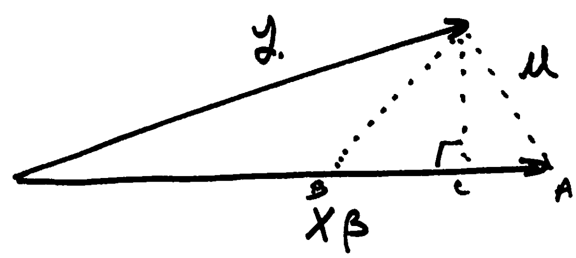
Orthogonal Projections

A *Projection* maps each point of E^n into a point in a subspace of E^n, while leaving all points in that space unchanged.



Orthogonal Projection maps a point to the point in the subspace that is closest to it. (e.g. C)

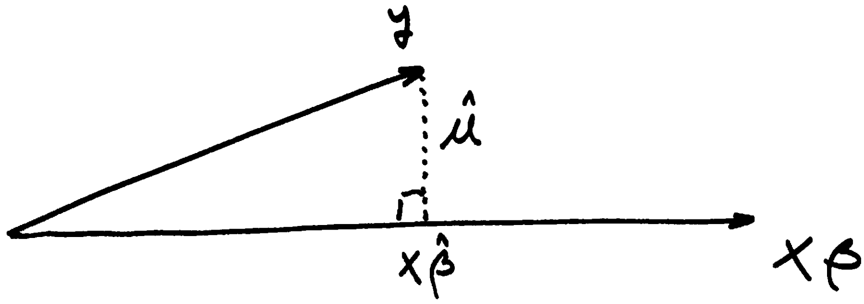
Algebraically, this can be accomplished by premultiplying the vector to be projected by a suitable *Projection Matrix*.

IN OLS The ORTHOG Projection Matrix is

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

$$P_X \cdot y = X(X^T X)^{-1} X^T y = X\hat{\beta}$$

where $\hat{\beta}$ is OLS.



$\hat{u}$ is obtained using the <u>orthogonal complement</u> to P_X

$$M_X = I - X(X^T X)^{-1} X^T$$

$P_X \cdot M_X = 0$

$M_X \cdot P_X = 0$

$\Rightarrow$ ORTHOgonal. "Annihilate" each other.

$$\begin{aligned} M_X y &= \left(I - X(X^T X)^{-1} X^T\right) y \\ &= y - X(X^T X)^{-1} X^T y \\ &= y - X\hat{\beta} = \hat{u} \end{aligned}$$

M_X is sometimes referred to as the "residual maker" matrix.

It removes from y everything that is correlated with X.

Terms:

P_x "projects onto" $S(X)$

M_x " " off of" $S(X)$

ORTHOG.
Projection matrices have 2 properties

— symmetric $P_x = P_x^T \quad M_x = M_x^T$

— Idempotent. $P_x P_x = P_x \quad M_x M_x = M_x$.

NOTE ALSO.

$$P_x + M_x = I$$

$$(P_x + M_x = I)\,\underset{\sim}{y}$$

$$P_x y + M_x y = \underset{\sim}{y} \qquad (*)$$

These are said to be <u>complementary</u> <u>projections</u> since $P_x + M_x = I$

they restore $\underset{\sim}{y}$.

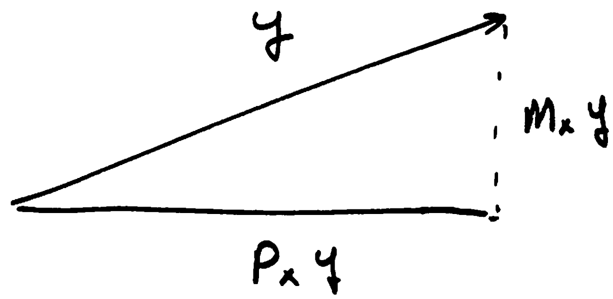
2.9 - 2.12

$*$ is sometimes called the <u>orthogonal</u> <u>decomposition</u> of $\underset{\sim}{y}$.

y $n \times 1$

$P_x y$ projects onto the k demensoinal subspace $S(X)$

$M_x y$ projects onto the $n-k$ demensoinal orthog complement $S^{\perp}(X)$.



By Pythagoras' Theorem

$$\|P_x y\|^2 + \|M_x y\|^2 = \|y\|^2$$

or

$$\hat{\beta}' X^T X \hat{\beta} + \hat{u}^T \hat{u} = y^T y$$

And it follows that

$$\|y\|^2 \geq \|P_x y\|^2$$

TSS $\geq$ ESS

General

P projects onto ______

M " off of ______

P_x projects onto $S(X)$ i.e.;

P_z " " $S(Z)$

$P_{X,W}$ projects onto $S(X,W)$

use Basis vectors of X & W

FWL

$$\underset{\sim}{y} = \underset{n \times k_1}{X_1} \beta_1 + \underset{n \times k_2}{X_2} \beta_2 + \underset{\sim}{u}$$

$$X = [X_1 \vdots X_2] \qquad k_1 + k_2 = k.$$
$n \times k.$

$$\beta = \begin{pmatrix} \beta_1 \\ - \\ \beta_2 \end{pmatrix}$$

OLS $\quad \hat{\beta} = (X^T X)^{-1} X^T \underset{\sim}{y} = \begin{pmatrix} \hat{\beta}_1^* \\ - \\ \hat{\beta}_2^* \end{pmatrix}$

~~FWL~~

Consider the "Residual Maker" M_1

$$M_1 = I - X_1 (X_1^T X_1)^{-1} X_1^T$$

and the Regression

$$M_1 \underset{\sim}{y} = M_1 X_2 \beta_2 + \text{res.}$$

$\hat{\beta}_2 =$

$$\hat{\beta}_2 = (X_2^T M_1^T M_1 X_2)^{-1} X_2^T M_1' M_1 \underset{\sim}{y}$$

$$= (X_2^T M_1 X_2)^{-1} X_2^T M_1 \underset{\sim}{y}$$

FWL $\Rightarrow \quad \beta_2^* = \hat{\beta}_2 \qquad$ also, residuals from both reg. are the same.

FWL

$$\underset{\sim}{y} = X_1\underset{\sim}{\beta_1} + X_2\underset{\sim}{\beta_2} + \underset{\sim}{u}$$

recall
$$I = P_X + M_X$$
$$\underset{\sim}{y} = P_X y + M_X y$$
$$= X\hat{\beta} + \hat{u}$$
$$= X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + M_X \underset{\sim}{y}$$

Premultiply by $X_2^T M_1$

$$X_2^T M_1 \underset{\sim}{y} = X_2^T M_1 X_1 \hat{\beta}_1 + X_2^T M_1 X_2 \hat{\beta}_2 + \cancel{M_X}$$
$$X_2^T M_1 M_X \underset{\sim}{y}$$
$$= X_2^T M_1 X_2 \hat{\beta}_2 + 0 + \circledast$$

$$X_2^T M_X \underset{\sim}{y}$$

$M_1 X_1 = 0$
$M_1 M_X = M_X$

$$= X_2^T M_1 X_2 \hat{\beta}_2 + \underbrace{X_2^T \hat{u}}_{=0 \text{ since } X^T\hat{u}=0}$$

$$\left(X_2^T M_1 X_2\right) \hat{\beta}_2 = X_2^T M_1 y$$

$$\hat{\beta}_2 = \left(X_2^T M_1 X_2\right)^{-1} X_2^T M_1 y$$

Also, the residuals from the 2 sets of regressions will be the same.

Applications

1) Seasonal dummies

$$\begin{array}{cc} & \begin{array}{cccc} d_1 & d_2 & d_3 & d_4 \end{array} \\ X_1 = & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ & & & \end{pmatrix} \end{array}$$

Remove seasonal variation in y & X_2 using M_1

(2) Time Trends

$$y = \gamma_1 + \gamma_2 t + \gamma_3 t^2 + X\beta + u.$$

$$X_1 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ \vdots & \vdots & \vdots \\ 1 & n & n^2 \end{pmatrix}$$

$$M_1 = I - X_1 (X_1^T X_1)^{-1} X_1^T$$

$$M_1 y \quad , \quad M_1 X_2$$

(3) Deviation from mean form

$$X_1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$M_1 y$ typical element $y_i - \bar{y}$

$M_1 X_2$ typical column elem $x_{ik} - \bar{x}_k$

(4) R^2 centered & uncentered.

In linear regression with an intercept.

$$\|y\|^2 = \|P_x y\|^2 + \|M_x y\|^2$$

$$y^T y = \hat{\beta}^T X^T X \hat{\beta} + \hat{u}^T \hat{u}$$

$$TSS = ESS + SSR$$

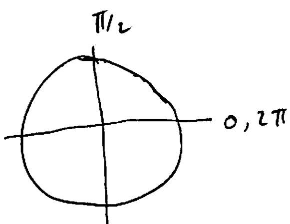
$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS} = \cos^2(\theta)$$

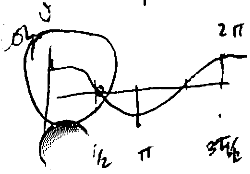
$$-1 \le \cos(\theta) \le 1 \Rightarrow 0 \le \cos^2(\theta) \le 1$$

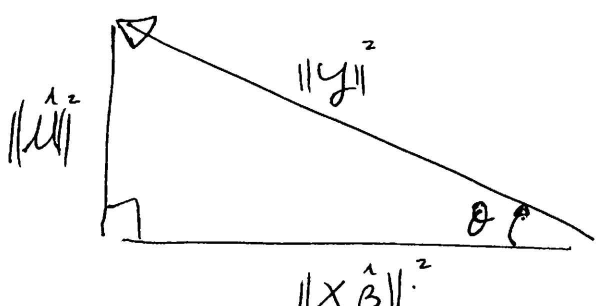
$\cos(\theta) = 0 \Rightarrow$ ~~Parallel~~ - ~~perfect fit~~

or Orthog. No Fit

$\cos(\theta) = 1 \Rightarrow$ perfect Fit.







$\|y\| \frac{\|P_x y\|}{\|y\|}$

$\frac{O}{H}$

$\cos\theta = \frac{A}{H} = \frac{\|X\hat{\beta}\|^2}{\|y\|^2} = \cos(\theta)$

$\theta \to 0 \quad \|\hat{u}\|^2 \to 0 \quad R^2 \to 1$

$\cancel{\cos\theta} = 0$

$\cos(0) = 1 \qquad \cos(1) =$

Also, for Right Triangle

$$\cos^2(\theta) + \sin^2(\theta) = 1$$

$$\Rightarrow \quad 1 - \sin^2(\theta) = \cos^2(\theta) = R^2$$

$$1 - \sin^2(\theta) = R^2$$

$$O/H$$

$$1 - \frac{\|u\|^2}{\|y\|^2} = 1 - \frac{SSR}{TSS}$$

Only True for Right Triangle

<u>Sometimes</u>

R_a^2 is invariant to changes in the scale of the variables or anything else that leaves the angle between y and $X\hat{\beta}$ unchanged.

So, we can multiply y & x's By constants and we are alright.

you cannot, however, add things to them.

$$y + \alpha \underset{\sim}{i} = X\beta + u.$$

$$P_X \; y + \alpha \underset{\sim}{i} = P_x(y + \alpha \underset{\sim}{i}) + M_x(y + \alpha \underset{\sim}{i})$$
$$= P_x \underset{-}{y} + \alpha \underset{\sim}{i} + M_x \underset{\sim}{y}$$

which follows from the fact that

$$P_X \underset{\sim}{i} = \underset{\sim}{i} \;, \quad M_X \underset{\sim}{i} = 0$$

when X includes a constant.

$$R_u^2 = \frac{\| P_x \underset{\sim}{y} + \alpha \underset{\sim}{i} \|^2}{\| y + \alpha \underset{\sim}{i} \|^2}$$

Letting $\alpha \to \infty$ we can make this as big as we wish. As αi term dominates the $\tilde{y}$. R^2 is big but it's all due to the constant!

Centered R^2

center y and X according to FWL then leaves residuals & estimates unchanged.

$$M_{\underset{\sim}{i}}(y + \alpha i) = M_i (X\beta) + \text{res.}$$

$$M_i \underset{\sim}{y} = M_i X\beta + \text{res.}$$

R_c^2

Then the second regress use P_x

$M_i Y = P_x M_i y + M_x M_i y$ $\quad R_c^2 = \dfrac{\| P_x M_i \underset{\sim}{y} \|^2}{\| M_i y \|^2} =$

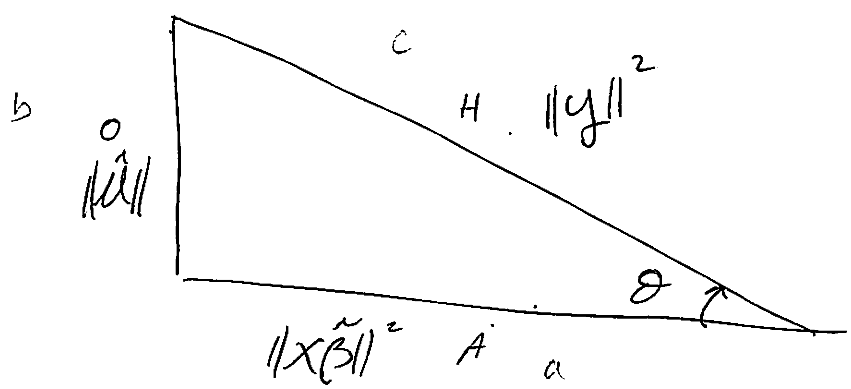
R^2 — use only with OLS & intercept
Be

Suppose something other than
OLS is used

Law of cosines

$$\|y\|^2 = \|X\hat{\beta}\|^2 + \|\hat{u}\|^2 - 2\|X\beta\|\|\hat{u}\|\cos\theta$$
$$-2\hat{y}^T\hat{u}$$

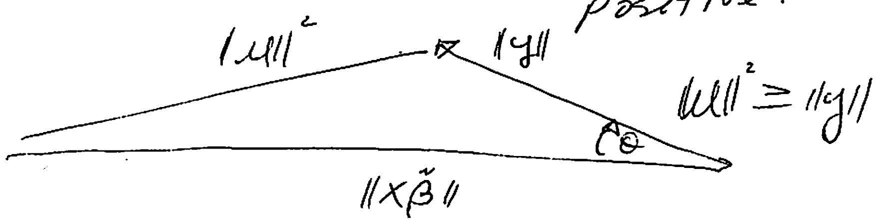
$\pm$



$$\frac{\|X\tilde{\beta}\|^2}{\|y\|^2} \neq 1 - \frac{\|y - X\tilde{\beta}\|^2}{\|y\|^2}$$

would be > 1

May not be positive.



$$P_X P_1 = P_X X_1 (X_1^T X_1)^{-1} X_1^T$$

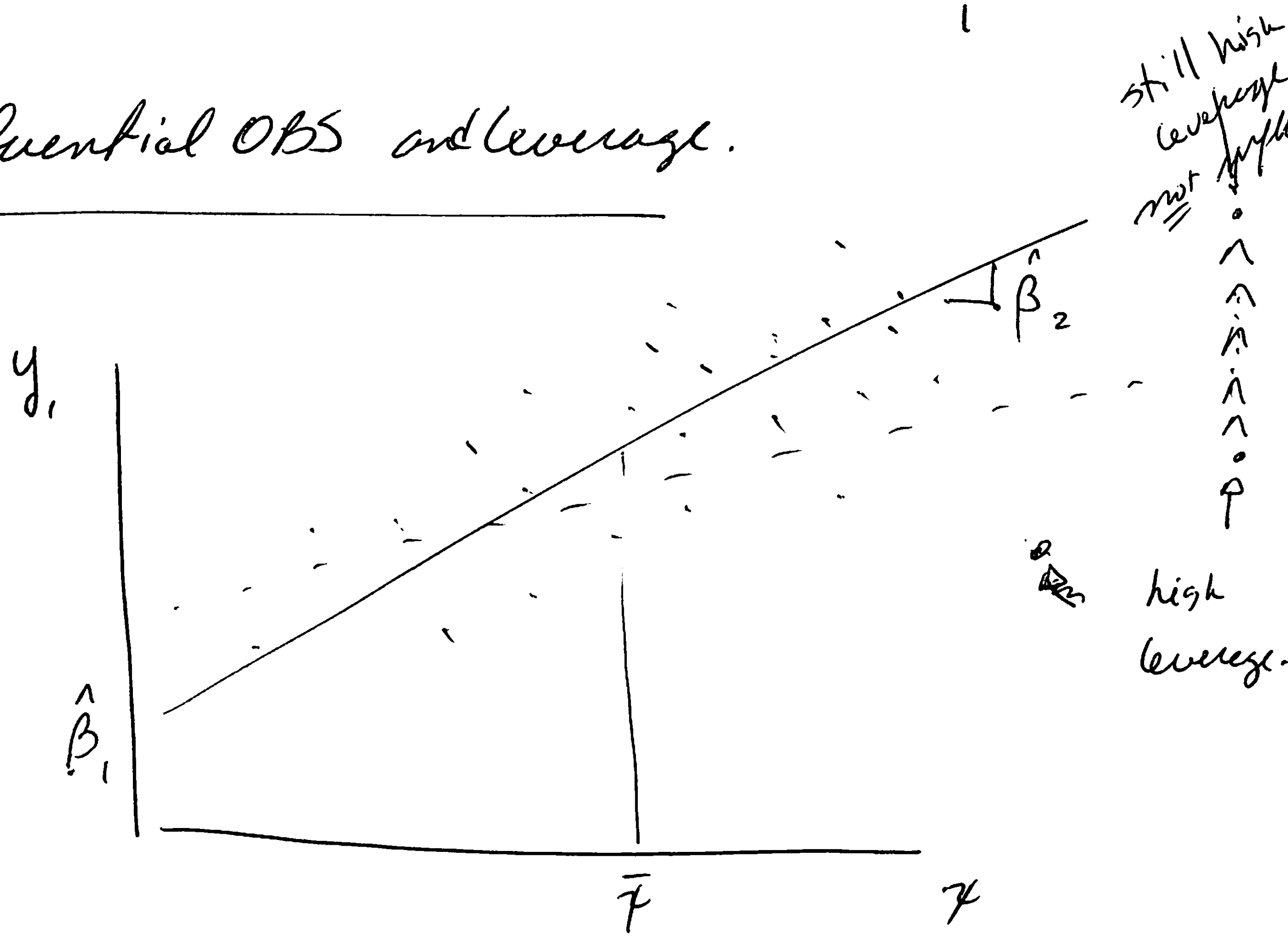
where $P_X = X(X^T X)^{-1} X^T$

Note: $P_X X = X$

this implies that projecting X onto itself yields X. X_1 is a subset of the cols of X. $\therefore$ $P_X X_1 (X_1^T X_1)^{-1} X_1^T = P_1$

$$X_1 (X_1^T X_1)^{-1} X_1^T$$

Influential OBS and leverage.



If the high leverage point is included in the regression, then the estimated slope will be smaller.

High leverage points are distant from $\bar{x}$
They have the potential to be influential if they also are a long way from the other points near the reg line in the y direction

i.e., if their omission changes the slope estimate $\Rightarrow$ Influential.

(Ia) $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{u}$

(IIa) $y_t = \beta_1 + \beta_2 X_{t2} + u_t$

Estimate (I) using all OBS
$= \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$

Then omit OBS t (high leverage)
and reestimate $= \hat{\beta}^{(t)}_{(B)}$
How much does $\hat{\beta}$ differ from $\hat{\beta}^{(t)}$?

Trick

Let $\underset{\sim}{e}_t = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ where 1 is in the t^{th} row — zero elsewhere.

$\underset{\sim}{y} = X\beta + \alpha \underset{\sim}{e}_t + \underset{\sim}{u}$

By including $\underset{\sim}{e}_t$ in the original model, we effectively remove the effect of OBS t on the results

OLS of $\underset{\sim}{\beta}$ yields $\hat{\beta}^{(t)}$!

Remove $\underset{\sim}{e}_t$ from results using FWL

$$M_t = I - \underset{\sim}{e}_t(\underset{\sim}{e}_t^T \underset{\sim}{e}_t)^{-1} \underset{\sim}{e}_t^T$$

$$M_t \underset{\sim}{y} = M_t X\beta + \alpha \overset{=0}{M_t e_t} + M_t \underset{\sim}{u}$$
$$= M_t X\beta + M_t \underset{\sim}{u}$$

$$M_t \underset{\sim}{y} = (I - \underset{\sim}{e}_t(\underset{\sim}{e}_t^T \underset{\sim}{e}_t)^{-1} \underset{\sim}{e}_t^T)\, y$$

$$\underset{\sim}{y} - \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 \cdots 0 \\ \vdots & & & & \\ 0 & & & & 0 \end{pmatrix} \underset{\sim}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ 0 \\ y_{t+1} \\ \vdots \\ y_n \end{pmatrix}$$

~~y_{t+1} etc~~

The t^{th} row of $M_t y = 0$!

Does the same to X!

Let $Z = [X \vdots \underset{\sim}{e}_t]$

$$P_Z = Z(Z^T Z)^{-1} Z^T$$
$$M_Z = I - P_Z$$

$$y = P_Z y + M_Z \underset{\sim}{y} = X\hat{\beta}^{(t)} + \hat{\alpha}\underset{-}{e}_t + M_Z \underline{y}$$

Premult by $\tilde{P}_X$

$$P_X \underset{\sim}{y} = P_X X\hat{\beta}^{(t)} + \hat{\alpha} P_X \underset{\sim}{e}_t + \underset{=0}{P_X M_Z \underset{\sim}{y}} \qquad P_X \hat{u} = 0$$

$$P_x \underset{\sim}{y} = X\hat{\beta} \quad \therefore$$

set the two equal.

$$X\hat{\beta} = P_x X\hat{\beta}^{(t)} + \hat{\alpha} P_x \underset{\sim}{e}_t$$

$$X\hat{\beta} = X\hat{\beta}^{(t)} + \hat{\alpha} P_x \underset{\sim}{e}_t$$

$$X(\hat{\beta} - \hat{\beta}^{(t)}) = \hat{\alpha} P_x \underset{\sim}{e}_t$$

<u>Find $\hat{\alpha}$</u>

Fin. 2/17/09

Use FWL. Remove X from Regression

$$M_x y = M_x X \beta^{(t)} + \alpha M_x \underset{\sim}{e}_t + M_x \underset{\sim}{u}$$

$$M_x \underset{\sim}{y} = \alpha M_x \underset{\sim}{e}_t + \text{res.}$$

Use OLS

$$\hat{\alpha} = (\underset{\sim}{e}_t^T M_x^T M_x \underset{\sim}{e}_t)^{-1} \underset{\sim}{e}_t^T M_x^T M_x \underset{\sim}{y}$$

$$\hat{\alpha} = \frac{\underset{\sim}{e}_t^T M_x \underset{\sim}{y}}{\underset{\sim}{e}_t^T M_x \underset{\sim}{e}_t} = \frac{\hat{u}_t}{\underset{\sim}{e}_t^T M_x \underset{\sim}{e}_t}$$

$M_x y = \hat{u}$ from all OBS

$\underset{\sim}{e}_t \hat{u}$ <u>picks</u> of t^{th} residual from original.

$$\underset{\sim}{e}_t^{*T} M_x \underset{\sim}{e}_t =$$

$$(0\ 0\ \cdot\ \cdot 0\ 1\ 0\ \cdot\cdot 0)\ M_x \begin{pmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}$$

Picks off the t^{th} DIAG element of $M_x \overset{= I - P_x}{=} 1 - h_t$

h_t is t^{th} Diag element of $X(X^T X)^{-1} X^T = P_x$

$X_t^T (X^T X)^{-1} X_t$

P_x is "hat Matrix" since.

X_t is t^{th} row of X.

$$P_x \underset{\sim}{y} = \hat{y}$$

$$\hat{\alpha} = \frac{\hat{u}_t}{1 - h_t}$$

$P_x e_t$

So, $X(\hat{\beta} - \hat{\beta}^{(t)}_*) = \frac{\hat{u}_t}{1-h_t} X(X^TX)^{-1}X^T e_t$

Premult By ~~$X(t)$~~ $(X^TX)^{-1}X^T$

$$(\hat{\beta} - \beta^{(t)}) = \frac{\hat{u}}{1-h_t} (X^TX)^{-1}\underbrace{X^T e_t}_{k\times 1 \,=\, t^{th} \text{ row of } X.}$$

Influence Depends on $\hat{u}_t, h_t$

when h_t is large, then $1-h_t$ small

$h_t \to 1$ $\hat{\alpha} \to \infty$ and we

Say that OBS t has high leverage.

$h_t \to 1$ - high leverage

$|u_t \gg 0$ high influence.

Average value of h_t is $\frac{k}{n}$

When each h_t is close to k/n Balanced Design

If some are much smaller and others greater

Unbalanced Design.