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

y_i = METRICS SKILLS

x_i = MEASURE THEORY CLASS

i^{TH} OBS := LAPO $\quad y_i = x_i y_i(1) + (1-x_i) y_i(0) = y_i(1)$

- 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

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

  $$= \int E[y_i(1) - y_i(0) \mid X_i = x_i]\, f_{x_i}(x_i)\, \lambda(dx_i) = \sum_j E[y_i(1) - y_i(0) \mid X = x_i] \cdot P(x_i)$$

  $$= P(x_i = 1)\, E[y_i(1) - y_i(0) \mid x_i = 1]$$

  $$+ (1 - P(x_i = 1))\, E[y_i(1) - y_i(0) \mid x_i = 0]$$

  $$= \pi\, \tau_{ATT} + (1-\pi)\tau_{ATU} \quad ///$$

- 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:

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

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:

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

DIFFERENCES BETWEEN TREATED AND UNTREATED are not entirely due to the Treatment

IF I WERE UNTREATED, I WOULD HAVE HIGHER METRICS SKILLS THAN THE UNTREATED (PEOPLE DOING MICRO)

EFFECT OF TREATMENT Maybe higher on TREATED Than UNTREATED

Randomized Control Trial

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

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

(1)

$E[y_i(0)|x_i=1] = E[y_i(0)]$

$E[y_i(0)|x_i=0] = E[y_i(0)]$

$\therefore E[y_i(0)|x_i=1] = E[y_i(0)|x_i=0]$

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

(2) SINCE

$E[y_i(1)|x_i=1] = E[y_i(1)] = E[y_i(1)|x_i=0]$

$\Rightarrow \tau_{ATT} = \tau_{ATU}$ USING (1)

- Estimation: hat instead of expectation.

(3): STRAIGHTFORWARD: (1) IMPLIES SELECTION BIAS = 0, (2) HET. TREAT = 0

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 $u_i(1) \sim (0, \sigma^2)$.

$$y_i(1) = E[y_i(1)] + u_i(1)$$

$$y_i(0) = E[y_i(0)] + u_i(0)$$

$$y_i = x_i\, y_i(1) + (1 - x_i)\, y_i(0) = y_i(0) + x_i\,(y_i(1) - y_i(0))$$

$$= E[y_i(0)] + u_i(0) + x_i\,(E[y_i(1)] - E[y_i(0)]) + x_i\,(u_i(1) - u_i(0))$$

$$= E[y_i(0)] + x_i\,(E[y_i(1)] - E[y_i(0)]) + \underbrace{\{x_i u_i(1) + (1 - x_i)\, u_i(0)\}}_{:= u_i}$$

$$y_i = \beta_0 + x_i \beta + u_i$$

Implication Missing at Random

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

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

$$E[\underbrace{x_i u_i(1)}_{} + \underline{(1-x_i)\, u_i(0)} \mid x_i]$$

$$E[x_i u_i(1) \mid x_i] = x_i E[u_i(1) \mid x_i]$$
$$= x_i E[y_i(1) - E[y_i(1)] \mid x_i]$$
$$= x_i E[y_i(1) - E[y_i(1)]]$$
$$= x_i \left(E[y_i(1)] - E[y_i(1)]\right) = 0 \quad ///$$

(2) $E[u_i x_i] = E[x_i E[\underbrace{u_i \mid x_i}_{=0}]] = 0 \quad ///$

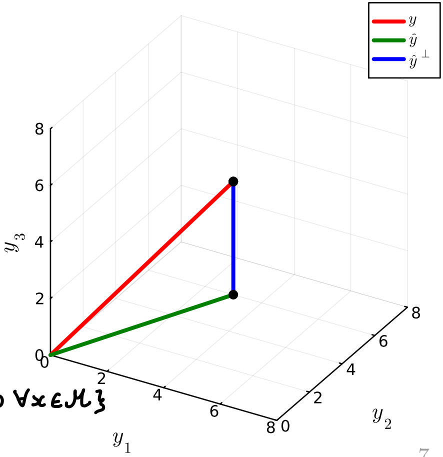
(3) $E[u_i] = E[E[\underbrace{u_i \mid x_i}_{=0}]] = 0 \quad ///$

WHY THOSE IMPORTANT: $E[u_i|x_i] = 0$ MARKOV GAUSS

$E[u_i x_i] = 0$ CONSISTENCY $\hat{\beta}_{OLS}$

Orthogonal Projection

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



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:

$$\hat{\mathbf{y}} = \arg\min_{x \in \mathcal{M}} \|\mathbf{y} - \mathbf{x}\|^2_{L^2} = \left(\sqrt{\sum_{i=1}^{3} (y_i - x_i)^2}\right)^2$$

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

$V = \mathbb{R}^3$

$\mathcal{M} \subseteq \mathbb{R}^2 \quad x \in \mathcal{M} \qquad \mathcal{M}^{\perp} := \{x^{\perp} : \langle x, x^{\perp} \rangle = 0 \ \forall x \in \mathcal{M}\}$

$\hat{y}^{\perp} \in \mathcal{M}^{\perp}$ ORTHOGONAL COMPLEMENT

WHAT IS THE VECTOR SUBSPACE? The one spanned by d-LINEAR INDEP. COLUMNS OF X.

Orthogonal Projection vs Linear Regression

CAN I ESTIMATE THE MODEL FOR $d < m$ ← dimension? d ←
$d = m$? $d > m$?

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:

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

$\hat{\mathbf{y}}$ 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**.

$$\beta := \arg\min_{b\in\mathbb{R}^d} E\left[\sum (y_i - x_i^T b)^2\right]$$ but in class

$$\beta := \arg\min_{b\in\mathbb{R}^d} E\left[(y_i - x_i^T \beta)^2\right]$$ ARE THOSE EQUIVALENT?

OLS Matrix Form

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

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

REMOVE EXPECTATION PLUG $\hat{\beta}$

$$\therefore \frac{\partial}{\partial \beta} E[(\mathbf{y} - Xb)^T(\mathbf{y} - Xb)] = 0 \implies \frac{1}{m}\frac{\partial \hat{e}^T \hat{e}}{\partial b} = \frac{1}{m}\hat{e}^T \frac{\partial \hat{e}}{b}$$

$$= \underbrace{\frac{1}{m}\hat{e}^T}_{(1 \times m)} \cdot \underbrace{(-X)}_{(m \times d)} = \underbrace{0}_{(1 \times d)}$$

SAMPLE MOMENT CONDITION HOLD

$X'\hat{e} = 0$

WE TAKE TRANSPOSE WE WANT A COLUMN VECTOR OF ZEROS

$$\longrightarrow \frac{1}{m}X'\hat{e} = \frac{1}{m}X'Y - \frac{1}{m}X'X\hat{\beta} = 0$$

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

OLS Vector Form

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

FOC:

[a] $\frac{1}{m} \sum 2(y_i - \hat{\alpha} - x_i^T \hat{\beta}) \cdot (-1) = 0$

$$-\frac{2}{m} \sum y_i + \frac{2}{m} \sum \hat{\alpha} + \frac{2}{m} \sum x_i^T \hat{\beta} = 0$$

$$\frac{1}{m} \sum \hat{\alpha} = \frac{1}{m} \sum y_i - \frac{1}{m} \sum x_i^T \hat{\beta} \qquad \hat{\alpha} = \bar{y} - \bar{x}^T \hat{\beta}$$

[b] $\frac{1}{m} \sum -2x_i (y_i - \hat{\alpha} - x_i^T \hat{\beta}) = 0$

$$-\frac{2}{m} \sum x_i y_i + \frac{2}{m} \sum x_i (\bar{y} - \bar{x}^T \hat{\beta}) + \frac{2}{m} \sum x_i x_i^T \beta = 0$$

$$-\frac{1}{m} \sum x_i y_i + \left(\frac{1}{m} \sum x_i\right) \bar{y} - \left(\frac{1}{m} \sum x_i\right) \bar{x}^T \hat{\beta} + \left(\frac{1}{m} \sum x_i x_i^T\right) \hat{\beta}$$

$$-\frac{1}{m} \sum x_i y_i + \bar{x}\bar{y} - \bar{x}\bar{x}^T \hat{\beta} + \left(\frac{1}{m} \sum x_i x_i^T\right) \hat{\beta} = 0$$

$$\hat{\beta} = \left[\frac{1}{m} \sum x_i x_i^T - \bar{x}\bar{x}^T\right]^{-1} \left[\frac{1}{m} \sum x_i y_i - \bar{x}\bar{y}\right]$$

OLS Vector Form

Regression: $y_i = \alpha + x_i^T\beta + u_i$ OLS: $(\alpha, \beta) = \underset{(a,b)}{\arg\min}\, E[\sum_{i=1}^n (y_i - a - x_i^t b)^2]$

$X'\hat{e}_i = \sum_{i=1}^{m} \begin{vmatrix} 1 \\ x_i \end{vmatrix} \hat{e}_i = 0_{[(d+1)\times 1]}$ where $x_i = |x_{i1} \dots x_{id}|^T$

DIMENSION 1

DIMENSION d

- IF WE HAVE INTERCEPT FIRST ELEMENT OF $\sum_{i=1}^{m} \begin{vmatrix} 1 \\ x_i \end{vmatrix} \hat{e}_i = 0$ IS $\sum \hat{e}_i = 0$

$\sum \hat{e}_i = \sum(y_i - \hat{\alpha} - x_i^T\hat{\beta}) = \sum(y_i - \bar{y} + \bar{x}^T\hat{\beta}_1 - x_i^T\hat{\beta})$

$= \sum y_i - \sum \bar{y} + \sum \bar{x}^T\beta - \sum x_i^T\beta = m\bar{y} - m\bar{y} + m\bar{x}^T\beta - m\bar{x}^T\beta = 0$ ✓

- OTHER ELEMENTS (2:d) elements are

$\sum x_i\hat{e}_i = \sum x_i(y_i - \hat{\alpha} - x_i^T\hat{\beta}) = 0$

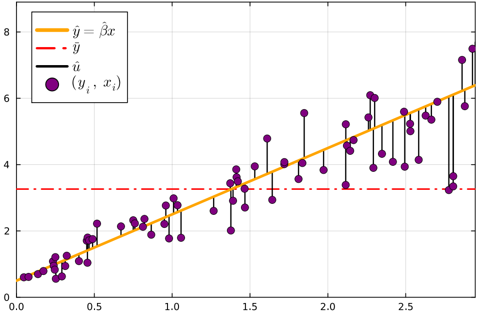
$\frac{1}{m}\sum x_i(y_i - \hat{\alpha} - x_i^T\hat{\beta}) = \frac{1}{m}\cdot 0$ FOC OLS ✓

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

REMEMBER $TSS = ESS + RSS$

$$y_i - \bar{y} = y_i - \hat{y}_i + \hat{y}_i - \bar{y} = \hat{u}_i + (\hat{y}_i - \bar{y})$$

TAKE SQUARE

$$(y_i - \bar{y})^2 = (\hat{u}_i)^2 + (\hat{y}_i - \bar{y})^2 + 2\hat{u}_i(\hat{y}_i - \bar{y})$$

TAKE SUM

$$\Sigma(\bar{y}_i - \bar{y})^2 = \Sigma(\hat{u}_i)^2 + \Sigma(\hat{y}_i - \bar{y})^2 + 2\{\Sigma\hat{u}_i\hat{y}_i - \Sigma\hat{u}_i\bar{y}\}$$

$$TSS = RSS + ESS + 2\{(\underbrace{\Sigma\hat{u}_i x_i^T}_{=0})\beta - \bar{y}\underbrace{\Sigma\hat{u}_i}_{=0}\}$$

BY PROPERTY OLS

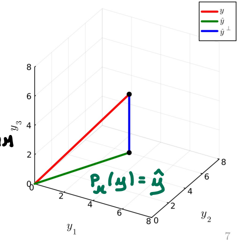
$$\therefore \quad \frac{TSS}{TSS} = \frac{RSS}{TSS} + \frac{ESS}{TSS} \longrightarrow R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Projection Matrix

Given our vector space $\overset{= \mathbb{R}^3}{\mathcal{V}}$ and $\overset{= \mathbb{R}^2}{\mathcal{M}} \subseteq V$, there is an orthogonal projection $P_{\mathcal{M}} : \mathcal{V} \to \mathcal{M}$ with the property that:

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

FROM $\mathbb{R}^3 \longrightarrow$ TO VECTOR IN $\mathbb{R}^2$
MINIMIZING $\|\cdot\|_{L2}$ NORM



(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 $P_{\mathcal{M}}(X) = X$

(i) $\hat{Y} = X\hat{\beta} = \underbrace{X(X'X)^{-1}X'}_{P_X}Y = P_X Y$

(ii) SYMMETRIC $P_X = P_X'$ $\longrightarrow$ $(X(X^TX)^{-1}X^T)^T = (X^T)^T([X^TX]^T)^{-1}(X)^T = X(X^TX)^{-1}X' = P_X$

INVERT ORDER $(AB)^T = B^TA^T$

IF A IS A SQUARED MATRIX $(A^{-1})^T = (A^T)^{-1}$

$\therefore [(X'X)^{-1}]^T = [(X'X)^T]^{-1}$

(iii) IDEMPOTENT $P_X P_X = P_X \longrightarrow X(\cancel{X'X})^{-1}\cancel{X'X}(X'X)^{-1}X' = X(X'X)^{-1}X'$

(iv) $P_X X = X \longrightarrow P_X X = X(X'X)^{-1}X'X = X \longrightarrow$ covariates unchanged after Projection

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 $P_{\mathcal{M}} + M_{\mathcal{M}} = I_n$

(i) ANNIHILATOR MATRIX : $\hat{u} = Y - \hat{Y} = Y - P_x Y = (I - P_x)Y = M_x Y$

(ii) SYMMETRIC $M_x = M_x' \longrightarrow M_x^T = (I - P_x)^T = (I^T - P_x^T) = (I - P_x) = M_x$

(iii) IDEMPOTENT $M_x \cdot M_x = (I - P_x)(I - P_x) = I - I \cdot P_x - P_x I + P_x P_x = I - 2P_x + P_x = (I - P_x) = M_x$

(iv) $M_x \hat{u} = \hat{u} \longrightarrow (I - X(X'X)^{-1}X')\hat{u} = \hat{u} - X(X'X)^{-1}\cancel{X'\hat{u}}^{\,0} = \hat{u}$

(v) $M_x \cdot P_x = 0 \longrightarrow (I - P_x)P_x = P_x - P_x P_x = P_x - P_x = 0$

(vi) $(P_x + M_x) = I \longrightarrow P_x + (I - P_x) = I$

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.

$$M_b = (I - P_b) = (I - X_b(X_b' X_b)^{-1} X_b') \longrightarrow (n \times b) \text{ dimension.}$$

$Y = X_A \hat{\beta}_A + X_B \hat{\beta}_B + \hat{u}$ Let's multiply by M_b

$$M_b Y = M_b X_a \hat{\beta}_a + M_b X_b \hat{\beta}_b + M_b \hat{u}$$

$$M_b Y = M_b X_a \hat{\beta}_a + (I - X_b(X_b' X_b)^{-1} X_b') X_b \hat{\beta}_b + \hat{u}$$

$M_b Y = M_b X_a \hat{\beta}_a + (X_b - X_b(X_b' X_b)^{-1} X_b' X_b) \hat{\beta}_b + \hat{u}$ Now let's multiply by X_a'

$X_a' M_b Y = (X_a' M_b X_a) \hat{\beta}_a + X_a' \hat{u} \longrightarrow 0$ by construction

We can rewrite the expression

$$\hat{\beta}_a = (X_a' M_b X_a)^{-1} (X_a' M_b Y)$$

M_b IDEMPOTENT

$$\hat{\beta}_a = (X_a' M_b M_b X_a)^{-1} (X_a' M_b M_b Y) \longrightarrow$$

$$\hat{\beta}_a = (X_a' M_b' M_b X_a)^{-1} (X_a' M_b' M_b Y)$$

M_b SYMMETRIC

$$\hat{\beta}_a = (W'W)^{-1}(W'V)$$

$M_b Y$ residual from $Y \sim X_b$ regression $= W$

$M_b X_a$ residual from $X_a \sim X_b$ regression $= V$

$$(I - X_b(X_b' X_b)^{-1} X_b') X_a = V$$

$$X_a - X_b[(X_b' X_b)^{-1} X_b' X_a] = V$$

$X_a - X_b \delta = V$ regression $X_a = X_b \delta + V$