Geometry of OLS and Frisch-Waugh-Lowell th.

↓ also known as theorem of the partitioned

$y = X\beta + \varepsilon$

$\hat{\beta} = (X'X)^{-1}X'y$

$\hat{y} = X\hat{\beta}$

$\hat{\varepsilon} = y - X\hat{\beta}$

$P_x = X(X'X)^{-1}X'$

$(N \times N)$ very large matrix

$$\boxed{P_x \cdot y = \hat{y}}$$

$M_x y = (I_N - P_x) \cdot y = \hat{\varepsilon}$

→ $\hat{y}$ is just the projection of y onto the column space of X

$P_x \cdot P_x$

$X(X'X)^{-1}\cancel{X'X(X'X)^{-1}}X'$

$X(X'X)^{-1}X'$

⇓

IDEMPOTENT MATRIX

$M_x \cdot y = \hat{\varepsilon}$

$$y = X\hat{\beta} + \hat{\varepsilon} = P_x y + M_x y = (\underset{N \times N}{P_x} + \underset{N \times N}{M_x})y$$

$y = (P_x + M_x)y$

$P_x + M_x = I_N$

the rank of the matrix P_x is k and the eigenvalues are either 1 or 0

k eigenvalues equal to 1 and $N-k$ eigenvalues $= 0$

the opposite is true for M_x → rank is $N-k$, the eigenvalues are either 1 or 0, $N-k$ are equal to 1 and k are equal to zero

$\text{trace}(P_x) = k$ $\qquad$ $\text{trace}(M_x) = (N-k)$

trace = sum (eigenvalues)

$$\underset{(N\times N)(N\times N)}{P_x M_x} = P_x \underset{(N\times N)}{(I_N - P_x)} = P_x - P_x P_x = \underset{N\times N}{(0)} \longrightarrow$$ the two matrix are orthogonal

$\hat{y}$ is then orthogonal to $\hat{\varepsilon}$

$$\hat{y} \perp \hat{\varepsilon}$$

Graphical illustration to these properties

$$y = \hat{y} + \hat{\varepsilon} \longrightarrow \hat{y} \perp \hat{\varepsilon}$$

$Z \in$ col. space (X) $\quad Z = Xw$

hence Z is generated by linear combination of X_j

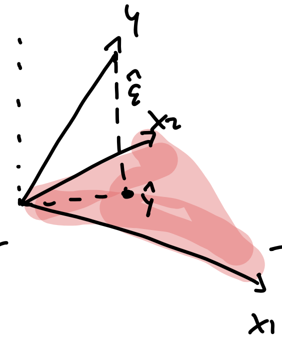
$$P_x \cdot Z = X\cancel{(X'X)^{-1}X'}Xw$$

$$= Xw = Z$$

$$P_x X = X$$

↓

projection of X on itself



→ this is the column space, is the span of that two vectors

if y is orthogonal then x_1, x_2 do not explain anything

in the opposite case $\hat{y} = y$ and the error term is zero

↓

a linear combination of x_1, x_2 can perfectly explain y

Pytagorean theorem

↓

we now derive R squared

$$y'y = \underbrace{\hat{y}'\hat{y}}_{} + \underbrace{\hat{\varepsilon}'\hat{\varepsilon}}_{} \longrightarrow$$ we can prove thanks to projection matrix

$\hat{y}'\hat{y}$ → explain of the sum of squares $\sum_{i=1}^{N} \hat{y}_i^2$

$\hat{\varepsilon}'\hat{\varepsilon}$ → $\sum_{i=1}^{N} \varepsilon_i^2$ residual sum of squares

$y'y$ = $\sum_{i=1}^{N} y_i^2$ TSS total sum of squares

$$y'y = y'(\hat{y}+\hat{\varepsilon})$$

$$= (\hat{y}'+\hat{\varepsilon}')(\hat{y}+\hat{\varepsilon})$$

$$= \hat{y}'\hat{y} + \underbrace{\hat{y}'\hat{\varepsilon}+\hat{\varepsilon}'\hat{y}} + \hat{\varepsilon}'\hat{\varepsilon}$$

they are the same and they are zero because orthogonal →

$$\hat{y}' = y'P_x$$

$$\hat{\varepsilon} = M_x \cdot y$$

$$\hat{y}'\hat{\varepsilon} = y'P_x M_x y$$

$$\parallel$$

$$0$$

Hence everything is zero

Hence

$$y'y = \hat{y}'y + \hat{\varepsilon}'\hat{\varepsilon}$$

TSS = ESS + RSS

$$\boxed{R^2 = 1 - \frac{RSS}{TSS} \quad \text{measure of fit}}$$

number $\in [0,1]$

in that case it is all error term

it is zero when the vector y is orthogonal

it is one when y is a linear combination of the xs

TSS* is computed on the recentered values of y

$$\downarrow \; (y-\bar{y})'(y-\bar{y})$$

in many packages the R^2 is calculated in this way

↓

Hence R^2 is more robust to the addition / subtraction of constants

otherwise one can make tss arbitrarily big and move the R^2 to 1

Very important result of the OLS estimator for the partitioned model

$\log(wage) = \beta_0 + \beta_1 educ + \varepsilon$ → compare 2 individuals, give one an extra year of education, watch what changes

$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 age + \varepsilon$ → β_1 is now different, compare two individuals, of the same age, the effect of an extra year of education is calculated by β_1

↓

Now the effect is CETERIS PARIBUS

FWL

$$y = x_1 \overset{k_1\times1}{\beta_1} + x_2 \overset{k_2\times1}{\beta_2} + \varepsilon \qquad \underset{N\times k}{X} = [\underset{N\times k_1}{x_1}\ \ \underset{N\times k_2}{x_2}]$$

$$y = \beta_0 + \beta_1 educ + \beta_2 age + \varepsilon$$

$\boxed{x_1 = [educ]}$ $x_2 = [iota\ ,\ age]$ → control matrix

possible way to partition the model

↓ we want to study the effect for this variable $\hat{\beta}_1$ ↓

FWL theorem

$$\hat{\beta} = \underbrace{(X'X)^{-1}X'y}_{k\times1} = \begin{array}{c|c} \left(\begin{array}{c} (X_1'M_{x_2}X_1)^{-1}(X_1'M_{x_2}y) \\ (X_2'M_{x_1}X_2)^{-1}(X_2'M_{x_1}y) \end{array}\right) & \begin{array}{c} k_1\times1 \\ k_2\times1 \end{array} \end{array}$$

→ i decompose the matrix into two submatrices

↳ $\hat{\beta}_2$

$M_{x_2} = I - X_2(X_2X_2')^{-1}X_2'$ residual maker matrix associated only related to the column space of x_2

$M_{x_2}y$ → residual of one hypothetical regression in which we regress y only on x_2 ⇒ $\hat{\varepsilon}_{x_2} = y - x_2\tilde{\beta}_2$

residuals and x_2 are orthogonal

↓

component of y that are not explained by x_2

$\tilde{\beta}_2 = (X_2'X_2)^{-1}X_2'y$

$\tilde{\beta}_1$ is the result of an hypothetical regression

the ~ is to remember that $\tilde{\beta}_2$ and $\hat{\beta}_1$ are different

$M_{X_2}\cdot X_1 \longrightarrow$ regression $X_1 = X_2\Gamma + \eta$

$\downarrow$

regressing each column of X_1 on X_2

$\hat{\eta}$ are the residuals of this regression

$\downarrow$

variability in X_1 that cannot be explained by X_2

$\hat{\eta} = M_{X_2}X_1$

variability not explained by X_2

$\downarrow$

$\hat{\beta}_1 = (X_1'M_{X_2}X_1)^{-1}X'M_{X_2}y$

$\downarrow$ it is like an OLS on transformed data

$\downarrow$ not explained by X_2

$$\boxed{M_{X_2}y = M_{X_2}X_1\beta_1 + w}$$

$\hat{\beta}_1 = (\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}'\tilde{y}$

we parted out only the effect of X_1

$= (X_1'M_{X_2}M_{X_2}X_1)^{-1}X_1'M_{X_2}M_{X_2}y$

regress residual of y on residuals of $X_1 = \hat{\eta}$

$$\boxed{\hat{\beta}_1 = (X_1'M_{X_2}X_1)^{-1}X_1'M_{X_2}y}$$

3 steps

- regress y on X_2 and obtain the residuals $\tilde{\varepsilon}_{X_2} = M_{X_2}y$
- regress X_1 on X_2 and obtain the residuals $\hat{\eta} = M_{X_2}\cdot X_1$
- regress $\hat{\varepsilon}_{X_2}$ on $\hat{\eta} \longrightarrow \hat{\beta}_1$

regress two things that does not depend on X_2

$\downarrow$

we lose the net effect of an extra X_1

$\downarrow$

we have isolated the effect

none of the two depends anymore on X_2 !!

$\downarrow$

so we are focusing on X_1

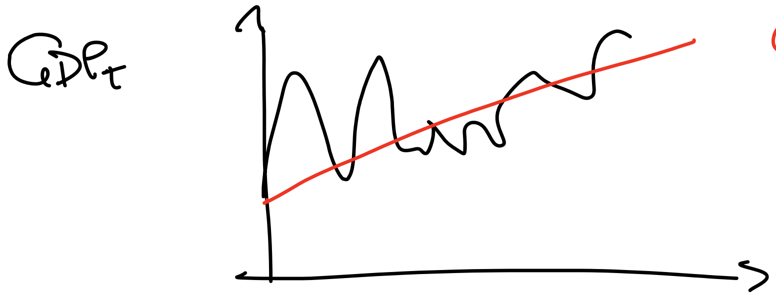
normal is

$\beta = (X'X)^{-1}X'y$

$\downarrow$

regress y on X

now we are regressing $\bigcirc\!\!\!\!M_{X_2}X_1$ on $M_{X_2}y$ $\rightarrow$ residuals on residuals



$GDP_t = \beta_0 + \beta_1 t + \varepsilon_\tau$

linear trend

$GDP_\tau^* = GDP_\tau - \hat{\beta}_0 - \hat{\beta}_1 \tau$

$\tau = \{1 ... T\}$

this is to de-trend

$$GDP_\tau = \beta_0 + \beta_1 \tau + \boxed{\beta_2}\pi_\tau + \varepsilon_\tau$$

this is a more difficult regression

if also inflation is subject to trend (we are analyzing β_2) do i need to detrend GDP and π to account for the effect of β_2

NO, there is no need to run also their regressions

How to prove this theorem?

$$\hat{\beta} = \arg\min_{\beta} (y - x\beta)'(y - x\beta)$$ this is the problem of OLS

Now i want to split in two

$$[\hat{\beta}_1 | \hat{\beta}_2] = \arg\min_{\beta_1, \beta_2} (y - x_1\beta_1 - x_2\beta_2)'(y - x_1\beta_1 - x_2\beta_2)$$

if i am interested only in β_1

$$\hat{\beta}_1 = \arg\min_{\beta_1} \left(\min_{\beta_2} (y - x_1\beta_1 - x_2\beta_2)'(y - x_1\beta_1 - x_2\beta_2) \right)$$

↳ if i solve for β_2 then the problem will only depend on β_1

it is like splitting the problem into two parts

i treat β_1 as known

Regress $y - x_1\beta_1$ on x_2 in the inner problem residual of this regression is $M_{x_2}(y - x_1\beta_1)$

$\downarrow$

the minimum is reached in the sum of squared residuals

FWL
the regression coefficient of the multivariate regression is equal to the regression coefficient in a bivariate model where residualised outcome is regressed on the residualised component of the predictor

$$\boxed{(y - x_1\beta_1)'\underbrace{M_{x_2}M_{x_2}}_{=M_{x_2}}(y - x_1\beta_1)}$$

only depends on β_1

$\Downarrow$

$$(y - x_1\beta_1)'M_{x_2}(y - x_1\beta_1)$$

$\searrow$ without M_{x_2} the problem is the same of a simple regression

Now the outer problem is a problem in β_1

$$\hat{\beta}_1 = \underset{\beta_1}{\arg\min}\left[(y - x_1\beta_1)'M_{x_2}(y - x_1\beta_1)\right]$$

$$y'M_{x_2}y - \beta_1'x_1'M_{x_2}y - y'M_{x_2}x_1\beta_1 + \beta_1'x_1'M_{x_2}x_1\beta_1$$

derivative with respect to β_1

$$0 - \cancel{2}x_1'M_{x_2}y + \cancel{2}(x_1'M_{x_2}x_1)\hat{\beta}_1 = 0$$

$$\boxed{\hat{\beta}_1 = (x_1'M_{x_2}x_1)^{-1}x_1'M_{x_2}y}$$

Proof of the theorem

the claim is

$$\boxed{\begin{aligned} y &= x_1\hat{\beta}_1 + x_2\hat{\beta}_2 + \hat{\varepsilon} \\ M_1 y &= M_1 x_2\hat{\beta}_2 + \hat{\varepsilon} \end{aligned}}$$

Example

$$sales = \beta_0 + \hat{\beta}_1 coupon + \hat{\beta}_2 income + \hat{\varepsilon}$$

if we only regress sales on coupon, we may violate the error to exclude the fact that people with high income maybe use less coupon and spend generally more

↓

income has an effect on sales which is now passing through coupons, we want to partial out this effect

↓

we are "partialling out" the effect of income on sales

any predictor in a multivariate regression of k coefficients explains variations of y not explained by the other k-1 ones neither by their relation with that predictor

$$Y = x_1\hat{\beta}_1 + x_2\hat{\beta}_2 + \hat{\varepsilon}$$

we want to partial out the effect of β_1

$$M_1 Y = M x_2 \hat{\beta}_2 + \hat{\varepsilon}$$

→ residual from regressing x_2 on x_1

↓

residuals from regression of y on x_1