Linear Regression

data: $\{y_i, k_i, l_i\}_{i=1}^{N}$

$$\exp(y_i) = \exp(k_i)^{\alpha} \exp(l_i)^{\beta} \exp^{\varepsilon_i}$$

Possibility 1: ε_i independent of k_i, l_i

* Possibility 2: "Mean independence"

$$\mathbb{E}[\varepsilon \mid k, l] = 0$$

$N \times 1$ vectors

$y_i = \alpha k_i + \beta l_i + \varepsilon_i$ for all i

Rewrite in matrix form

$$Y = X\theta + \epsilon$$

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \quad N \times 1 \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

$$X = \begin{bmatrix} k_1 & l_1 \\ \vdots & \vdots \\ k_N & l_N \end{bmatrix} \quad N \times 2$$

$$\theta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

$$\varepsilon = Y - X\theta$$

$$\mathbb{E}[Y - X\theta \mid X] = \mathbb{E}[\varepsilon \mid X] = 0$$

$$X' \mathbb{E}[Y - X\theta \mid X] = 0$$

$\mathbb{E}[g(w) f(z) \mid z]$

$= f(z) \mathbb{E}[g(w) \mid z]$

$\mathbb{E}[f(z) \mid z]$

$$E[X'Y - X'X\theta \mid X] = 0$$

$= f(z)$ a.s.

$$E[X'Y] = (X'X)\theta$$

2×1 matrix

$$\theta = \underbrace{(X'X)^{-1}}_{2\times 2} E[\underbrace{X'Y}_{2\times 1} \mid X]$$

Assume X is full rank

$$X = \begin{bmatrix} k_1 & l_1 \\ k_2 & l_2 \\ \vdots & \vdots \\ k_N & l_N \end{bmatrix} \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \quad X'Y = \begin{bmatrix} \sum_{i=1}^{N} k_i Y_i \\ \sum_{i=1}^{N} l_i Y_i \end{bmatrix}$$

$$X'X = \begin{bmatrix} \sum_{i=1}^{N} k_i^2 & \sum_{i=1}^{N} k_i l_i \\ \sum_{i=1}^{N} k_i l_i & \sum_{i=1}^{N} l_i^2 \end{bmatrix}$$

Propose an estimator

$$\hat{\theta} = (X'X)^{-1} X'Y$$

Ordinary Least Squares (OLS)

Prod expectation.

Applies in general to the model

$$y = X\beta + \varepsilon$$

with y $N\times 1$, X $N\times K$, β $K\times 1$

and $E[\varepsilon \mid X] = 0$

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

Properties: 1: Unbiasedness

An estimator $\hat{\Theta}$ of Θ is unbiased if

$$E[\hat{\Theta}] = \Theta$$

"finite sample property"

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

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

$$= \underbrace{(X'X)^{-1} X'X}_{I_K} \beta + (X'X)^{-1} X'\varepsilon$$

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

$$E[\hat{\beta}] = E_X[E[\hat{\beta} | X]]$$

Law of iterated expectations

L.I.E.

$$= \beta + E_X[E[(X'X)^{-1} X'\varepsilon | X]]$$

$$= \beta + E_X[(X'X)^{-1} X' E[X'\varepsilon | X]]$$

$$E[X'\varepsilon | X] = \begin{bmatrix} E[\sum_{i=1}^{N} k_i \varepsilon_i | X] \\ E[\sum_{i=1}^{N} l_i \varepsilon_i | X] \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} E[k_i \varepsilon_i | X] \\ \sum_{i=1}^{N} E[l_i \varepsilon_i | X] \end{bmatrix}$$

use $E[\varepsilon | X] = 0$

$$= X' E[\varepsilon | X] = X' 0 = 0$$

$$E[\hat{\beta}] = \beta$$, which is unbiasedness

Property 2: Variance

$$\mathrm{Var}(Z) := E[(Z - E[Z])(Z - E[Z])']$$

if Z is a vector

$$[\mathrm{Var}(Z)]_{jj} = \mathrm{Var}(Z_j) = E[(Z_j - E[Z_j])^2]$$

$$[\mathrm{Var}(Z)]_{jk} = \mathrm{Cov}(Z_j, Z_k) = E[[Z_j - E[Z_j]]\cdot[Z_k - E[Z_k]]]$$

Conditional variance

$$\mathrm{Var}(Z|W) := E[(Z - E[Z|W])(Z - E[Z|W])' | W]$$

Conditional variance of $\hat{\beta}$

Note $E[\hat{\beta}|X] = \beta$

$$\mathrm{Var}(\hat{\beta}|X) = E((\hat{\beta} - \beta)(\hat{\beta} - \beta)'|X]$$

$$= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}|X]$$

$$= (X'X)^{-1}X' \underbrace{E[\varepsilon\varepsilon'|X]}_{\Sigma} X(X'X)^{-1}$$

Σ conditional variance of the residuals

- Common to assume a structure for Σ

Example: $\Sigma = \sigma^2 I_N$ Homoskedasticity

says $\{\varepsilon_i\}_{i=1}^N$ are uncorrelated

$$I_N = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & & 1 \end{pmatrix}$$ and have $var(\varepsilon_i) = \sigma^2$ for all i

Then $$var(\hat{\beta}|X) = (X'X)^{-1} X' \sigma^2 I X (X'X)^{-1}$$
$$= \sigma^2 (X'X)^{-1} X'X (X'X)^{-1}$$
$$= \sigma^2 (X'X)^{-1}$$

Property: OLS prediction

We can use OLS to predict Y
Call vector of predictions $\hat{Y}$

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

orthogonal projection matrix

Properties:

Symmetry: $P_X = P_X'$

adjective: idempotent

Idempotence: $P_X P_X = P_X$

$(AB)' = B'A'$

Proofs:

- $P_X' = X\left[(X'X)^{-1}\right]' X'$

use $(A^{-1})' = (A')^{-1}$

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

- $P_X P_X = X(X'X)^{-1} \underbrace{X'X(X'X)^{-1}} X'$

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

Properties of P_X

Define vector of "predicted residuals" $\hat{\varepsilon} = \begin{pmatrix} \hat{\varepsilon}_1 \\ \vdots \\ \hat{\varepsilon}_N \end{pmatrix}$

$$= Y - \hat{Y} = Y - P_X Y = (I - P_X) Y$$

$$:= M_X Y \qquad \text{"residualizing matrix"}$$

M_X is also a projection

$$M_X M_X = (I - P_X)(I - P_X)$$

$$= I - 2P_X + P_X P_X$$

$$= I - 2P_X + P_X = I - P_X = M_X$$

$$M_X' = (I - P_X)' = I' - P_X' = I - P_X$$

$$Y = \underbrace{P_X Y}_{\hat{Y}} + \underbrace{M_X Y}_{\hat{\varepsilon}}$$

because this is $P_X Y + (I - P_X) Y$

$$= Y + (P_X - P_X) Y$$

$$= Y$$

$\hat{Y}$ and $\hat{\varepsilon}$ are orthogonal

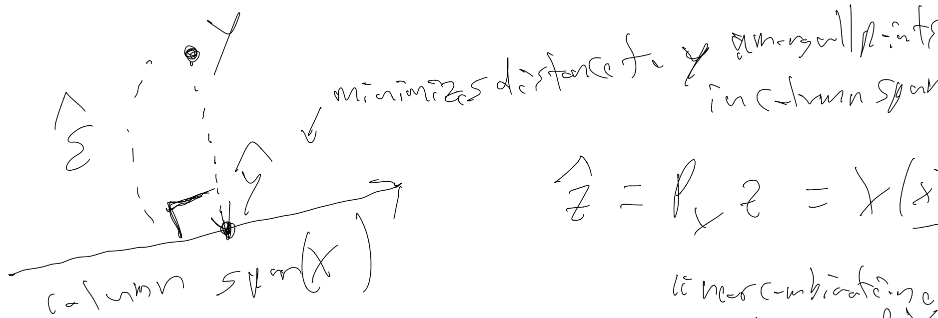
$$\Sigma_{i=1}^N \hat{Y}_i \hat{\varepsilon}_i = \hat{Y}' \hat{\varepsilon} = (P_X Y)' M_X Y$$

$$= Y' P_X' M_X Y = Y' P_X M_X Y$$

$$= Y'(P_X - P_X P_X) Y$$

$= y'(P_X - P_X)y = 0$

so residuals are orthogonal to $\hat{y}$



$\hat{z} = P_X z = X\underbrace{(X'X)^{-1}X'z}_{\text{vector}}$

linear combination of columns of X

Optimality of $\hat{\beta}$

Gauss-Markov Theorem

"OLS is BLUE"

Best Linear Unbiased Estimator

- OLS is optimal within a limited class (linear unbiased estimators" by the measure of "smallest variance"

We use the positive semidefinite partial order on matrices to rank variance

Symmetric $A \in \mathbb{R}^{k \times k}$ is positive semidefinite if $v'Av \geq 0$ for all v in $\mathbb{R}^k$

any linear combination of $v'\hat{\beta}$ has weakly smaller variance

A is positive definite if $v'Av > 0$ for all $v \neq 0$

Partial order $A \geq B$ in partial order

if $A - B$ is positive semidefinite

Will show $Var(\hat{B}^{OLS}) \leq Var(\tilde{B})$

for any $\tilde{B}$ in a class

Means variance of any linear combination is weakly smaller in variance

Condition: $\Sigma = \sigma^2 I$ homoskedasticity

Linear estimators are of form

$$\tilde{B} = Q(X)Y$$. OLS has $Q(X) = (X'X)^{-1}X'$

Can always write

$$Q(X) = (X'X)^{-1}X' + R'$$

by defining $R' = Q(X) - (X'X)^{-1}X'$

Let $\tilde{B}$ also be unbiased: $E[\tilde{B}|X] = B$

Then $E[\tilde{B}|X] = E[(X'X)^{-1}X'Y + R'Y|X]$

$$= E[(X'X)^{-1}X'(XB+\varepsilon) + R'(XB+\varepsilon)|X]$$

$$= B + R'XB$$

for $\tilde{B}$ to be unbiased need $E[\tilde{B}] = B$

so need $R'XB = 0$

For this to be true for any β
need $R'X = 0$

$$Var(\hat{\beta} | X) = E\Big[\underbrace{[(X'X)^{-1}X' + R']\varepsilon}_{\hat{\beta}-\beta} \underbrace{\varepsilon'(R + X(X'X)^{-1})}_{(\hat{\beta}-\beta)'} \Big| X\Big]$$

$$= \sigma^2 \big((X'X)^{-1}X' + R'\big)\big(R + X(X'X)^{-1}\big)$$

$$= \sigma^2\big[(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{I} + R'R\big]$$

$$= \sigma^2(X'X)^{-1} + \sigma^2 R'R$$

Compare $Var(\hat{\beta}^{OLS} | X) = \sigma^2(X'X)^{-1}$

so $Var(\hat{\beta} | X) - Var(\hat{\beta}^{OLS} | X) = \sigma^2 R'R$

but now $\sigma^2(v'R')Rv$ is a quadratic form and so ≥ 0

so $Var(\hat{\beta} | X) \geq Var(\hat{\beta} | X)$

in positive semidefinite partial order.

Estimate σ^2 to know what variance is

σ^2 is the variance of ε_i

$$\sigma^2 = E(\varepsilon_i - E(\varepsilon_i))^2 = \frac{1}{n}E(\varepsilon - E[\varepsilon])'(\varepsilon - E[\varepsilon])$$

$$= \frac{1}{n}E[\varepsilon'\varepsilon]$$

$$= \frac{1}{n} E\left[(y - XB)'(y - XB)\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^{n} (y_i - x_i' B)^2\right]$$

Propose: sample version (get rid of expectation and replace B by $\hat{B}$)

Try $\hat{\sigma}^2 = \frac{1}{n} (y - X\hat{B})'(y - X\hat{B})$

$$= \frac{1}{n} (M_X y)' M_X y = \frac{1}{n} y' M_X' M_X y$$

$$= \frac{1}{n} y' M_X y$$

This is a biased estimator and so we will want to modify it

Proof: $E[\hat{\sigma}^2 | X] = n^{-1} E[y' M_X y | X]$

(note $M_X y = M_X(X B + \varepsilon) = M_X \varepsilon$)

$$= n^{-1} E[\varepsilon' M_X \varepsilon | X]$$

Use properties of Trace: if A is a square matrix

$$Tr(A) = \sum_i A_{ii} \quad \text{sum of diagonal entries}$$

Trace is linear: $Tr(aA + bB) = a\,Tr(A) + b\,Tr(B)$

and has cyclic property $Tr(AB) = Tr(BA)$ when conformable

meaning AB and BA are of dimension to be well defined.

$\varepsilon' M_X \varepsilon$ is 1×1 scalar so $= Tr(\varepsilon' M_X \varepsilon)$

$$E[\varepsilon' M_X \varepsilon \mid X] = E[Tr(\varepsilon' M_X \varepsilon) \mid X]$$

$$= E[Tr(M_X \varepsilon\varepsilon') \mid X] \quad \text{by cyclic property}$$

$$= Tr(M_X\, E[\varepsilon\varepsilon' \mid X])$$

$$= Tr(M_X \sigma^2 I_N) = \sigma^2 Tr(M_X)$$

$$Tr(M_X) = Tr(I_N - P_X)$$

$$= N - Tr(P_X)$$

note $Tr(P_X) = Tr(X(X'X)^{-1}X')$

$$= Tr((X'X)^{-1}X'X) \quad (\text{cyclic})$$

$$= Tr(I_K) \quad \text{since } X \text{ is } N\times K$$

$$= K$$

$$E[\hat{\sigma}^2 \mid X] = \frac{n-k}{n}\sigma^2$$

So we use instead $\tilde{\sigma}^2 = \dfrac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k}$ which is $\dfrac{n}{n-k}\hat{\sigma}^2$

This is un biased!

This is the residual variance estimate

Can use $\tilde{\sigma}^2 (X'X)^{-1}$ as estimate of

$$\operatorname{Var}(\hat{\beta} \mid X).$$