Lecture 3: Geometry of OLS and partitioned regression

Economics 527: Econometric Methods

Vadim Marmer, UBC

Geometry of OLS

Least squares as a distance

- OLS, first lecture:

$$\hat{\beta} = \arg\min_{b \in \mathbb{R}^k} \underbrace{(Y - Xb)^\top (Y - Xb)}_{= \|Y - Xb\|^2}$$

- Norm of $a \in \mathbb{R}^n$:

$$\begin{aligned} \|a\| &= \sqrt{\sum_{i=1}^{n} a_i^2} \\ &= \sqrt{a^\top a} \end{aligned}$$

The column space $S(X)$

- Column space of X:

$$S(X) = \left\{ v \in \mathbb{R}^n : v = \underset{n \times k}{X} \underset{k \times 1}{b} \quad \text{for some } b \in \mathbb{R}^k \right\}$$

- Partition X column-wise:

$$\underbrace{X}_{n \times k} = [\ \underbrace{\underline{X}_1}_{n \times 1} \ \cdots \ \underbrace{\underline{X}_k}_{n \times 1}\ ],$$

  where $\underline{X}_j$ is the $n \times 1$ vector of the observations on the j-th regressor.

- What the elements of $S(X)$ look like, in terms of the columns of X:

$$\begin{aligned} \underbrace{X}_{n \times k} \underbrace{b}_{k \times 1} &= \underbrace{[\underline{X}_1 \ \cdots \ \underline{X}_k]}_{= X} \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix} \\ &= b_1 \underline{X}_1 + \cdots + b_k \underline{X}_k \\ &= \sum_{j=1}^{k} b_j \underline{X}_j \end{aligned}$$

- Each $\underline{X}_j \in S(X)$, since we can take $b_j = 1$, $b_i = 0$ for $i \neq j$:

$$\begin{aligned} Xb &= [\underline{X}_1 \ \cdots \ \underline{X}_k] \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \\ &= \underline{X}_j \in S(X) \end{aligned}$$

- $S(X) \subset \mathbb{R}^n$: k-dimensional subspace; $\operatorname{rank}(X) = k < n$.

$\hat{Y}$: closest to Y in $S(X)$

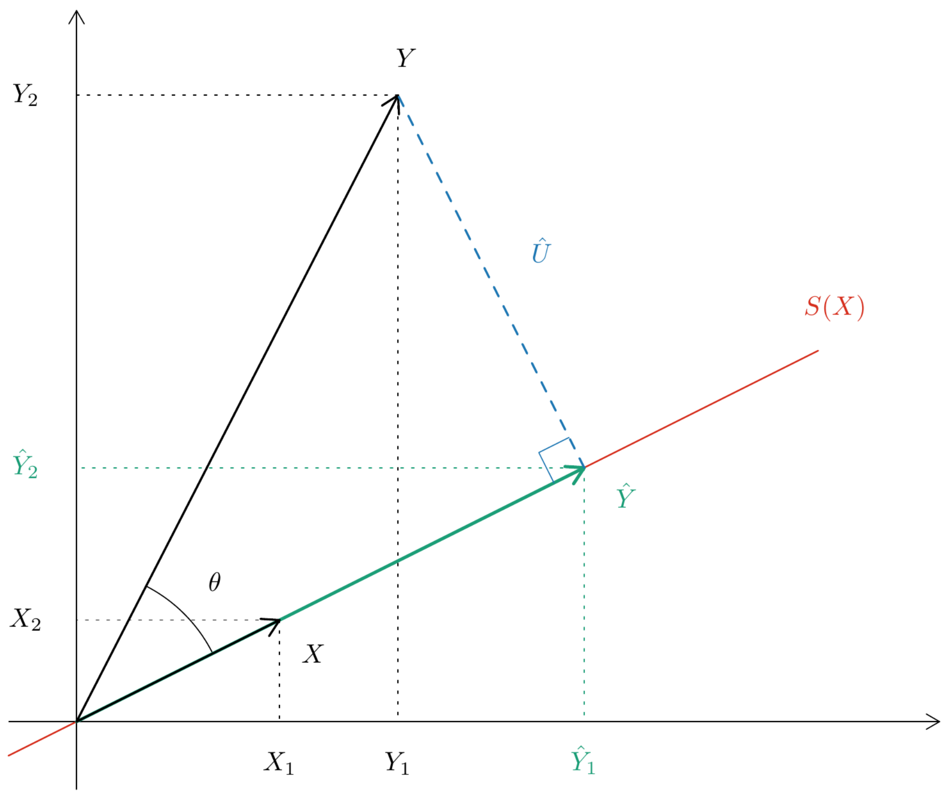
- We can re-state the OLS problem in terms of $S(X)$:

$$\hat{\beta} = \arg\min_{b \in \mathbb{R}^k} \|Y - Xb\|^2, \qquad \hat{Y} = X\hat{\beta}$$

$$\iff$$

$$\hat{Y} = \arg\min_{v \in S(X)} \|Y - v\|^2$$

- When $k = 1$, $n = 2$, we can visualize the solution: You can find the closest element by dropping a perpendicular from Y to $S(X)$.
- $Y = (Y_1, Y_2)^\top$, $X = (X_1, X_2)^\top$, and $S(X) = \{(bX_1, bX_2)^\top : b \in \mathbb{R}\}$.



- The coefficient $\hat{\beta}$ is determined by the angle θ between X and Y:

$$\begin{aligned}
\cos\theta &= \frac{X^\top Y}{\|X\|\,\|Y\|} \\
\hat{\beta} &= \frac{X^\top Y}{\|X\|^2} \\
&= \cos\theta \cdot \frac{\|Y\|}{\|X\|}
\end{aligned}$$

- Residuals, perpendicular to $S(X)$:

$$\hat{U} = Y - \hat{Y} = Y - X\hat{\beta}$$

Dropping the perpendicular: the algebra

- Fitted values:

$$
\begin{aligned}
\hat{Y} &= X\hat{\beta} \\
&= \underbrace{X(X^\top X)^{-1}X^\top}_{P_X} Y \\
&= P_X Y
\end{aligned}
$$

- Residuals:

$$
\begin{aligned}
\hat{U} &= Y - \hat{Y} \\
&= Y - X\hat{\beta} \\
&= Y - X(X^\top X)^{-1}X^\top Y \\
&= Y - P_X Y \\
&= \underbrace{(I_n - P_X)}_{= M_X} Y \\
&= M_X Y
\end{aligned}
$$

- Both $n \times n$:

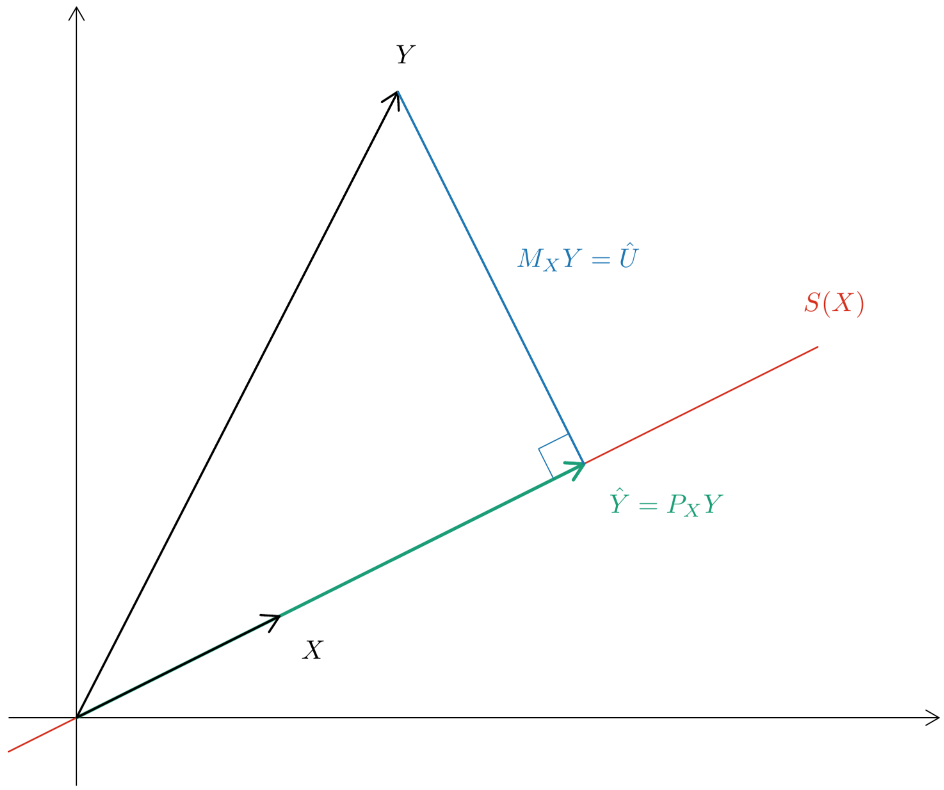
$$
\boxed{P_X = \underset{n\times k}{X}\ \underset{k\times k}{(X^\top X)}^{-1} \underset{k\times n}{X}^\top \qquad M_X = \underset{n\times n}{I_n} - \underset{n\times n}{P_X}}
$$

Projection matrices

$P_X Y$ and $M_X Y$

- For $Y \in \mathbb{R}^n$:

$$
\begin{aligned}
& P_X + M_X = I_n \\
\implies \quad & P_X Y + M_X Y = Y
\end{aligned}
$$



Properties of P_X

- Symmetric, by $(AB)^\top = B^\top A^\top$:

$$
\begin{aligned}
P_X^\top &= \left(X(X^\top X)^{-1} X^\top\right)^\top \\
&= X(X^\top X)^{-1} X^\top \\
&= P_X
\end{aligned}
$$

- $P_X X = X$:

$$
\begin{aligned}
P_X X &= \underbrace{X(X^\top X)^{-1} X^\top}_{P_X} X \\
&= X
\end{aligned}
$$

- Column by column, $P_X \underline{X}_j = \underline{X}_j$ for $j = 1, \dots, k$:

$$
\begin{aligned}
P_X X &= P_X [\, \underset{n \times 1}{\underline{X}_1} \ \cdots \ \underset{n \times 1}{\underline{X}_k} \,] \\
&= [\, P_X \underline{X}_1 \ \cdots \ P_X \underline{X}_k \,] \\
&= [\, \underline{X}_1 \ \cdots \ \underline{X}_k \,],
\end{aligned}
$$

where the last equality holds because $P_X X = X$.

- Idempotent:

$$
\begin{aligned}
P_X^2 &= P_X P_X \\
&= \underbrace{P_X X}_{=X} (X^\top X)^{-1} X^\top \\
&= P_X
\end{aligned}
$$

P_X: p.s.d., trace k

- For $a \in \mathbb{R}^n$:

$$\begin{aligned} a^\top P_X a &= \underbrace{a^\top X}_{w^\top} \underbrace{(X^\top X)^{-1}}_{\text{p.d.}} \underbrace{X^\top a}_{w} \\ &\geq 0 \end{aligned}$$

- We will use: $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

$$\begin{aligned} \operatorname{tr}(P_X) &= \operatorname{tr}\left(X(X^\top X)^{-1}X^\top\right) \\ &= \operatorname{tr}\left((X^\top X)^{-1} \underbrace{\underset{k\times n}{X^\top}\ \underset{n\times k}{X}}_{k\times k}\right) \\ &= \operatorname{tr}(I_k) \\ &= k \end{aligned}$$

- P_X: $n \times n$, symmetric, p.s.d., idempotent, rank k.
- $P_X : \mathbb{R}^n \longrightarrow S(X)$.

Properties of M_X

- Symmetric:

$$\begin{aligned} M_X^\top &= I_n^\top - P_X^\top \\ &= I_n - P_X \\ &= M_X \end{aligned}$$

- $M_X X = 0$:

$$\begin{aligned} M_X X &= (I_n - P_X)X \\ &= I_n X - P_X X \\ &= X - \underbrace{P_X X}_{=X} \\ &= X - X \\ &= 0 \end{aligned}$$

- $v = Xb \in S(X)$, $b \in \mathbb{R}^k$:

$$\begin{aligned} M_X v &= M_X X b \\ &= 0 \end{aligned}$$

$M_X P_X = 0$ and idempotency

- $M_X P_X = 0$ and $P_X M_X = 0$:

$$\begin{aligned} M_X P_X &= \underbrace{M_X X}_{=0} (X^\top X)^{-1} X^\top \\ &= 0 \\ P_X M_X &= (M_X P_X)^\top \\ &= 0 \end{aligned}$$

- Idempotent:

$$\begin{aligned} M_X M_X &= (I_n - P_X) M_X \\ &= M_X - \underbrace{P_X M_X}_{=0} \\ &= M_X \end{aligned}$$

The orthogonal complement $S^{\perp}(X)$

- Orthogonal complement of $S(X)$:

$$S^{\perp}(X) = \{\, v \in \mathbb{R}^n : v^{\top} X = 0 \,\}$$

- Normal equations, first lecture:

$$\begin{aligned} X^{\top}\hat{U} &= X^{\top}(Y - X\hat{\beta}) \\ &= X^{\top}Y - X^{\top}X \underbrace{\hat{\beta}}_{=(X^{\top}X)^{-1}X^{\top}Y} \\ &= X^{\top}Y - X^{\top}Y \\ &= 0 \\ \implies \quad \hat{U}^{\top}X &= 0 \\ \implies \quad \hat{U} &= M_X Y \in S^{\perp}(X) \end{aligned}$$

- $M_X : \mathbb{R}^n \longrightarrow S^{\perp}(X)$.

M_X: rank $n-k$, p.s.d.

- Rank:

$$\begin{aligned} \operatorname{rank}(M_X) &= \operatorname{tr}(M_X) \\ &= \operatorname{tr}(I_n - P_X) \\ &= \operatorname{tr}(I_n) - \operatorname{tr}(P_X) \\ &= n - \operatorname{tr}(P_X) \\ &= n - \operatorname{tr}\left(X(X^{\top}X)^{-1}X^{\top}\right) \\ &= n - \operatorname{tr}\big(\underbrace{(X^{\top}X)^{-1}X^{\top}X}_{I_k}\big) \\ &= n - k \end{aligned}$$

- M_X: $n \times n$; no inverse.
- **Claim.** M_X: p.s.d. For $a \in \mathbb{R}^n$ and $w = M_X a$:

$$\begin{aligned} a^{\top}M_X a &= a^{\top}M_X M_X a \quad \text{(idempotent)} \\ &= \underbrace{a^{\top}M_X^{\top}}_{w^{\top}} \underbrace{M_X a}_{w} \quad \text{(symmetric)} \\ &= w^{\top}w \\ &= \|w\|^2 \\ &\geq 0 \end{aligned}$$

$Y = \hat{Y} + \hat{U}$

- Decomposition:

$$\underset{\in \mathbb{R}^n}{Y} = \underset{\in S(X)}{\hat{Y}} + \underset{\in S^{\perp}(X)}{\hat{U}}, \qquad S(X) + S^{\perp}(X) = \mathbb{R}^n$$

- Fitted values:

$$\begin{aligned} \hat{Y} &= X\hat{\beta} \\ &= P_X Y \in S(X), \quad \text{dimension } k \end{aligned}$$

- Residuals:

$$\begin{aligned}\hat{U} &= Y - \hat{Y} \\ &= M_X Y \in S^{\perp}(X), \quad \text{dimension } n-k\end{aligned}$$

- Orthogonality:

$$\begin{aligned}\hat{Y}^\top \hat{U} &= Y^\top P_X^\top M_X Y \\ &= Y^\top \underbrace{P_X M_X}_{=0} Y \\ &= 0\end{aligned}$$

Estimation of σ^2

Two estimators of σ^2

- Assumptions 2 and 3:

$$\begin{aligned}\operatorname{Var}(U \mid X) &= \mathrm{E}\left[UU^\top \mid X\right] \\ &= \sigma^2 I_n\end{aligned}$$

- Unknown parameter:

$$\begin{aligned}\sigma^2 &= \mathrm{E}\left[U_i^2 \mid X\right] \\ &= \mathrm{E}\left[U_i^2\right]\end{aligned}$$

- Estimators:

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} \hat{U}_i^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} (Y_i - X_i^\top \underset{k\times 1}{\hat{\beta}})^2 \\ &= \frac{1}{n}\hat{U}^\top \hat{U} \\ s^2 &= \frac{1}{n-k}\sum_{i=1}^{n} \hat{U}_i^2\end{aligned}$$

- **Claim.** Under Assumptions 1 to 4:

$$\mathrm{E}\left[s^2 \mid X\right] = \sigma^2$$

Proof, step 1: residuals to errors

- Residuals:

$$\begin{aligned}\hat{U} &= M_X Y \\ &= M_X(X\beta + U) \quad \text{(Assumption 1)} \\ &= M_X U \quad (M_X X = 0)\end{aligned}$$

- Dimensions:

$$\underset{n\times 1}{\hat{U}} = \underset{n\times n}{M_X}\ \underset{n\times 1}{U}, \qquad \operatorname{rank}(M_X) = n-k$$

- Sum of squared residuals:

$$\begin{aligned}
\sum_{i=1}^{n} \hat{U}_i^2 &= \|\hat{U}\|^2 \\
&= \hat{U}^\top \hat{U} \\
&= (M_X U)^\top (M_X U) \\
&= U^\top M_X^\top M_X U \\
&= U^\top M_X U \quad \text{(symmetric, idempotent)}
\end{aligned}$$

Proof, step 2: a quadratic form

- To find:

$$\mathrm{E}\left[\hat{U}^\top \hat{U} \mid X\right] = \mathrm{E}\left[\overbrace{\underset{1\times n}{U^\top} \underset{n\times n}{M_X} \underset{n\times 1}{U}}^{1\times 1} \mid X\right]$$

- Does not factor:

$$\mathrm{E}\left[U^\top M_X U \mid X\right] \;\neq\; \mathrm{E}\left[U^\top \mid X\right] \, M_X \, \mathrm{E}\left[U \mid X\right]$$

- Quadratic form:

$$\underset{1\times n}{U^\top} \underset{n\times n}{A} \underset{n\times 1}{U} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} U_i U_j$$

- $\mathrm{E}\left[U_i U_j\right] \neq \mathrm{E}\left[U_i\right] \mathrm{E}\left[U_j\right]$ unless uncorrelated; $\mathrm{E}\left[U_i^2\right] \neq \mathrm{E}\left[U_i\right] \mathrm{E}\left[U_i\right]$.

Proof, step 3: the trace trick

- We will use:
  - 1×1 matrix: its own trace;
  - $\mathrm{tr}(AB) = \mathrm{tr}(BA)$;
  - $\mathrm{E}\left[\sum \cdot\right] = \sum \mathrm{E}\left[\cdot\right]$;
  - M_X: known given X.
- We have:

$$\begin{aligned}
\mathrm{E}\left[\hat{U}^\top \hat{U} \mid X\right] &= \mathrm{E}\left[U^\top M_X U \mid X\right] \\
&= \mathrm{E}\left[\mathrm{tr}\left(U^\top M_X U\right) \mid X\right] \\
&= \mathrm{E}\left[\mathrm{tr}\left(M_X U U^\top\right) \mid X\right] \\
&= \mathrm{tr}\left(\mathrm{E}\left[M_X U U^\top \mid X\right]\right) \\
&= \mathrm{tr}\left(M_X \underbrace{\mathrm{E}\left[U U^\top \mid X\right]}_{=\sigma^2 I_n}\right) \\
&= \sigma^2 \, \mathrm{tr}(M_X) \\
&= \sigma^2 (n-k)
\end{aligned}$$

$\hat{\sigma}^2$: biased

- Residuals:

$$\begin{aligned}
&\mathrm{E}\left[\sum_{i=1}^{n} \hat{U}_i^2 \mid X\right] = \sigma^2 (n-k) \\
\implies \quad &\mathrm{E}\left[\sum_{i=1}^{n} \hat{U}_i^2\right] = \sigma^2 (n-k) \quad \text{(LIE)}
\end{aligned}$$

- Errors:

$$\begin{aligned}\mathrm{E}\left[U^\top U\right] &= \mathrm{E}\left[\sum_{i=1}^n U_i^2\right] \\ &= n\sigma^2\end{aligned}$$

- Biased, by a known factor:

$$\begin{aligned}\mathrm{E}\left[\hat{\sigma}^2 \mid X\right] &= \frac{1}{n}\,\mathrm{E}\left[\hat{U}^\top \hat{U} \mid X\right] \\ &= \frac{n-k}{n}\,\sigma^2 \\ &\neq \sigma^2\end{aligned}$$

s^2: unbiased

- Correction:

$$\begin{aligned}\frac{n}{n-k}\,\mathrm{E}\left[\hat{\sigma}^2 \mid X\right] = \sigma^2 \\ s^2 &= \frac{n}{n-k}\,\hat{\sigma}^2 \\ &= \frac{\hat{U}^\top \hat{U}}{n-k} \\ &= \frac{\sum_{i=1}^n \hat{U}_i^2}{n-k} \\ &= \frac{\sum_{i=1}^n (Y_i - X_i^\top \hat{\beta})^2}{n-k} \\ &= \frac{\|Y - X\hat{\beta}\|^2}{n-k}\end{aligned}$$

- Unbiased:

$$\begin{aligned}&\boxed{\mathrm{E}\left[s^2 \mid X\right] = \sigma^2} \\ \overset{\text{LIE}}{\Longrightarrow}\quad &\mathrm{E}\left[s^2\right] = \sigma^2\end{aligned}$$

Standard errors

- Second lecture: $\mathrm{Var}\left(\hat{\beta} \mid X\right) = \sigma^2 (X^\top X)^{-1}$. Estimated:

$$\widehat{\mathrm{Var}}(\hat{\beta} \mid X) = s^2 (X^\top X)^{-1}$$

- Standard error against standard deviation:

$$\begin{aligned}\text{std. err}(\hat{\beta}_j) &= \sqrt{s^2 \left[(X^\top X)^{-1}\right]_{jj}} \\ &= \sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_j \mid X)} \\ \text{std. dev}(\hat{\beta}_j) &= \sqrt{\mathrm{Var}\left(\hat{\beta}_j \mid X\right)}\end{aligned}$$

Partitioned regression

The partition

- Two blocks of regressors, $k_1 + k_2 = k$:

$$\underset{n\times k}{X} = \left[\ \underset{n\times k_1}{X_1} \quad \underset{n\times k_2}{X_2}\ \right], \qquad \beta = \begin{bmatrix}\beta_1 \\ \beta_2\end{bmatrix} \begin{matrix} k_1 \times 1 \\ k_2 \times 1\end{matrix}$$

- Model:

$$\begin{aligned} Y &= \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + U \\ &= X_1\beta_1 + X_2\beta_2 + U \end{aligned}$$

- OLS:

$$\begin{aligned} \underset{k\times 1}{\hat{\beta}} &= \begin{bmatrix} \underset{k_1\times 1}{\hat{\beta}_1} \\ \underset{k_2\times 1}{\hat{\beta}_2} \end{bmatrix} \\ &= (X^\top X)^{-1}X^\top Y \end{aligned}$$

The Frisch-Waugh-Lovell theorem

- **Claim.**

$$\hat{\beta}_1 = (X_1^\top M_2 X_1)^{-1} X_1^\top M_2 Y, \qquad M_2 = I_n - P_2, \quad P_2 = X_2(X_2^\top X_2)^{-1}X_2^\top$$

- M_2: symmetric, idempotent, $M_2X_2 = 0$.
- $M_2 : \mathbb{R}^n \longrightarrow S^\perp(X_2)$.
- $M_1 = I_n - X_1(X_1^\top X_1)^{-1}X_1^\top$:

$$\hat{\beta}_2 = (X_2^\top M_1 X_2)^{-1} X_2^\top M_1 Y$$

Proof

- We will use:

$$\begin{aligned} & \hat{U} = Y - X\hat{\beta} \\ & \phantom{\hat{U}} = Y - X_1\hat{\beta}_1 - X_2\hat{\beta}_2 \\ \iff \quad & Y = X\hat{\beta} + \hat{U} \\ & = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{U} \end{aligned}$$

- We have:

$$\begin{aligned} &(X_1^\top M_2 X_1)^{-1} X_1^\top M_2 Y \\ &= (X_1^\top M_2 X_1)^{-1} X_1^\top M_2 (X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{U}) \\ &= \underbrace{(X_1^\top M_2 X_1)^{-1} X_1^\top M_2 X_1}_{=I_{k_1}} \hat{\beta}_1 \\ &\qquad + (X_1^\top M_2 X_1)^{-1} X_1^\top \underbrace{M_2 X_2}_{=0} \hat{\beta}_2 \\ &\qquad + (X_1^\top M_2 X_1)^{-1} X_1^\top M_2 \hat{U} \\ &= \hat{\beta}_1 + (X_1^\top M_2 X_1)^{-1} X_1^\top \underbrace{M_2\hat{U}}_{=\hat{U}} \\ &= \hat{\beta}_1 + (X_1^\top M_2 X_1)^{-1} \underbrace{X_1^\top \hat{U}}_{=0} \\ &= \boxed{\hat{\beta}_1} \end{aligned}$$

$M_2\hat{U} = \hat{U}$

- Normal equations, block by block:

$$
\begin{aligned}
0 &= X^\top \hat{U} \\
&= \begin{bmatrix} X_1 & X_2 \end{bmatrix}^\top \hat{U} \\
&= \begin{bmatrix} X_1^\top \\ X_2^\top \end{bmatrix} \hat{U} \\
&= \begin{bmatrix} X_1^\top \hat{U} \\ X_2^\top \hat{U} \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\end{aligned}
$$

- Hence:

$$
\begin{aligned}
M_2\hat{U} &= (I_n - P_2)\hat{U} \\
&= \hat{U} - P_2\hat{U} \\
&= \hat{U} - X_2(X_2^\top X_2)^{-1} \underbrace{X_2^\top \hat{U}}_{=0} \\
&= \hat{U}
\end{aligned}
$$

- Hence:

$$
\begin{aligned}
X_1^\top M_2 \hat{U} &= X_1^\top \hat{U} \\
&= 0
\end{aligned}
$$

Residuals on residuals

- We will use: $M_2^\top = M_2$ and $M_2 M_2 = M_2$.
- We have:

$$
\begin{aligned}
\hat{\beta}_1 &= (X_1^\top M_2 X_1)^{-1} X_1^\top M_2 Y \\
&= (X_1^\top M_2^\top M_2 X_1)^{-1} X_1^\top M_2^\top M_2 Y \\
&= \left((M_2X_1)^\top (M_2X_1)\right)^{-1} (M_2X_1)^\top (M_2Y) \\
&= (\tilde{X}_1^\top \tilde{X}_1)^{-1} \tilde{X}_1^\top \tilde{Y}
\end{aligned}
$$

- $\tilde{X}_1 = M_2X_1$, $\tilde{Y} = M_2Y$, $\tilde{U} = M_2U$:

$$
\begin{aligned}
Y &= X_1\beta_1 + X_2\beta_2 + U \\
M_2Y &= M_2X_1\beta_1 + \underbrace{M_2X_2}_{=0} \beta_2 + M_2U \\
&= M_2X_1\beta_1 + M_2U \\
\tilde{Y} &= \tilde{X}_1\beta_1 + \tilde{U}
\end{aligned}
$$

$\tilde{X}_1$ and $\tilde{Y}$: OLS residuals

- Residuals from the regression of the columns of X_1 against X_2:

$$
\begin{aligned}
\tilde{X}_1 &= M_2 X_1 \\
&= (I_n - P_2) X_1 \\
&= X_1 - P_2 X_1 \\
&= X_1 - X_2 \underbrace{(X_2^\top X_2)^{-1} X_2^\top X_1}_{\hat{\gamma}} \\
&= X_1 - X_2 \hat{\gamma} \\
\iff \quad X_1 &= X_2 \hat{\gamma} + \tilde{X}_1
\end{aligned}
$$

- $\tilde{X}_1^\top X_2 = 0$.
- Residuals from the regression of Y against X_2:

$$
\begin{aligned}
\tilde{Y} &= M_2 Y \\
&= Y - X_2 \underbrace{(X_2^\top X_2)^{-1} X_2^\top Y}_{\hat{\eta}} \\
\iff \quad Y &= X_2 \hat{\eta} + \tilde{Y}
\end{aligned}
$$

- In general: $\hat{\eta} \neq \hat{\beta}_2$, $\tilde{Y} \neq \hat{U}$.

The regression in three steps

1. Remove from X_1 the part collinear with X_2: $\tilde{X}_1$.
2. Remove from Y the part collinear with X_2: $\tilde{Y}$.
3. $\hat{\beta}_1$: the OLS coefficient from the regression of $\tilde{Y}$ against $\tilde{X}_1$.

- $\hat{\beta}_1$: the estimated effect of X_1 after controlling for X_2.
- Step 2 optional:

$$
\begin{aligned}
\hat{\beta}_1 &= (X_1^\top M_2 X_1)^{-1} X_1^\top M_2 Y \\
&= (X_1^\top M_2^\top M_2 X_1)^{-1} X_1^\top M_2^\top Y \\
&= \left((M_2 X_1)^\top (M_2 X_1)\right)^{-1} (M_2 X_1)^\top Y \\
&= (\tilde{X}_1^\top \tilde{X}_1)^{-1} \tilde{X}_1^\top Y
\end{aligned}
$$

- Step 1 is not optional:

$$
(X_1^\top X_1)^{-1} X_1^\top \tilde{Y} \;\neq\; (\tilde{X}_1^\top \tilde{X}_1)^{-1} \tilde{X}_1^\top \tilde{Y} \quad \text{in general}
$$

Partitioned regression and the intercept

The intercept as a regressor

- One regressor and an intercept:

$$
Y_i = \underset{1\times 1}{\beta_1} + \underset{1\times 1}{\beta_2} \underset{1\times 1}{X_{i,2}} + U_i
$$

- Matrix form, with ℓ the $n \times 1$ vector of ones:

$$
\begin{aligned}
X &= \begin{bmatrix} 1 & X_{1,2} \\ \vdots & \vdots \\ 1 & X_{n,2} \end{bmatrix} \\
&= \begin{bmatrix} \ell & X_2 \end{bmatrix} \\
Y &= \underset{1\times 1}{\beta_1} \underset{n\times 1}{\ell} + \underset{1\times 1}{\beta_2} \underset{n\times 1}{X_2} + U
\end{aligned}
$$

- Partitioned regression:

$$
\begin{aligned}
\underset{1\times 1}{\hat{\beta}_2} &= \big(\underset{1\times 1}{X_2^\top M_\ell X_2}\big)^{-1} \underset{1\times n}{X_2^\top} \underset{n\times 1}{M_\ell Y} \\
M_\ell &= I_n - \ell\big(\underset{1\times 1}{\ell^\top \ell}\big)^{-1} \ell^\top
\end{aligned}
$$

P_ℓ and M_ℓ

- $\ell^\top \ell$:

$$
\begin{aligned}
\ell^\top \ell &= \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}^\top \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \\
&= \sum_{i=1}^{n} 1 \cdot 1 \\
&= n
\end{aligned}
$$

- Hence:

$$
\begin{aligned}
P_\ell &= \ell(\ell^\top \ell)^{-1} \ell^\top \\
&= \frac{\ell \ell^\top}{n} \\
M_\ell &= I_n - \frac{\ell \ell^\top}{n}
\end{aligned}
$$

- For $v = (v_1, \dots, v_n)^\top$:

$$
\begin{aligned}
(\ell^\top \ell)^{-1} \ell^\top v &= \frac{\ell^\top v}{n} \\
&= \frac{\sum_{i=1}^n \ell_i v_i}{n} \\
&= \frac{\sum_{i=1}^n v_i}{n} \\
&= \bar{v}
\end{aligned}
$$

P_ℓ: averaging

- For $v \in \mathbb{R}^n$:

$$
\begin{aligned}
P_\ell v &= \ell(\ell^\top \ell)^{-1} \ell^\top v \\
&= \frac{\ell \ell^\top}{n} v \\
&= \ell \left(\frac{\sum_{i=1}^n 1 \cdot v_i}{n} \right) \\
&= \ell\, \bar{v} \\
&= \begin{pmatrix} \bar{v} \\ \vdots \\ \bar{v} \end{pmatrix}
\end{aligned}
$$

- Column space of ℓ:

$$
\begin{aligned}
S(\ell) &= \{\, w \in \mathbb{R}^n : w = c\,\ell,\ c \in \mathbb{R} \,\} \\
&= \left\{\, w \in \mathbb{R}^n : w = \begin{pmatrix} c \\ \vdots \\ c \end{pmatrix},\ c \in \mathbb{R} \,\right\}
\end{aligned}
$$

M_ℓ: deviations from the mean

- For $v \in \mathbb{R}^n$:

$$
\begin{aligned}
M_\ell v &= v - P_\ell v \\
&= v - \ell\,\bar{v} \\
&= \begin{pmatrix} v_1 - \bar{v} \\ \vdots \\ v_n - \bar{v} \end{pmatrix}
\end{aligned}
$$

- Orthogonal complement of $S(\ell)$:

$$
\begin{aligned}
S^\perp(\ell) &= \left\{\, v \in \mathbb{R}^n : \underbrace{v^\top \ell}_{=\sum_{i=1}^n v_i} = 0 \,\right\} \\
&= \left\{\, v \in \mathbb{R}^n : \sum_{i=1}^n v_i = 0 \,\right\} \\
&= \{\, v \in \mathbb{R}^n : \bar{v} = 0 \,\}
\end{aligned}
$$

$\tilde{X}_2 = M_\ell X_2$

- Mean of the regressor:

$$
\begin{aligned}
(\ell^\top \ell)^{-1} \ell^\top X_2 &= \frac{\ell^\top X_2}{n} \\
&= \frac{1}{n} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}^\top \begin{bmatrix} X_{1,2} \\ \vdots \\ X_{n,2} \end{bmatrix} \\
&= \frac{1}{n} \sum_{i=1}^n 1 \cdot X_{i,2} \\
&= \frac{1}{n} \sum_{i=1}^n X_{i,2} \\
&= \bar{X}_2
\end{aligned}
$$

- The regressor:

$$
\begin{aligned}
\tilde{X}_2 &= M_\ell X_2 \\
&= X_2 - \ell\,\frac{\ell^\top X_2}{n} \\
&= X_2 - \bar{X}_2\,\ell \\
&= \begin{pmatrix} X_{1,2} - \bar{X}_2 \\ \vdots \\ X_{n,2} - \bar{X}_2 \end{pmatrix}
\end{aligned}
$$

- The dependent variable: $M_\ell Y = (Y_1 - \bar{Y}, \dots, Y_n - \bar{Y})^\top$.

Slope: deviations from the mean

- We will use: $M_\ell^\top = M_\ell$, $M_\ell M_\ell = M_\ell$, and $X_2^\top M_\ell X_2$: 1×1.
- We have:

$$\begin{aligned}
\hat{\beta}_2 &= (X_2^\top M_\ell X_2)^{-1} X_2^\top M_\ell Y \\
&= \frac{X_2^\top M_\ell Y}{X_2^\top M_\ell X_2} \\
&= \frac{(M_\ell X_2)^\top Y}{(M_\ell X_2)^\top (M_\ell X_2)} \\
&= \frac{\tilde{X}_2^\top Y}{\tilde{X}_2^\top \tilde{X}_2} \\
&= \frac{(X_2 - \bar{X}_2\, \ell)^\top Y}{(X_2 - \bar{X}_2\, \ell)^\top (X_2 - \bar{X}_2\, \ell)} \\
&= \frac{\sum_{i=1}^n (X_{i,2} - \bar{X}_2)\, Y_i}{\sum_{i=1}^n (X_{i,2} - \bar{X}_2)^2} \\
&= \boxed{\frac{\sum_{i=1}^n (X_{i,2} - \bar{X}_2)(Y_i - \bar{Y})}{\sum_{i=1}^n (X_{i,2} - \bar{X}_2)^2}} \qquad (X_2^\top M_\ell Y = X_2^\top M_\ell^\top M_\ell Y)
\end{aligned}$$

Several regressors and an intercept

- Intercept β_1 and slope coefficients β_2:

$$\underset{n\times k}{X} = [\ \underset{n\times 1}{\ell} \quad \underset{n\times(k-1)}{X_2}\ ], \qquad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \qquad Y = \beta_1 \ell + X_2 \beta_2 + U$$

- Each regressor minus its mean:

$$\begin{aligned}
\tilde{X}_2 &= M_\ell X_2 \\
&= M_\ell [\, \underline{X}_2 \ \cdots \ \underline{X}_k \,] \\
&= [\, M_\ell \underline{X}_2 \ \cdots \ M_\ell \underline{X}_k \,] \\
&= [\, \tilde{\underline{X}}_2 \ \cdots \ \tilde{\underline{X}}_k \,]
\end{aligned}$$

- Slope coefficients:

$$\begin{aligned}
\hat{\beta}_2 &= (X_2^\top M_\ell X_2)^{-1} X_2^\top M_\ell Y \\
&= (\tilde{X}_2^\top \tilde{X}_2)^{-1} \tilde{X}_2^\top Y
\end{aligned}$$

- Variation of X_2 in the sample, with $X_{i,2}$ now $(k-1) \times 1$ and $\bar{X}_2 = \frac{1}{n} \sum_{i=1}^n X_{i,2}$:

$$\begin{aligned}
\tilde{X}_2^\top \tilde{X}_2 &= \sum_{i=1}^n \tilde{X}_{i,2} \tilde{X}_{i,2}^\top \\
&= \sum_{i=1}^n (X_{i,2} - \bar{X}_2)(X_{i,2} - \bar{X}_2)^\top
\end{aligned}$$

Intercept, 1: deviations from means

- In the sample:

$$\begin{aligned}
Y_i &= \beta_1 + X_{i,2}^\top \beta_2 + U_i \\
\bar{Y} &= \beta_1 + \bar{X}_2^\top \beta_2 + \bar{U} \\
\underbrace{Y_i - \bar{Y}}_{\tilde{Y}_i} &= \underbrace{(X_{i,2} - \bar{X}_2)^\top}_{\tilde{X}_{i,2}^\top} \beta_2 + \underbrace{(U_i - \bar{U})}_{\tilde{U}_i} \\
\tilde{Y}_i &= \tilde{X}_{i,2}^\top \beta_2 + \tilde{U}_i
\end{aligned}$$

- In the population:

$$\begin{aligned}
Y_i &= \beta_1 + X_{i,2}^\top \beta_2 + U_i \\
\mathrm{E}\left[Y_i\right] &= \beta_1 + \mathrm{E}\left[X_{i,2}\right]^\top \beta_2 \\
Y_i - \mathrm{E}\left[Y_i\right] &= (X_{i,2} - \mathrm{E}\left[X_{i,2}\right])^\top \beta_2 + U_i
\end{aligned}$$

Intercept, 2: non-zero error mean

- $\mathrm{E}\left[U_i \mid X_i\right] = \mu \neq 0$:

$$\begin{aligned}
\mathrm{E}\left[Y_i\right] &= \beta_1 + \mathrm{E}\left[X_{i,2}\right]^\top \beta_2 + \mu \\
Y_i - \mathrm{E}\left[Y_i\right] &= (X_{i,2} - \mathrm{E}\left[X_{i,2}\right])^\top \beta_2 + (U_i - \mu)
\end{aligned}$$

- Then:

$$\begin{aligned}
Y_i &= \beta_1 + X_{i,2}^\top \beta_2 + U_i \\
&= \beta_1 + X_{i,2}^\top \beta_2 + (U_i - \mu) + \mu \\
&= \underbrace{(\beta_1 + \mu)}_{= \beta_1^*} + X_{i,2}^\top \beta_2 + \underbrace{(U_i - \mu)}_{= U_i^*} \\
&= \beta_1^* + X_{i,2}^\top \beta_2 + U_i^*
\end{aligned}$$

- $\mathrm{E}\left[U_i^* \mid X_i\right] = 0$.
- Estimable: β_2 and $\beta_1^* = \beta_1 + \mu$.

Summary

Projections

- $P_X = X(X^\top X)^{-1} X^\top$ and $M_X = I_n - P_X$: symmetric, idempotent, p.s.d.
- $P_X X = X$, $M_X X = 0$, $P_X M_X = 0$.
- $\mathrm{tr}(P_X) = k$ and $\mathrm{rank}\,(M_X) = \mathrm{tr}(M_X) = n - k$.
- $Y = \hat{Y} + \hat{U}$ with $\hat{Y} = P_X Y \in S(X)$, $\hat{U} = M_X Y \in S^\perp(X)$, $\hat{Y}^\top \hat{U} = 0$.

s^2 and partitioned regression

- $\hat{U} = M_X U$, $\hat{U}^\top \hat{U} = U^\top M_X U$, $\mathrm{E}\left[\hat{U}^\top \hat{U} \mid X\right] = \sigma^2 \,\mathrm{tr}(M_X) = \sigma^2 (n - k)$.
- $\mathrm{E}\left[\hat{\sigma}^2 \mid X\right] = \frac{n-k}{n} \sigma^2$; $s^2 = \hat{U}^\top \hat{U}/(n - k)$ with $\mathrm{E}\left[s^2 \mid X\right] = \sigma^2$; $\widehat{\mathrm{Var}}(\hat{\beta} \mid X) = s^2 (X^\top X)^{-1}$.
- Frisch-Waugh-Lovell: $\hat{\beta}_1 = (X_1^\top M_2 X_1)^{-1} X_1^\top M_2 Y = (\tilde{X}_1^\top \tilde{X}_1)^{-1} \tilde{X}_1^\top \tilde{Y}$, with $\tilde{X}_1 = M_2 X_1$ and $\tilde{Y} = M_2 Y$.
- Intercept, one regressor: $\hat{\beta}_2 = \sum_i (X_{i,2} - \bar{X}_2) Y_i \big/ \sum_i (X_{i,2} - \bar{X}_2)^2$.