Multiple Regression Model: I

Suppose the data are generated according to

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + u_i \quad i = 1 \ldots n$$

Define

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} x_{11} & \cdots & x_{1K} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} \quad u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$$

So $y \in \mathbb{R}^n,\ X \in \mathbb{R}^{nxK},\ \beta \in \mathbb{R}^K,\ u \in \mathbb{R}^n$

Rks:

- In many applications, the first column of X is a vector of ones, so $(\forall i)\ x_{i1} = 1. \quad \therefore\ \beta_1$ is an intercept.

- Wooldridge likes to label the intercept β_0 and he always includes it in X. In his notation, k indicates the number of explanatory variables in addition to the intercept, so his X matrix has $k+1$ columns. Be aware of this difference in notation!

:Multiple Regression Model in Matrix notation

$$y = X\beta + u$$

Rk: I'll often write $y_i = x_i\beta + u_i$ as the typical observation

: Definition of the OLS estimator $\widehat{\beta}$

$$(\widehat{\beta}_1, \cdots \widehat{\beta}_K)' = \arg \min_{(\widetilde{\beta}_1, \cdots \widetilde{\beta}_K)'} \sum_{i=1}^{n} (y_i - \widetilde{\beta}_1 x_{i1} - \cdots - \widetilde{\beta}_K x_{iK})^2$$

or

$$\widehat{\beta} = \arg \min_{\widetilde{\beta} \in \mathbb{R}^K} (y - X\widetilde{\beta})'(y - X\widetilde{\beta})$$

:Normal equations

$$\sum_{i=1}^{n} x_{ij}(y_i - \widehat{\beta}_1 x_{i1} - \cdots - \widehat{\beta}_K x_{iK}) = 0 \quad j = 1, \cdots, K$$

or in matrix notation

$$X'(y - X\widehat{\beta}) = 0 \Leftrightarrow X'\widehat{u} = 0$$

:Expression for $\widehat{\beta}$

$$\begin{aligned} X'(y - X\widehat{\beta}) &= 0 \\ &\Leftrightarrow (X'X)\widehat{\beta} = X'y \end{aligned}$$

Need to show

1. a solution always exists

2. solution is unique if $\det(X'X) \neq 0 \Leftrightarrow (X'X)$ is invertible $\Leftrightarrow$ $(X'X)$ is nonsingular $\Leftrightarrow rank(X'X) = K \Leftrightarrow rank(X) = K$.

Under any of the conditions 2. above, we get

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

and

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

$$\widehat{u} = y - \widehat{y} = (I - P)y \equiv My$$

:Geometric interpretation of OLS

Think of y as a vector in $\mathbb{R}^n$ and $\widetilde{y} = X\widetilde{\beta}$ as a vector in $Sp(X) \subset \mathbb{R}^n$. The OLS problem can be written as

Find:

$$\widehat{y} = \arg \min_{\widetilde{y} \in Sp(X)} (y - \widetilde{y})'(y - \widetilde{y})$$

Solution: Let $\widehat{y}$ denote the orthogonal projection of y onto $Sp(X)$, that is the vector that generates a residual $\widehat{u} \perp Sp(X)$

Rks:

- $\widehat{u}$ is the vector $y - \widehat{y}$

- $\hat{u} \perp Sp(X)$ means that $\hat{u}'\tilde{y} = 0$ for all $\tilde{y} \in Sp(X)$
- $\hat{u}'\tilde{y} = 0 \Leftrightarrow \hat{u}'X\tilde{\beta} = 0$ for all $\tilde{\beta} \in \mathbb{R}^K \Leftrightarrow \hat{u}'X = 0$

1. The existence of $\hat{y}$ is geometrically obvious (for a proof use $Sp(X)$ is closed or $Range(X') = Range(X'X)$)

2. Given that exists an orthogonal projection, it's easy to show that it solves OLS problem and is unique. Consider any $\tilde{y} \in Sp(X)$. We have

$$y = \tilde{y} + \tilde{u} = \hat{y} + \hat{u}$$

$$\Leftrightarrow \tilde{u} = \hat{u} + (\hat{y} - \tilde{y})$$

Therefore

$$\begin{aligned}\tilde{u}'\tilde{u} &= \hat{u}'\hat{u} + (\hat{y} - \tilde{y})'(\hat{y} - \tilde{y}) + 2\hat{u}'(\hat{y} - \tilde{y}) \\ &= \hat{u}'\hat{u} + (\hat{y} - \tilde{y})'(\hat{y} - \tilde{y}) \quad \because \ (\hat{y} - \tilde{y}) \in Sp(X) \\ &\geq \hat{u}'\hat{u} \quad \text{with equality iff } (\hat{y} - \tilde{y}) = 0\end{aligned}$$

Rks:

- The orthogonal projection operator **P** is linear, i.e.
  $$\mathbf{P}(c_1 y_1 + c_2 y_2) = c_1 \mathbf{P}(y_1) + c_2 \mathbf{P}(y_2)$$
  so given a basis, it can be represented by a matrix P, i.e. $\widehat{y} = Py$.
- Because projections satisfy $\mathbf{P}(\mathbf{P}(y_1)) = \mathbf{P}(y_1)$, we must have $P\widehat{y} = \widehat{y} \Leftrightarrow P^2 = P$ so P must be idempotent
- For orthogonal projections, we get the additional property that $P = P'$
- $M = (I - P)$ corresponds to the projection onto $Sp(X)^{\perp}$, i.e. the set of vectors orthogonal to $Sp(X)$.
- $P^2 = P$ and $P = P'$ implies that all the eigenvalues of P are either 0 or 1.

:Back to $\widehat{\beta}$

From the properties of $\hat{y}$ above we get

1. existence of $\hat{y} \in Sp(X) \Rightarrow \hat{y} = X\widehat{\beta}$ for some $\widehat{\beta} \in \mathbb{R}^K$

2. If columns of X are linearly independent, then

- $\widehat{\beta}$ is unique
- $P = X(X'X)^{-1}X'$

:Analysis of Variance

Define $M = I - P$. We have the decomposition

$$y = Py + My$$
$$= \hat{y} + \hat{u}$$

$$y'y = \hat{y}'\hat{y} + \hat{u}'\hat{u} \qquad \text{since } \hat{y}'\hat{u} = 0$$

$$\text{or } \; SST_0 = SSE_0 + SSR_0$$

where the subscript “0” is used to indicate “about the origin”

Define $R_0^2 = SSE_0/SST_0 = 1 - SSR_0/SST_0$.

Properties

- $0 \leq R_0^2 \leq 1$
- $\min \tilde{u}'\tilde{u} \Leftrightarrow \max R_0^2$
- Let θ_0 denote the angle between y and $\hat{y}$.

$$\cos^2(\theta_0) = \left[\frac{y'\hat{y}}{\sqrt{y'y \cdot \hat{y}'\hat{y}}} \right]^2 = \frac{\hat{y}'\hat{y}}{y'y} = R_0^2$$

where we have used $y'\hat{y} = (\hat{y} + \hat{u})'\hat{y} = \hat{y}'\hat{y}$

We don't usually use R_0^2 if X contains an intercept.

:Coefficient of Determination R^2

Define

$$A = I_n - \frac{1}{n}\iota\iota' \quad \text{where } \iota = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^n$$

Rks:

- $\overline{y} = \frac{1}{n}\iota' y$
- $Ay = y - \overline{y}\iota$
- $A = A^2 = A' \qquad A\iota = 0$ and $Az = z$ if $z'\iota = 0$
- If $\iota \in Sp(X)$, then $\iota'\hat{u} = 0$ and

$$A\hat{u} = \hat{u} - \frac{1}{n}\iota\iota'\hat{u} = \hat{u}$$

- $\iota'\hat{u} = 0 \Leftrightarrow \overline{\hat{u}} = 0 \Leftrightarrow \overline{\hat{y}} = \overline{y}$

Assuming $\iota \in Sp(X)$, we can write

$$Ay = A\hat{y} + A\hat{u} = A\hat{y} + \hat{u}$$

Therefore,

$$y'A'Ay = \hat{y}'A'A\hat{y} + \hat{u}'\hat{u} + 2\hat{y}'A'\hat{u} \quad \Leftrightarrow$$

$$y'Ay = \hat{y}'A\hat{y} + \hat{u}'\hat{u} \quad \Leftrightarrow$$

$$SST = SSE + SSR$$

where the absence of a subscript denotes "about the mean".

Define $R^2 = SSE/SST$

Properties (assuming $\iota \in Sp(X)$)

- $0 \leq R^2 \leq 1$
- $\min \tilde{u}'\tilde{u} \Leftrightarrow \max R^2$
- Let θ denote the angle between Ay and $A\hat{y}$.

$$\cos^2(\theta) = \left[\frac{y' A \hat{y}}{\sqrt{y' A y \bullet \hat{y}' A \hat{y}}} \right]^2 = \frac{\hat{y}' A \hat{y}}{y' A y} = r_{y\hat{y}}^2 = R^2$$

Rk: For $A = I_n - \frac{1}{n}\iota\iota'$, we get

$$(Ax)' Ay = (Ax)' y = x'(Ay) \qquad \Leftrightarrow$$

$$\sum (x_i - \overline{x})(y_i - \overline{y}) = \sum (x_i - \overline{x}) y_i = \sum x_i (y_i - \overline{y})$$

:Changing units

What happens if I change the units of measurement in

(a) the dependent variable?

(b) the independent variable?

a) Let $y_v = cy + a$

where $c \in \mathbb{R}$ and $a \in Sp(X) \Leftrightarrow a = X\gamma$ for some γ

$$\hat{y}_v = P\hat{y}_v = P(cy + a)$$

$$= cPy + Pa \quad \because \text{ projection is linear}$$

$$= c\hat{y} + a \quad \because \; a \in Sp(X)$$

Therefore,

$$\hat{y}_v \equiv X\hat{\beta}_v = cX\hat{\beta} + X\gamma = X(c\hat{\beta} + \gamma)$$

$$\Leftrightarrow \hat{\beta}_v = c\hat{\beta} + \gamma$$

Using $\hat{u}_v = y_v - \hat{y}_v = c(y - \hat{y}) = c\hat{u}$, we get

$$\hat{u}_v'\hat{u}_v = c^2 u'u$$

The relationship between R_v^2 and R^2 can be complicated, but if $a \in Sp(i)$, then $Aa = 0$ so

$$R_v^2 = \frac{\hat{y}_v'A\hat{y}_v}{y_v'Ay_v} = \frac{c^2\hat{y}'A\hat{y}}{c^2y'Ay} = R^2$$

b) Let $X_v = XD$ where D is invertible. This allows us to consider an arbitrary change of basis for $Sp(X)$ as well as a change in units (special case $D = diag(c_1, \cdots, c_K)$).

Because D is invertible, any vector $z = X\lambda$ can also be written as $z = X_v\lambda_v$ and vice-versa (use $z = (XD)(D^{-1}\lambda) \equiv X_v\lambda_v$). Therefore, $Sp(X_v) = Sp(X)$. It follows immediately that

$$\begin{aligned} P_v y &= Py \quad \Leftrightarrow \\ X_v\widehat{\beta}_v &= X\widehat{\beta} \quad \Leftrightarrow \\ X\left(D\widehat{\beta}_v - \widehat{\beta}\right) &= 0 \\ &\Leftrightarrow \widehat{\beta}_v = D^{-1}\widehat{\beta} \end{aligned}$$

Exercise: Show that the residual sum of squares and the R^2 are unchanged if replace the regressors X with X_v.

Before completing our discussion of the algebra of OLS, I need to introduce another quick piece of matrix algebra.

Definition: Let $A \in \mathbb{R}^{mxm}$. The *trace* of A is the sum of its diagonal components, i.e.

$$tr(A) \equiv \sum_{i=1}^{m} a_{ii}$$

Properties:

1. $(\forall c \in \mathbb{R})\ \ tr(cA) = c \bullet tr(A);\ tr(A+B) = tr(A) + tr(B)$

2. $tr(A) = tr(A')$

3. if both products exist, $tr(AB) = tr(BA)$

4. if A is idempotent, $tr(A) = rk(A)$

: Consequences of adding an observation

Express X in terms of its rows, i.e.

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{where } x_i \in \mathbb{R}^{1xK} \text{ is the } i^{th} \text{ row of } X$$

Recall $\hat{y} = Py$ where $P = X(X'X)^{-1}X'$. Define h_{ii} as the i^{th} diagonal element of P. We have

$$h_{ii} = [X(X'X)^{-1}X']_{ii} = x_i(X'X)^{-1}x_i'$$

By definition

$$\sum_i h_{ii} = tr(P) = tr(X(X'X)^{-1}X')$$

$$= tr((X'X)^{-1}X'X) = tr(I_K) = K$$

Therefore, $\overline{h} = K/n$.

Definition: If h_{ii} is 'high' (say $h_{ii} > 2\overline{h}$), then observation i is called a leverage point.

Leverage points are observations whose explanatory variables have the potential to exert an unusually strong effect on the fitted model. To see why, it is useful to understand how the regression coefficients change if we add a new observation.

Exercise: If $A \in \mathbb{R}^{nxn}$ is invertible, and $c \in \mathbb{R}^{n}$, then

$$(A + cc')^{-1} = A^{-1} - \frac{A^{-1}cc'A^{-1}}{1 + c'A^{-1}c}$$

It's easy to see that

$$X'X = \sum_i x_i' x_i \quad \text{and } X'y = \sum_i x_i' y_i$$

Therefore

$$\widehat{\beta} = \left(\sum_i x_i' x_i\right)^{-1} \sum_i x_i' y_i$$

The following allows us to see how $(X'X)^{-1}$ changes when we add observation j to the rest of the sample:

$$\left(\sum_i x_i' x_i\right)^{-1} = \left(\sum_{i \neq j} x_i' x_i\right)^{-1} - \frac{\left(\sum_{i \neq j} x_i' x_i\right)^{-1} x_j' x_j \left(\sum_{i \neq j} x_i' x_i\right)^{-1}}{1 + x_j \left(\sum_{i \neq j} x_i' x_i\right)^{-1} x_j'}$$

Rk: Use this result to show that $0 \leq h_{ii} \leq 1$.

Let $\widehat{\beta}(j)$ denote the OLS estimator if we drop obs j i.e.

$$\widehat{\beta}(j) = \left(\sum_{i \neq j} x_i' x_i\right)^{-1} \sum_{i \neq j} x_i' y_i$$

Using the result above, we obtain

$$\widehat{\beta} = \widehat{\beta}(j) + (X'X)^{-1} x_j' \frac{y_j - x_j \widehat{\beta}(j)}{1 - h_{jj}}$$

So has h_{jj} gets bigger, the effect of the j^{th} observation becomes potentially bigger. An *influential* observation is one that has a big effect (not just a potentially big effect). We see that this depends on h_{jj} and also $y_j - x_j \widehat{\beta}(j)$ (the error we would get forecasting the j^{th} observation).

: Consequences of adding many observations

Write

$$y = \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \end{bmatrix} \text{ and } X = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix}$$

where $\underline{y}_s \in \mathbb{R}^{n_s}$ and $\underline{X}_s \in \mathbb{R}^{n_s x K}$ for $s = 1,2$ where both n_s are large enough that both $\underline{X}_s$ has full column rank. Then we can define the OLS estimators from the two subsamples as

$$\widehat{\beta}_1 = (\underline{X}_1' \underline{X}_1)^{-1} \underline{X}_1' \underline{y}_1 \text{ and } \widehat{\beta}_2 = (\underline{X}_2' \underline{X}_2)^{-1} \underline{X}_2' \underline{y}_2$$

When we combine the two samples, we get

$$\begin{aligned}\widehat{\beta} &= (X'X)^{-1}X'y \\ &= (\underline{X}_1'\underline{X}_1 + \underline{X}_2'\underline{X}_2)^{-1}(\underline{X}_1'\underline{y}_1 + \underline{X}_2'\underline{y}_2) \\ &= (\underline{X}_1'\underline{X}_1 + \underline{X}_2'\underline{X}_2)^{-1}((\underline{X}_1'\underline{X}_1)\widehat{\beta}_1 + (\underline{X}_2'\underline{X}_2)\widehat{\beta}_2)\end{aligned}$$

Recall that $V(\widehat{\beta}_s) = \sigma^2(\underline{X}_s'\underline{X}_s)^{-1} \equiv H_s^{-1}$ where H_s is called the *precision matrix*. So we can write

$$\widehat{\beta} = (H_1 + H_2)^{-1}(H_1\widehat{\beta}_1 + H_2\widehat{\beta}_2)$$

so $\widehat{\beta}$ is a *precision weighted average*.

: Frisch-Waugh Theorem

Suppose we write

$$
\begin{aligned}
y &= X\widehat{\beta} + \widehat{u} \\
&= X_1\widehat{\beta}_1 + X_2\widehat{\beta}_2 + \widehat{u} \quad (*)
\end{aligned}
$$

where X_1 denotes the first K_1 columns of X, and X_2 denotes the remaining K_2 columns (with $K_1 + K_2 = K$).

Define $M_2 = I_n - X_2(X_2'X_2)^{-1}X_2,$ so for any vector z, M_2z gives the part of z that is normal to $Sp(X_2)$.

Premultiplying both sides of (*) by $X_1'M_2$ gives

$$
\begin{aligned}
X_1'M_2y &= X_1'M_2X_1\widehat{\beta}_1 + X_1'M_2X_2\widehat{\beta}_2 + X_1'M_2\widehat{u} \\
&= X_1'M_2X_1\widehat{\beta}_1 + X_1'0 + X_1'\widehat{u} \\
&= X_1'M_2X_1\widehat{\beta}_1
\end{aligned}
$$

But if X has rank K, then $X_1' M_2 X_1$ must have rank K_1.

$$\therefore \quad \widehat{\beta}_1 = (X_1' M_2 X_1)^{-1} X_1' M_2 y$$

This formula has a relatively simple interpretation.

Let $\widetilde{X}_1 = M_2 X_1$ and $\widetilde{y} = M_2 y$. $\widetilde{X}_1$ replaces each column of X_1 with the part of it that is normal to $Sp(X_2)$. So to compute $\widehat{\beta}_1$, we regress y on part of $X_1 \perp Sp(X_2)$. Equivalently, regress part of $y \perp Sp(X_2)$ on part of $X_1 \perp Sp(X_2)$.

Of course, the formula is symmetric and we can also write

$$\therefore \quad \widehat{\beta}_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 y$$

Exercise: Show that if $X_1' M_2 X_1$ is singular, then $\exists\ c_1 \neq 0$ we have $X_1 c_1 = X_2 c_2$

The Frisch-Waugh theorem has two important uses:

1. It provides a computational trick. This was important historically and still useful when dealing with panel data sets and individual effects.

2. It provides an understanding of how OLS controls for X_2 when it estimates the partial response of y to X_1. ONLY the part of $X_1 \perp Sp(X_2)$ is used to estimate $\widehat{\beta}_1$.

Special cases:

- Suppose X_1 is a column of ones (i.e. an intercept). Then $M_1 = I - (1/n)\iota\iota'$ (which we called A above). $M_1 y$ and $M_1 X_2$ are just deviations from means. So if $K_2 = 1$

$$\widehat{\beta}_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 y$$

$$= \frac{\sum (x_{i2} - \overline{x}_2)(y_i - \overline{y})}{\sum (x_{i2} - \overline{x}_2)^2}$$

- More generally, suppose K is arbitrary but still we are interested in a single coefficient. w.l.o.g., order the columns of X so that it corresponds to β_1. Define $\widehat{r} = M_2 X_1$.

$$\widehat{\beta}_1 = \frac{\sum \widehat{r}_i y_i}{\sum \widehat{r}_i^2}$$

Sampling Properties of $\widehat{\beta}$ and $\widehat{\sigma}^2$: Some moments

Suppose the data generation process satisfies the following assumptions:

- MLR.1 $y_i = \underline{x}_i\beta + u_i \quad i = 1 \ldots n$
- MLR.2 $\{(\underline{x}_i, y_i), i = 1..n\}$ is a random sample
- MLR.3 In the sample, there are no exact linear combinations among the independent variables
- MLR.4 $E(u_i|\underline{x}_i) = 0 \quad i = 1 \ldots n$
- MLR.5 $V(u_i|\underline{x}_i) = \sigma^2 \quad i = 1 \ldots n$

Rks: I've written $\underline{x}_i$ for the vector of explanatory variables corresponding to observation i.

Rewrite these assumptions in matrix notation:

- S1 $y = X\beta + u$
- S2 $X'X$ is invertible
- S3 $E(u|X) = 0$
- S4 $V(u|X) = \sigma^2 I_n$

where we have

- SLR.1⇔S1
- SLR.3⇔S2
- SLR.2 and SLR.4⇒S3
- SLR.2 and SLR.5⇒S4

Define $L = (X'X)^{-1}X'$. Under S1 and S2,

$$\begin{aligned} \widehat{\beta} &= (X'X)^{-1}X'y \\ &= \beta + Lu \end{aligned}$$

:First Moment of $\widehat{\beta}$

$$E(\widehat{\beta}|X) = E(\beta + Lu) = \beta + LE(u|X)$$

$$= \beta \quad \text{by S3}$$

RK: $E(\widehat{\beta}) = E(E(\widehat{\beta}|X)) = \beta$

:Second Moment of $\widehat{\beta}$

$$V(\widehat{\beta}|X) = V(\beta + Lu|X) = LV(u|X)L'$$

$$= \sigma^2(X'X)^{-1} \quad \text{by S4}$$

From the Frisch-Waugh theorem, if we are interested in the variance of a subset of the coefficients, w.l.o.g call it $\widehat{\beta}_1$, we have

$$
\begin{aligned}
V(\widehat{\beta}_1|X) &= V(X_1'M_2X_1)^{-1}X_1'M_2y|X) \\
&= V((X_1'M_2X_1)^{-1}X_1'M_2u|X) \\
&= (X_1'M_2X_1)^{-1}X_1'M_2V(u|X)M_2X_1(X_1'M_2X_1)^{-1} \\
&= \sigma^2(X_1'M_2X_1)^{-1}
\end{aligned}
$$

If β_1 is a scalar case, we can write

$$\sigma^2(X_1'M_2X_1)^{-1} = \frac{\sigma^2(X_1'X_1)^{-1}}{1-R_0^2}$$

where R_0^2 is the R^2 about the origin from the regression of X_1 on X_2.

- Under S1-S4, $\widehat{\beta} = Ly$ is BLUE, i.e. it's the Gauss-Markov estimator (conditional on X).

: First moment of $\hat{\sigma}^2$

Define

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n-K} = \frac{u'Mu}{n-K}$$

where $M = I_n - X(X'X)^{-1}X'$. Note that

$$\begin{aligned} tr(M) &= tr(I_n - X(X'X)^{-1}X') \\ &= tr(I_n) - tr(X(X'X)^{-1}X') \\ &= tr(I_n) - tr((X'X)^{-1}X'X) \\ &= tr(I_n) - tr(I_K) = n - K \end{aligned}$$

So

$$E\left((n-K)\widehat{\sigma}^2|X\right) = E(u'Mu|X)$$

$$= E(tr(u'Mu)|X)$$

$$= E(tr(Muu')|X)$$

$$= tr(E(Muu')|X))$$

$$= tr(M\sigma^2 I_n) \text{ by S4}$$

$$= \sigma^2 tr(M)$$

Therefore, under S1-S4

$$E\left(\widehat{\sigma}^2|X\right) = \sigma^2$$

:Specification Error

Suppose you are interested in the model

$$y = X\beta + u$$

But instead you estimate by OLS the model

$$y = Z\gamma + v$$

- Let's assume $E(u|X, Z) = 0$, and identify γ with $E(v|Z) = 0$. What's the relationship between $\gamma/\widehat{\gamma}$ and $\beta/\widehat{\beta}$?

$$\begin{aligned}\widehat{\gamma} &= (Z'Z)^{-1}Z'y \\ &= (Z'Z)^{-1}Z'(X\widehat{\beta} + \widehat{u}) \\ &= (Z'Z)^{-1}Z'(X\beta + u)\end{aligned}$$

Rk: The second equality says that $\widehat{\gamma}_i = \sum_{k=1}^{K} \widehat{\delta}_{ik}\widehat{\beta}_k + \widehat{\lambda}_i$ where $\widehat{\delta}_{ik}$ is the OLS coefficient on Z_i from the regression of X_k on Z, and $\widehat{\lambda}_i$ is its coefficient from the regression of $\widehat{u}$.

1. From the third equality, we obtain

$$E(\widehat{\gamma}|X,Z) = (Z'Z)^{-1}Z'X\beta$$

2. The OLS estimator of the variance of v is given by

$$\widehat{\sigma}_v^2 = \frac{y'M_z y}{tr(M_z)}$$

Therefore, assuming $V(u|X,Z) = \sigma^2 I_n$

$$\begin{aligned} E(\widehat{\sigma}_v^2|X,Z) &= \frac{E((X\beta+u)'M_z(X\beta+u)|X,Z)}{tr(M_z)} \\ &= \frac{E(u'M_z u|X,Z) + E(\beta'XM_zX\beta)|X,Z)}{tr(M_z)} \\ &= \sigma^2 + \frac{\beta'XM_zX\beta}{tr(M_z)} \end{aligned}$$

Interpretation

A. Exclusion of relevant variables.

$$X = \begin{pmatrix} X_1 & X_2 \end{pmatrix} \; Z = X_1$$

1. Then

$$\begin{aligned} \widehat{\gamma} &= (X_1'X_1)^{-1}X_1'(X_1\widehat{\beta}_1 + X_2\widehat{\beta}_2 + \widehat{u}) \\ &= \widehat{\beta}_1 + (X_1'X_1)^{-1}X_1'X_2\widehat{\beta}_2 \end{aligned}$$

Rk: If X_2 is a scalar, then this expression can be written as

$$\widehat{\gamma} = \widehat{\beta}_1 + \widehat{\delta}\widehat{\beta}_2$$

where $\widehat{\delta}$ is the vector of OLS coefficients from the regression of X_2 on X_1.

2. The mean of the OLS estimate of $\widehat{\sigma}_v^2$ satisfies

$$
\begin{aligned}
E(\widehat{\sigma}_v^2 | X, Z) &= \sigma^2 + \frac{\beta' X M_1 X \beta}{tr(M_1)} \\
&= \sigma^2 + \frac{\beta_2' X_2 M_1 X_2 \beta_2}{n - K_1} \\
&\geq \sigma^2
\end{aligned}
$$

- Unless $X_1' X_2 = 0,$ the exclusion of relevant variables will lead to OLS coefficients on X_1 that are biased, and an estimated variance of the error that tends to overestimate σ^2.
- Notice that if X_1 is chosen by some random mechanism that is independent of the sample data (*random treatments)* we can guarantee $E(X_1' X_2) = 0$.

B. Inclusion of irrelevant variables

$$X = X_1 \quad Z = \begin{pmatrix} X_1 & X_2 \end{pmatrix}$$

1. Then

$$\widehat{\gamma} = (X'X)^{-1}X'(X_1\widehat{\beta}_1 + \hat{u}) \quad \text{so}$$

$$\begin{pmatrix} \widehat{\gamma}_1 \\ \widehat{\gamma}_2 \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_1 \\ 0 \end{pmatrix} + (X'X)^{-1}X'\hat{u}$$

Rk: It follows immediately from $E(u|X,Z) = 0$ that $E(\widehat{\gamma}_1|X,Z) = \beta_1$ and $E(\widehat{\gamma}_2|X,Z) = 0$. So including irrelevant variables does not bias the OLS estimator.

2. The mean of the OLS estimate of $\widehat{\sigma}_v^2$ satisfies

$$E(\widehat{\sigma}_v^2|X,Z) = \sigma^2 + \frac{\beta_1' X_1 M X_1 \beta_1}{tr(M)}$$

$$= \sigma^2$$

So the inclusion of irrelevant variables doesn't lead to a bias in the OLS estimator of σ^2.

3. The precision of the OLS estimator of the coefficients on X_1 is made worse by the inclusion of irrelevant variables.

$$\begin{aligned} V(\widehat{\gamma}_1|X,Z) &= \sigma^2(X_1'M_2X_1)^{-1} \\ &\geq \sigma^2(X_1'X_1)^{-1} = V(\widehat{\beta}_1|X,Z) \end{aligned}$$

The inequality reflects the fact that it's only the variation in X_1 that is linearly independent of X_2 that gets used to estimate $\widehat{\gamma}_1$. If X_1 has only one column, then

$$V(\widehat{\gamma}_1|X,Z) = \frac{V(\widehat{\beta}_1|X,Z)}{1 - R_0^2}$$

where R_0^2 is the R^2 about the origin from the regression of X_1 on X_2.