● **Theorem:**

Assume that D_i is independent of Y_i, given X_i, for all i.

Suppose that D_i, Y_i and X_i are identically distributed.

Then, we have the following theorem:

$$\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{D_i Y_i}{\pi(X_i)} - \frac{(1-D_i)Y_i}{1-\pi(X_i)}\Big) \xrightarrow{\ p\ } \mathrm{E}(Y^1 - Y^0)$$

Proof:

$$\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{D_i Y_i}{\pi(X_i)} - \frac{(1-D_i)Y_i}{1-\pi(X_i)}\Big) \xrightarrow{\ p\ } \mathrm{E}\Big(\frac{DY}{\pi(X)} - \frac{(1-D)Y}{1-\pi(X)}\Big) = \mathrm{E}\Big(\frac{DY^1}{\pi(X)} - \frac{(1-D)Y^0}{1-\pi(X)}\Big)$$

$$= \mathrm{E}\Big(\mathrm{E}\Big[\frac{DY^1}{\pi(X)} - \frac{(1-D)Y^0}{1-\pi(X)}\Big|X\Big]\Big) = \mathrm{E}\Big(\frac{\mathrm{E}(DY^1|X)}{\pi(X)} - \frac{\mathrm{E}[(1-D)Y^0|X]}{1-\pi(X)}\Big)$$

$$= \mathrm{E}\Big(\frac{\mathrm{E}(D|X)\mathrm{E}(Y^1|X)}{\pi(X)} - \frac{\mathrm{E}(1-D|X)\mathrm{E}(Y^0|X)}{1-\pi(X)}\Big) = \mathrm{E}\Big(\frac{\pi(X)\mathrm{E}(Y^1|X)}{\pi(X)} - \frac{[1-\pi(X)]\mathrm{E}(Y^0|X)}{1-\pi(X)}\Big)$$

$$= \mathrm{E}(\mathrm{E}[Y^1|X] - \mathrm{E}[Y^0|X]) = \mathrm{E}(\mathrm{E}[Y^1 - Y^0|X]) = \mathrm{E}(Y^1 - Y^0)$$

Line 1:

- $\xrightarrow{p}$ utilizes the **law of large number**

The 1st equality comes from the definition of Y, i.e., $Y = Y^1$ for $D = 1$ and $Y = Y^0$ for $D = 0$.

Line 2:

- The 1st equality holds by the **law of iterated expectation (繰り返し期待値の法則)**. The expectation inside indicates the conditional expectation given X. The expectation outside indicates the unconditional expectation with respect to X.
- The 2nd equality implies that $\pi(X)$ goes out of the expectation of X because $\pi(X)$ is a function of X. We have the conditional expectations given X in the numerater of two terms.

Line 3:

- The 1st equality utilizes the assumption that D given X is independent of Y given X.

  $\longrightarrow$ This assumption called **SUTVA** (stable unit value assumption).

- The 2nd equality replaces the conditional expectation $\mathrm{E}(D|X)$ by the propensity score $\pi(X)$.

Line 4:

- The 1st equality utilizes that the numerator and denominator are canceled out for the two terms.
- The 2nd equality indicates that the expectation of difference between two random variables is equivalent to difference betweem two expectations.
- The 3rd equality is obtained by the law of iterated expectation.

5 Generalized Method of Moments (GMM, 一般化積率法)

5.1 Method of Moments (MM, 積率法)

As $n \longrightarrow \infty$, we have the result: $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow \mathrm{E}(X) = \mu$.

$\Longrightarrow$ **Law of Large Number (大数の法則)**

$X_1, X_2, \cdots, X_n$ are n realizations of X.

[Review] Chebyshev's inequality (チェビシェフの不等式) is given by:

$$P(|X-\mu| > \epsilon) \le \frac{\sigma^2}{\epsilon^2} \quad \text{or} \quad P(|X-\mu| \le \epsilon) \ge 1 - \frac{\sigma^2}{\epsilon^2},$$

where $\mu = \mathrm{E}(X)$, $\sigma^2 = \mathrm{V}(X)$ and any $\epsilon > 0$.

Note that $P(|X-\mu| > \epsilon) + P(|X-\mu| \le \epsilon) = 1$.

Replace X, $\mathrm{E}(X)$ and $\mathrm{V}(X)$ by $\overline{X}$, $\mathrm{E}(\overline{X}) = \mu$ and $\mathrm{V}(\overline{X}) = \dfrac{\sigma^2}{n}$.

As $n \longrightarrow \infty$,

$$P(|\overline{X} - \mu| \leq \epsilon) \geq 1 - \frac{\sigma^2}{n\epsilon^2} \longrightarrow 1.$$

That is, $\overline{X} \longrightarrow \mu$ as $n \longrightarrow \infty$.

[End of Review]

$\overline{X}$ is an approximation of $\mathrm{E}(X) = \mu$.

Therefore, $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ is taken as an estimator of μ.

$\Longrightarrow$ $\overline{X}$ is MM estimator of $\mathrm{E}(X) = \mu$.

MM is applied to the regression model as follows:

Regression model: $y_i = x_i\beta + u_i$, where x_i and u_i are assumed to be stochastic.

Familiar Assumption: $\mathrm{E}(x'u) = 0$, called the **orthogonality condition (直交条件)**, where x is a $1 \times k$ vector and u is a scalar.

We consider that $(x_1, x_2, \cdots, x_n)$ and $(u_1, u_2, \cdots, u_n)$ are realizations generated from random variables x and u, respectively.

From the law of large number, we have the following:

$$\frac{1}{n}\sum_{i=1}^{n} x_i' u_i = \frac{1}{n}\sum_{i=1}^{n} x_i'(y_i - x_i\beta) \longrightarrow \mathrm{E}(x'u) = 0.$$

Thus, the MM estimator of β, denoted by β_{MM}, satisfies:

$$\frac{1}{n}\sum_{i=1}^{n} x_i'(y_i - x_i\beta_{MM}) = 0.$$

Therefore, β_{MM} is given by:

$$\beta_{MM} = \Big(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n} x_i' y_i\Big) = (X'X)^{-1}X'y,$$

which is equivalent to OLS and MLE.

Note that X and y are:

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

- However, β_{MM} is inconsistent when $\mathrm{E}(x'u) \neq 0$, i.e.,

$$\beta_{MM} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u = \beta + \Big(\frac{1}{n}X'X\Big)^{-1}\Big(\frac{1}{n}X'u\Big) \not\longrightarrow \beta.$$

Note as follows:

$$\frac{1}{n}X'u = \frac{1}{n}\sum_{i=1}^{n} x_i'u_i \;\longrightarrow\; \mathrm{E}(x'u) \neq 0.$$

In order to obtain a consistent estimator of β, we find the instrumental variable z which satisfies $\mathrm{E}(z'u) = 0$.

Let z_i be the ith realization of z, where z_i is a $1 \times k$ vector.

Then, we have the following:

$$\frac{1}{n}Z'u = \frac{1}{n}\sum_{i=1}^{n} z_i'u_i \;\longrightarrow\; \mathrm{E}(z'u) = 0.$$

The β which satisfies $\displaystyle\frac{1}{n}\sum_{i=1}^{n} z_i'u_i = 0$ is denoted by β_{IV}, i.e., $\displaystyle\frac{1}{n}\sum_{i=1}^{n} z_i'(y_i - x_i\beta_{IV}) = 0$.

Thus, β_{IV} is obtained as:

$$\beta_{IV} = \Big(\frac{1}{n}\sum_{i=1}^{n} z_i' x_i\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n} z_i' y_i\Big) = (Z'X)^{-1}Z'y.$$

Note that $Z'X$ is a $k \times k$ square matrix, where we assume that the inverse matrix of $Z'X$ exists.

Assume that as n goes to infinity there exist the following moment matrices:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} z_i' x_i &= \frac{1}{n}Z'X \longrightarrow M_{zx}, \\
\frac{1}{n}\sum_{i=1}^{n} z_i' z_i &= \frac{1}{n}Z'Z \longrightarrow M_{zz}, \\
\frac{1}{n}\sum_{i=1}^{n} z_i' u_i &= \frac{1}{n}Z'u \longrightarrow 0.
\end{aligned}$$

As n goes to infinity, β_{IV} is rewritten as:

$$\begin{aligned}\beta_{IV} &= (Z'X)^{-1}Z'y = (Z'X)^{-1}Z'(X\beta + u) = \beta + (Z'X)^{-1}Z'u \\ &= \beta + (\frac{1}{n}Z'X)^{-1}(\frac{1}{n}Z'u) \longrightarrow \beta + M_{zx} \times 0 = \beta,\end{aligned}$$

Thus, β_{IV} is a consistent estimator of β.

- We consider the asymptotic distribution of β_{IV}.

By the central limit theorem,

$$\frac{1}{\sqrt{n}}Z'u \longrightarrow N(0, \sigma^2 M_{zz})$$

Note that $\mathrm{V}(\frac{1}{\sqrt{n}}Z'u) = \frac{1}{n}\mathrm{V}(Z'u) = \frac{1}{n}\mathrm{E}(Z'uu'Z) = \frac{1}{n}\mathrm{E}\Big(\mathrm{E}(Z'uu'Z|Z)\Big)$

$= \frac{1}{n}\mathrm{E}\Big(Z'\mathrm{E}(uu'|Z)Z\Big) = \frac{1}{n}\mathrm{E}(\sigma^2 Z'Z) = \mathrm{E}(\sigma^2\frac{1}{n}Z'Z) \longrightarrow \mathrm{E}(\sigma^2 M_{zz}) = \sigma^2 M_{zz}.$

We obtain the following asymmptotic distribution:

$$\sqrt{n}(\beta_{IV} - \beta) = (\frac{1}{n}Z'X)^{-1}(\frac{1}{\sqrt{n}}Z'u) \longrightarrow N(0, \sigma^2 M_{zx}^{-1} M_{zz} M_{zx}^{-1\prime})$$

Practically, for large n we use the following distribution:

$$\beta_{IV} \sim N\Big(\beta, s^2(Z'X)^{-1}Z'Z(Z'X)^{-1\prime}\Big),$$

where $s^2 = \dfrac{1}{n-k}(y - X\beta_{IV})'(y - X\beta_{IV})$.

- In the case where z_i is a $1 \times r$ vector for $r > k$, $Z'X$ is a $r \times k$ matrix, which is not a square matrix. $\Longrightarrow$ **Generalized Method of Moments (GMM, 一般化積率法)**

5.2 Generalized Method of Moments (GMM, 一般化積率法)

In order to obtain a consistent estimator of β, we have to find the instrumental variable z which satisfies $\mathrm{E}(z'u) = 0$.

For now, however, suppose that we have z with $\mathrm{E}(z'u) = 0$.

Let z_i be the ith realization (i.e., the ith data) of z, where z_i is a $1 \times r$ vector and $r > k$.

Then, using the law of large number, we have the following:

$$\frac{1}{n}Z'u = \frac{1}{n}\sum_{i=1}^{n} z_i'u_i = \frac{1}{n}\sum_{i=1}^{n} z_i'(y_i - x_i\beta) \longrightarrow \mathrm{E}(z'u) = 0.$$

The number of equations (i.e., r) is larger than the number of parameters (i.e., k).

Let us define W as a $r \times r$ weight matrix, which is symmetric.

We solve the following minimization problem:

$$\min_{\beta} \Big(\frac{1}{n}\sum_{i=1}^{n} z_i'(y_i - x_i\beta)\Big)' W \Big(\frac{1}{n}\sum_{i=1}^{n} z_i'(y_i - x_i\beta)\Big),$$

which is equivalent to:

$$\min_{\beta} \Big(Z'(y - X\beta)\Big)' W \Big(Z'(y - X\beta)\Big),$$

i.e.,

$$\min_{\beta} (y - X\beta)'ZWZ'(y - X\beta).$$

Note that $\sum_{i=1}^{n} z_i'(y_i - x_i\beta) = Z'(y - X\beta)$.

W should be the inverse matrix of the variance-covariance matrix of $Z'(y-X\beta) = Z'u$.

Suppose that $\mathrm{V}(u) = \sigma^2\Omega$.

Then, $\mathrm{V}(Z'u) = \mathrm{E}(Z'u(Z'u)') = \mathrm{E}(Z'uu'Z) = Z'\mathrm{E}(uu')Z = \sigma^2 Z'\Omega Z = W^{-1}$.

The following minimization problem should be solved.

$$\min_{\beta}\ (y - X\beta)'Z(Z'\Omega Z)^{-1}Z'(y - X\beta).$$

The solution of β is given by the GMM estimator, denoted by β_{GMM}.

Remark: For the model: $y = X\beta + u$ and $u \sim (0, \sigma^2\Omega)$, solving the following minimization problem:

$$\min_{\beta}(y - X\beta)'\Omega^{-1}(y - X\beta),$$

GLS is given by:

$$b = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$

Note that b is the best linear unbiased estimator.

Remark: The solution of the above minimization problem is equivalent to the GLE estimator of β in the following regression model:

$$Z'y = Z'X\beta + Z'u,$$

where Z, y, X, β and u are $n \times r$, $n \times 1$, $n \times k$, $k \times 1$ and $n \times 1$ matrices or vectors. Note that $r > k$.

$y^* = Z'y$, $X^* = Z'X$ and $u^* = Z'u$ denote $r \times 1$, $r \times k$ and $r \times 1$ matrices or vectors, where $r > k$.

Rewrite as follows:

$$y^* = X^*\beta + u^*,$$

$\Longrightarrow$ r is taken as the sample size.

u^* is a $r \times 1$ vector.