ECON 468

CHAPTER 2 : GEOMETRY OF LINEAR REGRESSION

In this chapter, we consider the numerical properties of OLS, that arise as a consequence of how OLS estimates are obtained. Such properties hold for every set of data.

Topics covered:

- geometry of OLS estimation
- Frisch - Waugh - Lovell (FWL) Theorem
- Applications of FWL Theorem
- influential observations

2.2 Review on Vector Geometry

The length, or <u>norm</u>, of a vector $\underset{\sim}{x}$ is $\|x\|$.

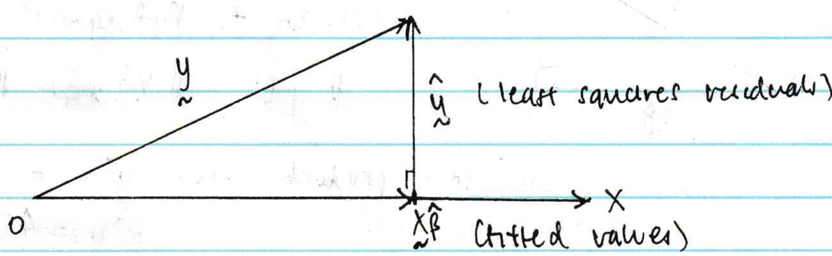
$$\|\underset{\sim}{x}\|^2 = (\underset{\sim}{x}^T \underset{\sim}{x}) = \sum_{i=1}^{n} x_i^2$$

Cosine rule : $\cos \angle(\underset{\sim}{a}, \underset{\sim}{b}) = \dfrac{\underset{\sim}{a} \cdot \underset{\sim}{b}}{\|\underset{\sim}{a}\| \|\underset{\sim}{b}\|}$

Cauchy - Schwartz inequality : $|\underset{\sim}{a}^T \underset{\sim}{b}| \leq \|\underset{\sim}{a}\| \|\underset{\sim}{b}\|$

In order to define the OLS estimator as $\hat{\beta} = (X^T X)^{-1} X^T y$, it is necessary to assume that $(X^T X)$, the $k \times k$ square matrix, is invertible. If $(X^T X)$ is invertible, then the columns of X are linearly independent.

2.3 Geometry of OLS Estimation

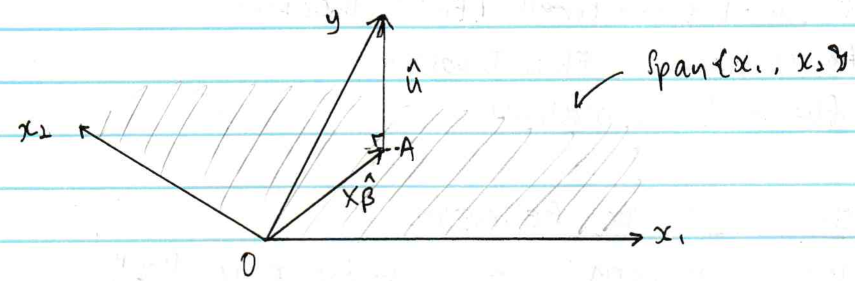


Orthogonality condition : $X^T u = 0 = X^T (y - X\hat{\beta})$

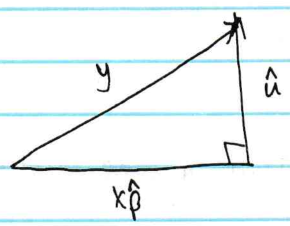
The <u>vector of least squares residuals</u> $\hat{u}$ is $u(\hat{\beta})$, the vector of residuals evaluated at $\hat{\beta}$, the least squares estimator. $\hat{u}$ is orthogonal to all the regressors $\Rightarrow$ $\hat{u}$ is orthogonal to every vector in $\text{Span}\{X\}$, the span of the regressors.

The vector $X\hat{\beta}$ is referred to as the vector of fitted values. As it lies in $\text{Span}\{X\}$, it is orthogonal to $\hat{u}$. In geometric terms, the vector $\hat{u}$ makes a right angle with the vector $X\hat{\beta}$.

<u>3 dimensional setup</u>: y projected on 2 regressors x_1 and x_2



In the 3D set-up, if $\hat{u}$ is orthogonal to the horizontal plane, it must itself be vertical. Thus, it is obtained by "dropping a perpendicular" from y to the horizontal plane. The shortest distance from y to the horizontal plane is obtained by descending vertically onto it, and the point A is the closest point in the plane to y. Thus $\|\hat{u}\|$ minimizes the norm of $u(\beta)$, $\|u(\beta)\|$ with respect to β. $SSR(\beta) = \|u(\beta)\|^2$. Since minimizing the norm of $u(\beta)$ is the same thing as minimizing the squared norm (ie SSR), it follows that $\hat{\beta}$ is ~~also~~ the OLS estimator.



according to Pythagoras' Theorem,

$$\|y\|^2 = \|X\hat{\beta}\|^2 + \|\hat{u}\|^2$$

in scalar product form: $y^T y = \hat{\beta}^T X^T X\hat{\beta} + \hat{u}^T\hat{u}$

$$y^T y = \hat{\beta}^T X^T X\hat{\beta} + (y - X\hat{\beta})^T(y - X\hat{\beta})$$

The total sum of squares (TSS) is equal to the explained sum of squares (ESS) plus the sum of squared residuals (SSR).

<u>Orthogonal projection</u>

Algebraically, an orthogonal projection onto a given subspace can be performed by premultiplying the vector to be projected by a suitable <u>projection matrix</u>.

In OLS, the projection matrix that yields the vector of fitted residuals is $P_X = X(X^TX)^{-1}X^T$, the projection onto X.

Explanation: recall that $\hat{\beta} = (X^TX)^{-1}X^Ty$

Hence $X\hat{\beta}$, vector of fitted values, can be expressed as

$$X\hat{\beta} = X(X^TX)^{-1}X^Ty$$
$$= P_X y, \text{ as } X\hat{\beta} \text{ is the projection of } y \text{ onto } X.$$

The projection matrix that yields the vector of residuals is $M_X = I - P_X$.

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

Explanation: M_X applied to y yields the vector of residuals $\hat{u}$.

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

<u>Properties of P_X and M_X</u>

- $P_X X = X \Rightarrow P_X Xb = Xb$, $b \in \mathbb{R}^k$
- P_X is idempotent : ie $P_X P_X = P_X$

$$\begin{aligned} P_X P_X &= X\cancel{(X^TX)^{-1}X^TX}(X^TX)^{-1}X^T \\ &= X(X^TX)^{-1}X^T = P_X \quad \blacksquare \end{aligned}$$

→ it follows that M_X is idempotent too.

$$\begin{aligned} M_X M_X &= (I - P_X)(I - P_X) \\ &= I - P_X - P_X + P_X^2 \\ &= I - P_X \cancel{- P_X + P_X} \quad \text{since } P_X \text{ idempotent} \\ &= I - P_X = M_X. \quad \blacksquare \end{aligned}$$

- P_X is symmetric ($P_X = P_X^T$)
- P_X and M_X are complementary ie $P_X + M_X = I$

$$P_X y + M_X y = y$$

- $P_X \cdot M_X = 0$

$$\left[\begin{aligned} P_X M_X &= P_X(I - P_X) \\ &= P_X - P_X^2 \\ &= P_X - P_X = 0 \quad \blacksquare \end{aligned} \right]$$

P_X and M_X annihilate each other.

Say we have $P_X z \cdot M_X w = (P_X z)^T(M_X w)$

$$= z^T \underbrace{P_X^T M_X}_{=0} w = 0.$$

The projection matrix M_X annihilates all points that lie in $\text{Span}\{X\}$, likewise P_X annihilates all points that lie in $\text{Span}^{\perp}\{X\}$.

Linear Transformations of Regressors

A nonsingular linear transformation: postmultiply X by any nonsingular $k \times k$ matrix A; let A be partitioned by its columns denoted $\underset{\sim}{a}_i$:

$$XA = X[\underset{\sim}{a}_1 \ \underset{\sim}{a}_2 \ \dots \ \underset{\sim}{a}_k]$$

$$= [X\underset{\sim}{a}_1 \ X\underset{\sim}{a}_2 \ \dots \ X\underset{\sim}{a}_k]$$

$X\underset{\sim}{a}_i$ can be thought of as a linear combination of the columns of X in an n-vector.

$\therefore \text{Span}\{X\} = \text{Span}\{XA\}$

Any element of $\text{Span}\{X\}$ can be described as $X\beta$, $\beta \in \mathbb{R}^k$.

Since A is nonsingular, ie its inverse exists,

$$X\beta = XAA^{-1}\beta = XA\underbrace{(A^{-1}\beta)}_{k \times 1}$$

$\therefore$ the expression is a linear combination of the columns of XA, ie it belongs to $\text{Span}\{XA\}$.

This implies that the orthogonal projections P_X and P_{XA} are the same.

Proof:

$$P_{XA} = XA((XA)^T XA)^{-1} (XA)^T$$

$$= XA(A^T X^T X A)^{-1} A^T X^T$$

$$\cancel{= XA(A^T X^T XA)^{-1}}$$

$$\cancel{= XA A^{-1} X^{-1} X^{T^{-1}} A^{T^{-1}} A^T X^T}$$

$$= X \cancel{A A^{-1}} (X^T X)^{-1} \cancel{A^{T^{-1}} A^T} X^T$$

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

recall rule of reversing inverses: $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$

2.4 FRISCH - WAUGH - LOVELL THEOREM

We have a linear regression model with 2 groups of regressors X_1 and X_2:

$$y = X_1\beta_1 + X_2\beta_2 + u,$$

- X_1 is $n \times k_1$, X_2 is $n \times k_2$ and $X = [X_1 \ X_2]$ is $n \times (k_1 + k_2)$.
- assume that X_1 and X_2 are orthogonal ie $X_1^T X_2 = 0 = X_2^T X_1$

Under this assumption, the vector of least squares estimates $\hat{\beta}_1$ is the same as the one obtained from $y = X_1\beta_1 + u$; same for $\hat{\beta}_2$.

In other words, if X_1 and X_2 are orthogonal, we can drop either set of regressors without affecting the coefficients of the other set. (Regressing y on only X_1, or on only X_2, will give us the same estimates as regressing it on both X_1 and X_2).

From $y = X_1\beta_1 + X_2\beta_2 + u$, vector of fitted values is $P_X y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2$
From $y = X_1\beta_1 + u$, vector of fitted values is $P_{X_1} y$.

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

$P_X P_{X_1} = P_{X_1}$ is true whether or not X_1 and X_2 are orthogonal.

Proof: $P_X P_{X_1} = P_X X_1 (X_1^T X_1)^{-1} X_1^T$
$= X_1 (X_1^T X_1)^{-1} X_1^T$ since $P_X X_1 = X_1$
$= P_{X_1}$. ∎

$P_{X_1} P_X = P_{X_1}$ is also true, we can obtain this by evaluating $(P_X P_{X_1})^T$.

Hence, the vector of fitted values $P_{X_1} y$ can be evaluated:

$$
\begin{aligned}
P_{X_1} y &= P_{X_1} P_X y \\
&= P_{X_1}(X_1\hat{\beta}_1 + X_2\hat{\beta}_2) \\
&= P_{X_1} X_1 \hat{\beta}_1 + P_{X_1} X_2 \hat{\beta}_2 \\
&= X_1 \hat{\beta}_1 \quad (\text{since } P_{X_1} X_1 = X_1,\ P_{X_1} X_2 = 0 \text{ because of orthogonality})
\end{aligned}
$$

the same OLS estimator in regression model with both X_1 and X_2.

<u>What if X_1 and X_2 are not orthogonal?</u>

We can create a set of variables from X_2 that are orthogonal to X_1 by acting on X_2 with the orthogonal projection $M_{X_1} = I - P_{X_1}$ to obtain $M_{X_1} X_2$.

$$
\begin{aligned}
y &= X_1\alpha_1 + M_1 X_2 \alpha_2 + u \\
&= X_1\alpha_1 + (I - X_1(X_1^T X_1)^{-1} X_1^T) X_2 \alpha_2 + u \\
&= X_1\alpha_1 + (X_2 - X_1(X_1^T X_1)^{-1} X_1^T X_2)\alpha_2 + u
\end{aligned}
$$

This is a regression model with 2 groups of regressors X_1 and $M_1 X_2$, which are mutually orthogonal. Therefore, if we omit X_1, $\hat{\alpha}_2$ will be unchanged: the regressions

$y = X_1\alpha_1 + M_1 X_2 \alpha_2 + u$ and $y = M_1 X_2 \alpha_2 + v$ must yield the same $\hat{\alpha}_2$. However, note that the residuals $u \neq v$.

If we replace y by $M_1 y$, we further obtain:

$$M_1 y = M_1 (M_1 X_2 \alpha_2) + \text{residuals}$$

$$M_1 y = M_1 X_2 \alpha_2 + \text{residuals} \quad (M \text{ is idempotent})$$

where $\hat{\alpha}_2$ should be the OLS estimate $\hat{\beta}_2$.

Formally derived the theorem:

FWL Theorem: 1. The OLS estimates β_2 from $y = X_1 \beta_1 + X_2 \beta_2 + u$ and $M_1 y = M_1 X_2 \beta_2 + \text{residuals}$ are numerically identical.

2. The residuals from the above 2 regressions are numerically identical.

Proof: By the standard formula $\hat{\beta} = (X^T X)^{-1} X^T y$,

$\hat{\beta}_2$ from $M_1 y = M_1 X_2 \beta_2 + u$ is

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

From $y = X_1 \beta_1 + X_2 \beta_2 + u$, let $\hat{\beta}_{OLS1}$ and $\hat{\beta}_{OLS2}$ denote the vectors of OLS estimates.

Then,
$$y = P_X y + M_X y$$
$$= X_1 \hat{\beta}_{OLS1} + X_2 \hat{\beta}_{OLS2} + M_X y \quad \text{——} \ (*)$$

Multiply both sides by $X_2^T M_1$ on the LHS to get:

$$X_2^T M_1 y = X_2^T \underbrace{M_1 X_1}_{=0} \hat{\beta}_{OLS1} + X_2^T M_1 X_2 \hat{\beta}_{OLS2} + \underbrace{X_2^T M_1 M_X y}_{=0}$$

$$\therefore X_2^T M_1 y = X_2^T M_1 X_2 \hat{\beta}_{OLS2}$$

Pre multiply by $(X_2^T M_1 X_2)^{-1}$ to obtain:

$$(X_2^T M_1 X_2)^{-1} X_2^T M_1 y = \hat{\beta}_{OLS2}$$

$$\hat{\beta}_2 = \hat{\beta}_{OLS2}. \quad (\text{Proof of \#1} \ \blacksquare)$$

Pre multiply $(*)$ by M_1:

$$M_1 y = \underbrace{M_1 X_1}_{=0} \hat{\beta}_1 + M_1 X_2 \hat{\beta}_2 + \underbrace{M_1 M_X}_{=M_X} y$$

$$= M_1 X_2 \hat{\beta}_2 + M_X y$$

$$\therefore u = M_X y \quad (\text{Proof of \#2} \ \blacksquare)$$

In general, for $y = X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2 + \hat{u}$, FWL Theorem says that $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{u}$ are the same for: $M_1 y = M_1 X_2 \hat{\beta}_2 + \hat{u}$

$$M_2 y = M_2 X_1 \hat{\beta}_1 + \hat{u}$$

2.5 APPLICATIONS OF FWL THEOREM

A regression in which the regressors are broken up into 2 groups can arise in many situations. We will look at 3 of these: seasonal dummy variables, time trends and measures of goodness of fit.

Seasonal dummy variables

Many economic activities are strongly affected by the season; use seasonal dummy variables to help capture the seasonal variation in our data. Consider quarterly data, so there are 4 seasonal dummy variables that each take "1" for just one of the four seasons.

$$\underset{\sim}{s}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix}, \quad \underset{\sim}{s}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ \vdots \end{bmatrix}, \quad \underset{\sim}{s}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ \vdots \end{bmatrix}, \quad \underset{\sim}{s}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ \vdots \end{bmatrix}$$

$s_i \perp s_j$, $i \neq j$ $\rightarrow$ $s_i^T s_j = 0$ (orthogonal)

$\underset{\sim}{s}_1 + \underset{\sim}{s}_2 + \underset{\sim}{s}_3 + \underset{\sim}{s}_4 = \underset{\sim}{i}$

Our regression model: $y = \alpha_1 \underset{\sim}{s}_1 + \alpha_2 \underset{\sim}{s}_2 + \alpha_3 \underset{\sim}{s}_3 + \alpha_4 \underset{\sim}{s}_4 + X\beta + u$

If observation t is in the 1st quarter, the t^{th} observation can be written as:

$$y_t = \alpha_1 + X_t\beta + u_t.$$

The introduction of seasonal dummies gives us a different constant for every season.

Let S denote an $n \times 4$ matrix, $S = [\underset{\sim}{s}_1 \ \underset{\sim}{s}_2 \ \underset{\sim}{s}_3 \ \underset{\sim}{s}_4]$

Then the regression can be written as $y = S\delta + X\beta + u$, when it is clear that there are 2 groups of regressors, as required for FWL!

From $y = S\delta + X\beta + u$, we apply FWL to find:

$M_S y = M_S X\hat{\beta} + M_S \hat{u}$, where $M_S = I - S(S^TS)^{-1}S^T$

* seasonally adjusting y, $M_S y$ is orthogonal to all seasonal variables.

$\beta = \hat{\beta}$ and $u = M_S \hat{u}$ according to FWL.

Time Trends

The linear time trend, represented by $T = [1 \vdots 2 \vdots 3 \vdots 4 \vdots \cdots]$.

Imagine we have a regression with a constant and linear time trend

$y = \gamma_1 \underset{\sim}{i} + \gamma_2 T + X\beta + u$,

observation t: $y_t = \gamma_1 + \gamma_2 T_t + X_t \beta + u_t$

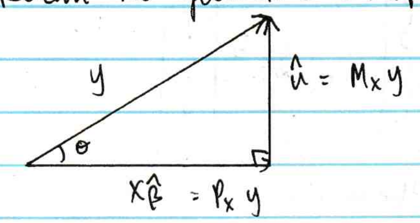
It is often desirable to make the time trend orthogonal to the constant centering it, by applying $M_{\underset{\sim}{i}}$.

$$M_i y = \underbrace{y - \bar{y} i}_{\text{deviations from the mean}}, \quad \text{where } \bar{y} = \frac{1}{n} \sum_{t=1}^{n} y_t$$

We can also project all other variables in a regression model off the time trend variables, to obtain detrended variables eg $M_T y$.

<u>Goodness of fit of a regression</u>

Recall the geometric interpretation of OLS:



$$\|y\|^2 = \|P_X y\|^2 + \|M_X y\|^2$$
$$= ESS + SSR = TSS$$

The measure of goodness of fit of a regression model is known as the coefficient of determination, denoted R^2.

$$R^2 = \frac{ESS}{TSS} = \cos^2 \theta$$

For any angle θ, $-1 \leq \cos\theta \leq 1 \Rightarrow 0 \leq R^2 \leq 1$

If $\theta = 0$, y and $X\hat{\beta}$ coincide: perfect fit, $R^2 = 1$.

If θ = right angle, y coincides with $\hat{u}$: $R^2 = 0$ (no fit).

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

- uncentered $R^2 = R_u^2 = \dfrac{\|P_X y + \alpha i\|^2}{\|y + \alpha i\|^2}$

Choosing a large α brings R_u^2 closer to 1, however it might be misleading if α dominates $P_X y$ and y.

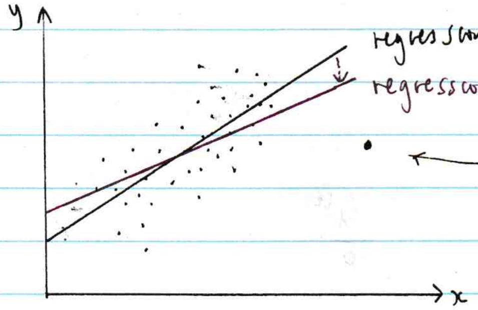
- centered R^2: $R_c^2 = \dfrac{\|P_X M_i y\|^2}{\|M_i y\|^2} = 1 - \dfrac{\|M_X y\|^2}{\|M_i y\|^2}$

2.6 INFLUENTIAL OBSERVATIONS

An important feature of OLS estimation is that each element of the vector of parameter estimates, $\hat{\beta}$, is a weighted average of the elements of the vector y.

$$\hat{\beta}_i = \underbrace{i^{th} \text{ row of } (X^T X)^{-1} X^T}_{\text{"weight"}} \cdot y$$

Because each element of $\hat{\beta}$ is a weighted average, some observations may affect the value of $\hat{\beta}$ much more than others do.



x coordinate

Because of its position well to the right of the other observations, this point exerts a good deal of leverage on the regression line, pulling it down towards itself.

Influence: how much the predicted scores for other observations would differ if the observation in question were not included.

Leverage: how much the observation's value on the predictor variable differs from the mean of the predictor variable. The greater an observation's leverage, the more potential it has to be an influential observation.

Hence, for the example above, it is the x-coordinate that gives the point its position of high leverage; but the y-coordinate determines whether the high leverage position will be exploited, resulting in substantial influence on the regression line.

Influence of a single observation t on $\hat{\beta}$: $\hat{\beta} - \hat{\beta}^{(t)}$ ← estimate of β when the t^{th} observation is omitted

We remove the effect of the t^{th} observation by using a dummy variable,

$$e_t = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \leftarrow t^{th} \text{ observation}$$

and including it as a regressor.

$$y = X\beta + \alpha e_t + u$$

We have 2 regressors X and e_t. Apply FWL theorem to obtain

$$M_t y = M_t X\beta + \text{residuals}, \quad M_t \equiv M_{e_t} = I - e_t(e_t^T e_t)^{-1} e_t^T$$

* y with its t^{th} component replaced by 0.

$$M_t y = (I - P_t)\, y$$
$$= y - e_t \underbrace{(e_t^T e_t)^{-1}}_{=1} e_t^T y$$
$$= y - e_t \underbrace{e_t^T y}_{t^{th}\text{ component of } y}$$
$$= y - e_t y_t$$

$$y = X\hat{\beta}^{(t)} + \hat{\alpha} e_t + \hat{u}^{(t)}$$

Premultiply by P_x:

$$P_x y = P_x X \hat{\beta}^{(t)} + P_x \hat{\alpha} e_t + \underbrace{P_x \hat{u}^{(t)}}_{=0}$$
$$= X\hat{\beta}^{(t)} + \hat{\alpha} P_x e_t$$

Using $P_x y = X\hat{\beta}$, we get the following equality:

$$X\hat{\beta} = X\hat{\beta}^{(t)} + \hat{\alpha} P_x e_t$$
$$X(\hat{\beta} - \hat{\beta}^{(t)}) = \hat{\alpha} P_x e_t$$

← the measure of influence by observation t.

To compute $\hat{\alpha}$, use FWL theorem, which tells us that $\hat{\alpha}$ in $y = X\hat{\beta} + \hat{\alpha} e_t + \hat{u}$ is the same as in $M_x y = M_x \hat{\alpha} e_t + \text{residuals}$.

$$\hat{\alpha} = \frac{e_t^T M_x y}{e_t^T M_x e_t}$$

$$= \frac{t^{th}\text{ element in vector of residuals}}{t^{th}\text{ diagonal element in } M_x}$$

$$= \frac{\hat{u}_t}{1 - h_t}, \quad h_t = t^{th}\text{ diagonal element in } P_x.$$

Recall that $X(\hat{\beta} - \hat{\beta}^{(t)}) = \hat{\alpha} P_x e_t$.

Premultiply both sides by $(X^T X)^{-1} X^T$ to obtain:

$$\hat{\beta} - \hat{\beta}^{(t)} = (X^T X)^{-1} X^T \hat{\alpha} P_x e_t$$
$$= \hat{\alpha} (X^T X)^{-1} X^T X (X^T X)^{-1} X^T e_t$$
$$= \hat{\alpha} (X^T X)^{-1} X_t^T = \frac{\hat{u}_t}{1 - h_t} (X^T X)^{-1} X_t^T$$

When either $\hat{u}_t$ is large or h_t is large, or both, the effect ~~abv~~ the t^{th} observation on at least some elements of $\hat{\beta}$ is likely to be substantial.
The influence of an observation depends on both $\hat{u}_t$ and h_t.
If $\hat{u}_t$ is large (this is related to the y-coordinate), influence is greater.
If h_t is large (x-coord), it has high leverage, or potential influence.

<u>Properties of h_t</u>

- $h_t = e_t^T P_x e_t = \| P_x e_t \|^2$

  (because $e_t^T P_x^T P_x e_t = e_t^T P_x e_t$)

- $0 \le h_t \le 1$

  ($\| P_x e_t \|^2 \ge 0$, and Pythagoras' theorem tells us $\| e_t \|^2 = \| P_x e_t \|^2 + \| M_x e_t \|^2$
  $\Rightarrow 1 = \| P_x e_t \|^2 + \| M_x e_t \|^2$
  As $\| M_x e_t \|^2 \ge 0$, $\| P_x e_t \|^2 \le 1$.)

If A is a square $n \times n$ matrix, its trace, denoted $tr(A)$ is the sum of the elements on the principal diagonal:

$$tr(A) = \sum_{i=1}^{n} A_{ii}$$

The trace is invariant under a cyclic permutation of the factors, ie
$tr(ABC) = tr(BCA) = tr(CAB)$

$$\begin{aligned} \sum_{t=1}^{n} h_t = tr(P_x) &= tr(X(X^TX)^{-1}X^T) \\ &= tr(X^TX(X^TX)^{-1}) \\ &= tr(I_k) \\ &= k \end{aligned}$$

} cyclic permutation

The average of h_t is $\frac{1}{n}\sum_{t=1}^{n} h_t = \frac{k}{n}$. When h_ts are close to the average value $\frac{k}{n}$, it implies that no observation has very much leverage; X is said to have a balanced design