1. identifiable. a parameterization for $EY \in \mathbb{R}^n$

$$\boxed{EY} = f(\beta) = X\beta.$$

Def: β is identifiable for $\forall \beta_1, \beta_2$, if $f(\beta_1) = f(\beta_2) \Rightarrow \beta_1 = \beta_2$.

$g(\beta)$ is identifiable. $\Leftrightarrow$ if $f(\beta_1) = f(\beta_2) \Rightarrow g(\beta_1) = g(\beta_2)$.

$$\boxed{Y} = \boxed{X}\beta + e.$$ knowing EY is not sufficient to tell us the value of $\beta / g(\beta)$.

$\Rightarrow$ $\beta / g(\beta)$ not identifiable.

Regression model. $X^{n\times p}$. $\text{rank}(X) = p$. $Y = X\beta + e$.

β is identifiable.

$$X\beta_1 = X\beta_2.$$

$$\Rightarrow \quad (X'X)^{-1}X' \cdot X\beta_1 = (X'X)^{-1}X' \cdot X\beta_2$$

$$\Rightarrow \quad \beta_1 = \beta_2.$$

$X^{n\times p}$ $\text{rank}(X) < p$. $Y = X\beta + e$.

β is no longer identifiable.

$(X'X)^{-1}$. not exist!

Example: n patients two drugs $\rightarrow$ effect on levels of sth. in blood.

n patients $\downarrow$ μ. two drugs $\downarrow$ β_1, $\downarrow$ β_2.

$$y = \mu + \left\langle \begin{matrix} \beta_1 \\ \beta_2 \end{matrix} \right. + e.$$

$n \rightarrow$ first drug

$n \rightarrow$ second drug.

t.

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \\ y_{n+1} \\ \vdots \\ y_{2n} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & \vdots \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix}}_{X} \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \end{pmatrix} + e.$$

$R(X) \in \mathbb{R}^3$. rank = 2.

$\boxed{\text{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}}$.

$\boxed{g(\beta)} = \Lambda'\beta$.

$\text{rank}(X) = 2 < 3$.

$\vec{\beta}$ is identifiable.

is μ identifiable? No!

$$(\mu, \beta_1, \beta_2) = \begin{cases} \begin{array}{ccc} \mu & \beta_1 & \beta_2 \\ (1, & 10, & 100) \\ (5, & 6, & 96). \end{array} \end{cases} \qquad \begin{cases} \boxed{EY_1 = 11} \\ EY_2 = 101. \end{cases}$$

$\mu + \beta_1 = 11 \qquad \beta_2 - \beta_1 = 90.$

Def: linear function of β, $\Lambda'\beta$ is estimable if $\exists P$.

$$\Lambda'\beta = P'X\beta.$$

$\boxed{\Lambda' \in R(X).}$

estimable $\Rightarrow$ identifiable.

$\not\Leftarrow$.

$$\frac{f_1(\beta)}{f_2(\beta)}$$

μ is estimable? $\quad \mu = (1, 0, 0)\begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \end{pmatrix}.$

$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}. \qquad \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$ No!

$\mu + \beta_1$ is estimable? $\quad \mu + \beta_1 = (1\ 1\ 0)\begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \end{pmatrix}.$

$\beta_2 - \beta_1$ is estimable? $\beta_2 - \beta_1 = (0 \ -1 \ 1)\begin{pmatrix}\mu \\ \beta_1 \\ \beta_2\end{pmatrix}$. ✓

$$\begin{pmatrix}0\\-1\\1\end{pmatrix} = \begin{pmatrix}1\\0\\1\end{pmatrix} - \begin{pmatrix}1\\1\\0\end{pmatrix} \in R(X).$$

X not full rank. $\hat{\beta}$ has infinite solution

to minimize $\|Y - X\beta\|^2 = \sum (y_i - \beta^T x_i)^2$.

Prop: $\cdot$ $X\beta = \sum_{k=1}^{p} \vec{X_k}\beta_k$. $X = \left(\vec{X_1} \ \vec{X_2}, \cdots, \vec{X_p}\right)$.

$\boxed{\beta_i}$ is <u>NOT</u> identifiable (estimable). $\boxed{\Leftrightarrow}$ $\exists \alpha_j \in \mathbb{R}$. s.t.

$X_i = \sum_{j \neq i} \alpha_j X_j$. $\leftarrow$

Covered by effect provided by others.

$r(X) < p$. multicollinearity.

consider linear dependence between columns of X.

Procedure: $Y \sim X_1, \cdots, X_p$.

for $i = 1, \cdots, p$.

$X_i = \beta_0 + \beta_1 X_1 + \cdots + \beta_{i-1} X_{i-1} + \beta_{i+1} X_{i+1} + \cdots + \beta_p X_p + e$.

R_i^2 $\quad VIF_i = \dfrac{1}{1 - R_i^2} \to \infty$ as $R_i^2 \to 1$.

find maximal $\{VIF_i\}_i \longrightarrow$ remove X_i.

for $i = 1, \cdots, p-1$.

$X_i = \beta_0 + \cdots + e$.

rule of thumb $\boxed{\max_i \{VIF_i\}_i > 10}$