}} = P_{\mathcal{M}}$

(4) Show that $P_{\mathcal{M}}(X) = X$

Annihilator Matrix

Given our vector space $\mathcal{V}$ and $\mathcal{M} \subseteq V$, there is an annihilator matrix $M_{\mathcal{M}} : \mathcal{V} \to \mathcal{M}^{\perp}$ with the property that:

$$M_{\mathcal{M}}(\mathbf{y}) = \hat{\mathbf{u}} = \mathbf{y} - X\hat{\beta}$$

(1) Construct the annihilator matrix $M_{\mathcal{M}}(\cdot)$

(2) Show that $M_{\mathcal{M}}(\cdot)$ is symmetric, i.e. $M_{\mathcal{M}} = M'_{\mathcal{M}}$

(3) Show that $M_{\mathcal{M}}(\cdot)$ is idempotent, i.e. $M_{\mathcal{M}} \cdot M_{\mathcal{M}} = M_{\mathcal{M}}$

(4) Show that $M_{\mathcal{M}}(\hat{\mathbf{u}}) = \hat{\mathbf{u}}$

(5) Show that $P_{\mathcal{M}} \cdot M_{\mathcal{M}} = 0$

(6) Show that $P_{\mathcal{M}} + M_{\mathcal{M}} = I_n$

Frisch–Waugh–Lovell Theorem

If the regression we are interested in is expressed in terms of two separate sets of predictor variables (partitioned regression):

$$Y = X_A\beta_A + X_B\beta_B + u$$

then the estimate of $\hat{\beta}_A$ will be the same as the estimate of it from a modified regression of the form:

$$M_BY = M_BX_A\beta_A + M_Bu$$

Partialling out effect: by including additional regressors (X_B), the coefficients of β_A explains the variation between Y and X_A not explained by the other regressor.