MLE, a special case of
\[ \frac{1}{n} \sum_{i=1}^{n} h(z_i, \hat{\beta}) = 0 \]

where
\[ h(z_i, \beta) = \frac{-\ln f(z_i, \beta)}{\sigma \beta}. \]

\[ \sqrt{n} \left( \hat{\beta}(\tau) - \beta \right) = \left( \mathbb{E} \left( \frac{\partial h(z_i, \hat{\beta})}{\partial \beta} \right) \right)^{-1} \frac{\sqrt{n}}{n} \sum_{i=1}^{n} h(z_i, \beta) \]

\[ D(\tau) = \frac{\partial}{\partial \beta} \mathbb{E} h(z_i, \beta) \]

\[ m(\tau) = h(z_i, \beta_0) \]

\[ m(\tau) = \frac{\partial \ln f(z_i, \beta_0)}{\partial \beta} \]

Efficiency condition
\[ \mathbb{E} \left[ \frac{\partial h(z_i, \beta_0)}{\partial \beta} \right] = \mathbb{E} h(z_i, \beta_0) \frac{\partial \ln f(z_i, \beta_0)}{\partial \beta} \]

\[ D(\tau) = \mathbb{E} m(\tau) m(\tau)^\top \]

for all \( h(\cdot) \)
Efficient Choice of Instruments

Consider \( Y_+ = X_+ \beta + \varepsilon_+ \). In what sense is \( \varepsilon_+ \) "orthogonal" to an instrument \( Z_+ \)?

1. \( E(\varepsilon_+ Z_+) = 0 \) \( \implies \) \( \text{IVLS} \)

2. \( E[\varepsilon_+ | Z_+] = 0 \) conditional moment model.

3. \( \varepsilon_+ \perp Z_+ \)

\( \circ \) is the common assumption.

More generally,

\[
\begin{align*}
E( Y_+ - X_+ \beta | Z_+) &= 0, \\
E( \Phi( Y_+, X_+, \beta ) | Z_+) &= 0.
\end{align*}
\]
If $E(y_r - x_i \beta | z_r) = E(\varepsilon_r | z_r) = 0$

then for any function $A(z_r)$

$$E A(z_r) \varepsilon_r = E \left[ E(\varepsilon_r | z_r) A(z_r) \right] = 0$$

$$\Rightarrow E A(z_r) (y_r - x_i \beta) = 0$$

or $E A(z_r) P(y_r, x_i, \beta) = 0$

Without loss of generality, set

$$\dim(A(z_r)) = \dim(P)$$

$$\frac{1}{n} \sum_{i=1}^{n} A(z_r) P(y_r, x_i, \beta) = 0$$

$$\sqrt{n} \left( \hat{\beta} - \beta \right) = \left( \frac{\partial}{\partial \beta} E A(z_r) P(y_r, x_i, \beta) \right)^{-1} \sqrt{n} \sum_{i=1}^{n} A(z_r) P(y_r, x_i, \beta) + o_p(1)$$

$$\Phi(T) = \Phi(A(t)) = D(T)^{-1} m(T)$$
Where

\[ D(\tau) = E \frac{\partial}{\partial \beta} A(z) \ P(y, x, \beta) \]

\[ = E A(z) \ \frac{\partial}{\partial \beta} (y, x, \beta) \]

if \ \Phi(\cdot) = y - x, \beta

\[ D(\tau) = E A(z) \cdot X \]

\[ m(\tau) = A(z) \ P(y, x, \beta) = A(z) \tilde{\Xi} \]

\[ E m(\tau) \ m(\tau) = E A(z) \tilde{\Xi} \tilde{\Xi}^t A(z)^t \]

\[ = E A(z) \ \Phi(z) \ A(z)^t \]

\[ = E \sigma^2(z) A(z) A(z)^t \]

if \ \tilde{\Xi} \text{ is a scalar}.
Efficiency condition.

\[ D(\tau) = \text{E} m(\tau) m(\tau) \]

\[ \frac{\partial}{\partial \beta} \left( y_+^2 + \beta^2 \right) \]

\[ \text{E} A(z_+)^2 \sigma^2(z_+) \]

\[ A(z_+) = \sigma^2(z_+) \frac{\partial}{\partial \beta} \left( y_+^2 + \beta^2 \right) \]

If \( \sigma^2(z_+) = \sigma^2(z_+) \)

\[ A(z_+) = \sigma^2(z_+) \frac{\partial}{\partial \beta} \left( y_+^2 + \beta^2 \right) \]

If also \( \rho(y_+, x_+, \beta) = y_+ - x_+ \beta \)

Then \( A(z_+) = \sigma^2(z_+) \frac{\partial}{\partial \beta} \left( y_+^2 + \beta^2 \right) \)

In principle, both

\[ \sigma^2(z_+) = \text{E} (\epsilon^2 | z_+) \]

and \( \text{E} (x_+ | z_+) \) has to be estimated nonparametrically.
Special cases

1. $\sigma^2(z_t) = \sigma^2$

\[A(z_t) = \mu_{z_t}X_t^t] \]

2SLS is efficient only if $E[z_{t}\mid X_t]$ happens to be linear.

2. $X_t = \epsilon_t$

\[A(z_t) = \sigma^2(X_t)^{-1}X_t \]

Weighted LS

3. $X_t = \epsilon_t$, $\sigma^2(X_t) = \sigma^2$

then OLS is actually efficient.

However, none of these situations guarantee that OLS is not efficient if $E \epsilon_t X_t$