Semiparametric Efficiency Bound

The asymptotic variance of any semiparametric estimator is no smaller than the supremum of the Cramer-Rao bounds for all parametric submodels, denoted $\mathcal{V}$.

$\mathcal{V}$: Semiparametric Efficiency Bound

Examples:

1. $\beta_0 = E(Z)$, $\beta_1 = \int z f(z|x) \, dz$.

2. $y_i = x_i \beta_0 + g(y_i, \gamma) + \varepsilon_i$.
   $\varepsilon_i \sim N(0, \sigma^2)$, $\sigma^2$ is known.
   $f(x), g(y) \text{ unknown}$.
   Submodels: $f(x, \eta), g(x, \eta)$.
An estimator \( \hat{\beta} \) is said to be regular in a parametric submodel if, for each \( \theta_0 \),
\[
\sqrt{n} (\hat{\beta} - \beta(\theta_0)) \quad \text{for} \quad \sqrt{n} (\theta_n - \theta_0) \quad \text{bounded,}
\]
has a limiting distribution not depending on the "local" data generating process (ruled out "Hodges" Estimators).

\( \hat{\beta} \) is "regular" if it is regular in every regular parametric submodel and the limiting distribution does not depend on the parametric submodel.

A smooth parametric submodel is one that satisfies a "mean-square" differentiability condition.

A regular parametric submodel is one that is smooth with a nonsingular informativity matrix.
Theorem 2.1. If $\beta$ is regular, then the limiting distribution $\sqrt{n}(\hat{\beta} - \beta_0)$ is equal to the distribution of $Y + U$

where $Y \sim N(0, V)$ and $U$ is independent of $Y$.

Asymptotically Linear Estimators: $(\hat{\beta})$.

$\sqrt{n}(\hat{\beta} - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Psi_i(z_i) + o_p(1)$

$E[\Psi_i] = 0 \quad E[\Psi_i \Psi_j]$ finite and nonsingular

$\Psi_i(z)$: Influence Function.

(Example: $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Psi_i(z_i, \hat{\beta}) = o_p(1)$)

Then

$\Psi_i(z) = -M^{-1} m(z, \beta_0)$

$M = \frac{\partial^2}{\partial \beta^2} E[m(z, \beta)] \bigg|_{\beta = \beta_0}$
Thm 2.2. Suppose that \( \hat{\beta} \) is an asymptotically linear estimator with influence function \( I \), and for all regular parametric submodels, \( \beta(\theta) \) is differentiable and \( E_\theta \| I \|^2 \) exists and is continuous on a neighborhood of \( \theta_0 \), then \( \hat{\beta} \) is regular if and only if, for all parametric submodels.

\[
\frac{\partial \beta(\theta)}{\partial \theta} = E \left[ I' S_\theta \right]
\]

\( \uparrow \)

pathwise derivative.

Example: \[
\gamma = -M^{-1} m(z; \beta_0)
\]

\[
M \cdot \frac{\partial \beta(\theta)}{\partial \theta} = E \left[ m(z; \beta_0) S_\theta \right] = 0
\]

Follows from Total Differentiating (4)

\[
M(\theta; 0) = E_\theta m(\theta; \beta(\theta)) = 0
\]
Calculating the Bound.

A "differentiable" parameter $\beta(t)$:

$\beta(t)$ is differentiable for all parametric submodels and $\exists d$, definite such that for all parametric submodels,

$$\frac{2}{d} \beta(t) = E \left[ d \ S_0' \right]$$

Example $\beta = E z$

$$\beta(t) = \int z f_0(z) \, dz$$

$$\frac{2}{d} \beta(t) = \int z \left( \frac{d}{d} \frac{f_0(z)}{f(z)} \right) f(z) \, dz$$

$S_0'$

$$= E \ z \cdot S_0'$$

so $d = z$
How to find \( d \):

One way is to find the influence function \( \hat{v} \) for an asymptotically linear regular estimator.

Then \( d = -\hat{v} \).

d is not unique

\( d = d + c \) also OK
for any constant \( c \).

For \( \beta \) & a parametric submodel \( \theta \), the Cramer-Rao bound for \( \beta \).

\[
V_{\beta, \theta} = \frac{\partial \beta(\theta)}{\partial \theta} \left( E S_0 S_0' \right)^{-1} \frac{\partial \beta(\theta)}{\partial \theta}'
\]

\[
= E \left[ d S_0' \right] E(S_0 S_0')^{-1} E(S_0 d')
\]

\[
= E d_0 d_0'
\]

For \( d_0 = E(d S_0') E(S_0 S_0')^{-1} S_0 \)

OLS projection of \( d \) on \( S_0 \) space
The tangent set $\mathcal{L}$:

the mean square closure of all linear combinations of scores $S_0$ for smooth parametric submodels:

$$\mathcal{L} = \{ \mathbf{a} \in \mathbb{R}^k : \mathbb{E}[\mathbf{a}^2] < \infty, \exists A_y S_0 \}$$

with

$$\lim_{j \to \infty} \mathbb{E}[\| \mathbf{a} - A_y S_0 \|^2] = 0$$

$$\sum_{j=1}^{\infty} \frac{1}{2^j} ??$$

$\mathcal{D}$

$\mathcal{L}$ is linear if $a_1 \mathbf{a}_1 + b_2 \mathbf{a}_2 \in \mathcal{L}$

for $\mathbf{a}_1, \mathbf{a}_2 \in \mathcal{L}$, $a_1, b_2 \in \mathbb{R}$.

Projection of $\mathbf{d}$ on $\mathcal{L}$ is $\Delta$

such that $\Delta \in \mathcal{L}$ and

$$\mathbb{E}[(\mathbf{d} - \Delta)^\prime \mathbf{a}] = 0 \quad \forall \mathbf{a} \in \mathcal{L}$$
Thm 3.1 Suppose that the parameter is differentiable, \( \mathcal{L} \) is linear, and \( \mathcal{E} \mathcal{S} \) is nonsingular. for the projection of \( \delta \) of \( \mathcal{L} \) onto \( \mathcal{V} \). then \( \mathcal{V} = \mathcal{E}[\delta \delta'] \)

\( \delta \) is called the "Efficient Influence Function"

Example of \( \mathcal{E} \mathcal{L} \) and \( \mathcal{T} \).

\( \beta = E \beta \) \hspace{1cm} \beta(0) = E \beta E \).

Without any restriction on \( S \).

\( \mathcal{L} \) is the space of all mean vector random vectors. \( \{ E S_0 = 0 \} \) \hspace{1cm} \forall \theta \).

\( d = 2 \) \hspace{1cm} \text{guess} \hspace{1cm} \delta = \beta - E \beta \).

Ok since \( \delta \in \mathcal{L} \) (mean zero).

(1) \( E(\delta - \delta') \delta = E(E \beta) \delta \).

\( = E \beta E \delta = 0 \) since \( E \delta = 0 \)
Now consider a "semiparametric" model with a parametric component $\beta$ and a nonparametric component $\gamma$.

$\Theta = (\beta, \gamma) \quad S_\theta = (S_\beta, S_\gamma)$

Now

$$\frac{\partial \beta(\theta)}{\partial \theta} = \begin{bmatrix} I_k & 0 \end{bmatrix}$$

In order for

$$\frac{\partial \beta(\theta)}{\partial \theta} = \begin{bmatrix} I_k & 0 \end{bmatrix}$$

$$E(d, S_\theta)$$

$$\begin{bmatrix} S_\beta & S_\gamma \end{bmatrix}$$

$\Rightarrow$

$$\begin{cases} E\left[ dS_\beta \right] = I_k, \\ E\left[ dS_\gamma \right] = 0 \end{cases}$$

How can such $d$ be found?

and if such $d \in L_2$, then

$$d = \delta : \text{ the efficient score}$$
Begin with \( S_\beta \), project \( S_\beta \) on the space spanned by \( S_\eta \) (can be \( \infty \) dimension).

Take the residual \( \Rightarrow \) call it \( S' \).

Then

\[
E S S_\eta' = 0 \quad \forall S_\eta.
\]

If \( E S S_\beta' \) nonsingular

Define \( d = E(S S_\beta')^{-1} S \).

\( d \) satisfies both

\[
\begin{cases}
E d S_\beta' = I_k \\
E d S_\eta' = 0
\end{cases}
\]

Ditto. Also \( d \) is a linear combination of \( S_\alpha \) and \( S_\eta \).

So \( d \in \mathcal{L} \).

Hence \( \delta = d \), Efficient score!
Tangent set in the nonparametric (1) dimension (only): \( T \)

linear span of \( S \).

\( S \): the residual of projecting \( S_\beta \) onto \( T \), so that

1. \( S_\beta - S \in T \)
2. \( E[SS'] = 0 \) for all \( l \in T \)

Thm 3.2. If \( f(z|p) \) is smooth with score \( S_\beta \), \( T \) is linear, \( S \) satisfies \( E[SS'] \) nonsingular, then \( \beta \) is a differentiable parameter and has efficient influence function

\[ d = E[SS']^{-1}S, \quad \text{with} \]

\[ V = (ESS')^{-1}. \]
Three key steps:

1. Set up a sub-parametric model. Compute the Scores $S_B$, $S_N$.

2. Characterize the tangent set $T$ of $S_N$'s.

3. Compute $S'$, the residual of projecting $S_B$ onto $T$.

To compute 3, need lemmas and special cases.
Lemma 3.4: If \( WU \) has finite 2nd moment and \( V \) and \( W \) are functions of some \( T \) such that \( E[UU' | T] \) is constant and positive definite, then the projection of \( WU \) on \( T \)
\[ T_V = \{ D(V)u : E[1D(V)^2] < \infty \} \]
is \( E[W | V]U \).

Proof:

1. \( E[W | V] U \in T_V \).

2. \[
E\left[(WU - E[W|V]U)D(V)\right]
= E\left(W - E[W|V]\right) E[uu' | T] D(V)' \\
= E\left[W - E[W|V]\right] C\cdot D(V)',
\]

\( = 0 \)
Example of a semiparametric regression

\[ y = x' \beta + g_0(v) + \varepsilon \]

\[ \varepsilon \sim N(0, \sigma_0^2) \quad \sigma_0^2 \text{ known.} \]

\[ g_0(v) \quad \text{f(\text{any}) unknown.} \]

\[ \text{Data} \quad g_0(v; \eta) \quad \text{f} \left( f_0(x, y, \eta) \right) \rightarrow \text{suppose known} \]

\[ \theta = \{ \beta, \eta \} \]

\[ \ln f(y, x, \theta) \]

\[ = C - \frac{1}{2} 2 \ln \sigma_0^2 - \frac{(y - x' \beta - g_0(v; \eta))^2}{2 \sigma_0^2} + \ln f_0(x, v) \]

then

\[ S_\beta = \frac{x' \varepsilon}{\sigma_0^2} \quad S_\eta = \frac{g_0' \varepsilon}{\sigma_0^2}. \]

\[ g_\eta = \frac{2 g_0(v; \eta_0)}{\sigma_0^2} \quad \varepsilon = y - x' \beta - g_0(v) \]

\[ \rightarrow \text{(unrestricted function) of } \nu. \]

\[ J = \{ \varepsilon \in D(\nu) : \quad \mathbb{E} [\varepsilon^2 (D(\nu))^2] < \infty \} \]

Sub model

\[ g_0(v; \eta) = g_0(v) + \eta' \left[ \sigma_0^2 D(v) \right]. \]
Apply Lemma 3.4.

\[ U = \Sigma, \quad W = \sigma_\xi^{-2} \pi, \quad V = \nu, \quad T = (x, \nu) \]

projection \( \tilde{S}_p = \frac{x \xi}{\sigma_\xi^2} \) on \( \mathcal{Y} = \{ \xi \in D(\nu) \} \) is

\[ \frac{\xi}{\sigma_\xi^2} E[x \mid \mathcal{W}] \]

So \( S = S_p - \text{projection} = \frac{\xi}{\sigma_\xi^2} (x - E[x \mid \nu]) \)

\( \downarrow \) the efficient score.

Allowing for \( \sigma_\xi^2 \) & \( f(x, \nu) \) unknown does not change the bound.
Example 2: Linear Regression with Independent Error.

\[ y = x \beta + \varepsilon \quad \varepsilon \perp \!
\perp x \]

\[ \varepsilon \sim f_\theta (\varepsilon) \quad \text{but unknown.} \]

\[
\ln f(y, x, \theta) = \ln f(y-x\beta, \eta) + \ln f(x, \eta) 
\]

\[ S_\beta = -x S(\varepsilon) = -x \frac{f_\varepsilon (\varepsilon)}{f(\varepsilon)} \]

\[ S_\eta = \frac{f_\eta (\varepsilon, \eta_0)}{f(\varepsilon)} + \frac{f_\eta (x, \eta_0)}{f(x)} \]

\[ = D(\varepsilon) + D(x) \]

Tangent set.

\[ Y = \left\{ D(\varepsilon) + D(x) \mid E(D(\varepsilon)) = 0, \ E(D(x)) = 0 \right\} \]

both in \( \mathbb{R}^2 \)

1. \( D(\varepsilon) \perp D(x) \)
2. \( S_\beta = -x S(\varepsilon) \perp D(x) \)
3. Project \( S_\beta \) onto \( D(\varepsilon) \)

\[ S^* = S_\beta - E(S^2 \mid \varepsilon) = -\{x - E(x \mid \varepsilon)\} S(\varepsilon) \]
Adaptive Estimation: if $\beta \perp \gamma$

or for a subversion \( \beta \) if $\beta = (\beta, \alpha)$

if

\[
\hat{S}_\beta = E[\hat{S}_\beta S_\alpha] E(s_\alpha s_\alpha)^{-1} S_\alpha \perp \gamma
\]

In the independent $\varepsilon$ Regression model

\( \alpha \rightarrow \text{Intercept} \)

\[
\hat{S}_\beta = -\alpha \hat{s}(\varepsilon) \quad S_\alpha = -\hat{s}(\varepsilon)
\]

\[
\hat{S}_\beta = E[\hat{S}_\beta S_\alpha] E(s_\alpha s_\alpha)^{-1} E S_\alpha
\]

\[
= \hat{s}(\varepsilon) \left\{ x - \frac{E(s(\varepsilon) x)}{E(s(\varepsilon)^2)} \right\}
\]

\[
= -\hat{s}(\varepsilon) \left\{ x - E(x) \right\} \text{ by independence}
\]

Same as $S$, therefore

\( \perp \gamma \)
Extension 1. (FIML).

\[ \rho(z; \beta_0) = \xi, \quad \chi \perp \xi. \]
\[ z = (y, \chi). \]
\[ \rho(z; \beta_0) \text{ is one-to-one function of } y. \]

Let \[ J(z; \beta) = \ln | \det (P_y(z; \beta)) |. \]
\[ S(\xi) = \frac{f_\xi(\xi)}{f(\xi)}. \]
\[ D(\xi) = \frac{f_\eta(\xi)}{f(\xi)} \quad D(x) = \frac{f_\eta(x)}{f(x)}. \]

\[ \ln f(z; \theta) = \underbrace{J(z; \beta)}_{\gamma / \xi \mid \chi} + \ln f(P(z; \beta), \eta) + \ln f(x, \eta) \]

\[ S_\beta = J_\beta (z) + \rho_\beta (z) S(\xi) \quad S_\eta = D(\xi) + D(x). \]

\[ D(\xi) \perp D(x) \quad \text{(for fixed } \gamma) \]

1. \[ S_\beta \perp D(x) \quad \text{since} \quad \mathbb{E} [ S_\beta \mid x ] = 0. \]

2. \[ S = S_\beta - \mathbb{E} [ S_\beta \mid \xi ] = J_\beta (z) - \mathbb{E} [ J_\beta (z) \mid \xi ] - S(\xi), \quad \rho_\beta (z) - \mathbb{E} [ \rho_\beta (z) \mid \xi ] \]
Example 2. ($\hat{L} / ML$).

Now let $Z = (y, u, x)$.

$P(Z; \beta) = \mathcal{E} \mathcal{Z} + x$.

If one to one function of $y$.

$\ln f(Z; \theta)$

$= J(Z; \beta) + \ln f(\hat{P}(Z; \beta), \eta)$

$+ \ln f(x, \eta) + \ln h(\omega | x, \hat{P}(z, \beta), \eta)$

Why is this the likelihood??
Example 3: binary choice

\[ y = \text{I}(x'\beta_0 - \varepsilon > 0) \quad \xi \perp x \]

\[ f_0(\varepsilon) \text{ unknown.} \]

path: \( F(\xi, \eta), f(x, \eta) \)

\[ f(y, x, \varepsilon) = y \ln F(x'\beta, \eta) + (1 - y) \ln (1 - F(x'\beta, \eta)) + f(x, \eta) \]

\[ S_\beta = \sigma(x'\beta)^{-2} (y - F(x'\beta, \eta)) f_0(x'\beta) \times \]

\[ \sigma(x'\beta) = \left( F(x'\beta) (1 - F(x'\beta)) \right)^{1/2} \]

\[ S_\eta = \sigma(x'\beta_0)^{-2} (y - F(x'\beta_0)) \frac{f_0(x'\beta_0, \eta_0)}{f(x)} \text{ D}(x) \]

\[ D(x) = \frac{f(x)}{f(x)} \text{ same function.} \]

i) \[ J = \left\{ \frac{(y - F(x'\beta)) \text{ D}(x)}{\sigma(x'\beta)^{-1}} \right\} \text{ (anything) (mean 0) } \]

Lemma 3.4 again

\[ U = \sigma(x'\beta)^{-1} (y - F(x'\beta)) \]

\[ W = \sigma(x'\beta)^{-1} f(x'\beta) x \]

\[ V = x'\beta \]
Also \( S_\beta \perp D_2 (x) \).

So,
\[
S = S_\beta - \text{Proj}[S_\beta | \frac{y - F(x_\beta)}{\sigma(x_\beta)} D_2 (x_\beta)]
\]
\[
= S_\beta - U \cdot \mathbb{E}[W | V]
\]
\[
= S_\beta - \frac{y - F(x_\beta)}{\sigma(x_\beta)} \cdot \frac{f(x_\beta)}{\sigma(x_\beta)} \mathbb{E}[x | x_\beta]
\]
\[
= \frac{(y - F(x_\beta)) f(x_\beta)}{\sigma^2(x_\beta)} \left( x - \mathbb{E}[x | x_\beta] \right).
\]

Another approach.

Chamberlain (1987)

Use multinomial distribution as approximation of unknown distributions.