6. The variance of β_{IV} is given by:

$$\mathrm{V}(\beta_{IV}) = s^2 (Z'X)^{-1} Z'Z (X'Z)^{-1},$$

where

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

12.3 Two-Stage Least Squares Method (2段階最小二乗法, 2SLS or TSLS)

1. Regression Model:

   $$y = X\beta + u, \quad u \sim N(0, \sigma^2 I),$$

   In the case of $\mathrm{E}(X'u) \neq 0$, OLSE is not consistent.

2. Find the variable Z which satisfies $\frac{1}{n} Z'u \longrightarrow M_{zu} = 0$.

3. Use $Z = \hat{X}$ for the instrumental variable.

$\hat{X}$ is the predicted value which regresses X on the other exogenous variables, say W.

That is, consider the following regression model:

$$X = WB + V.$$

Estimate B by OLS.

Then, we obtain the prediction:

$$\hat{X} = W\hat{B},$$

where $\hat{B} = (W'W)^{-1} W'X$.

Or, equivalently,

$$\hat{X} = W(W'W)^{-1} W'X.$$

$\hat{X}$ is used for the instrumental variable of X.

4. The IV method is rewritten as:

   $$\beta_{IV} = (\hat{X}'X)^{-1} \hat{X}'y = (X'W(W'W)^{-1} W'X)^{-1} X'W(W'W)^{-1} W'y.$$

   Furthermore, β_{IV} is written as follows:

   $$\beta_{IV} = \beta + (X'W(W'W)^{-1} W'X)^{-1} X'W(W'W)^{-1} W'u.$$

Therefore, we obtain the following expression:

$$\sqrt{n}(\beta_{IV} - \beta) = \Big(\big(\frac{1}{n} X'W\big) \big(\frac{1}{n} W'W\big)^{-1} \big(\frac{1}{n} XW'\big)' \Big)^{-1} \big(\frac{1}{n} X'W\big) \big(\frac{1}{n} W'W\big)^{-1} \big(\frac{1}{\sqrt{n}} W'u\big)$$
$$\longrightarrow N\Big(0, (M_{xz} M_{zz}^{-1} M_{xz}')^{-1}\Big).$$

5. Clearly, there is no correlation between W and u at least in the limit, i.e.,

   $$\mathrm{plim}\big(\frac{1}{n} W'u\big) = 0.$$

6. **Remark:**

   $$\hat{X}'X = X'W(W'W)^{-1} W'X = X'W(W'W)^{-1} W'W(W'W)^{-1} W'X = \hat{X}'\hat{X}.$$

Therefore,

$$\beta_{IV} = (\hat{X}'X)^{-1} \hat{X}'y = (\hat{X}'\hat{X})^{-1} \hat{X}'y,$$

which implies the OLS estimator of β in the regression model: $y = \hat{X}\beta + u$ and $u \sim N(0, \sigma^2 I_n)$.

13 Large Sample Tests

13.1 Wald, LM and LR Tests

$\theta : K \times 1$

$h(\theta) : G \times 1$ vector function, $G \leq K$

$\theta : K \times 1$

The null hypothesis $H_0 : h(\theta) = 0 \implies G$ restrictions

$\tilde{\theta} : k \times 1$, restricted maximum likelihood estimate

$\hat{\theta} : k \times 1$, unrestricted maximum likelihood estimate

$I(\theta) : k \times k$, information matrix, i.e.,

$$I(\theta) = -\mathrm{E}\Big(\frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta'}\Big).$$

$\log L(\theta)$: log-likelihood function

$R_\theta = \dfrac{\partial h(\theta)}{\partial\theta'} : G \times k$

$F_\theta = \dfrac{\partial \log L(\theta)}{\partial\theta} : k \times 1$

1. **Wald Test (ワルド検定):** $\quad W = h(\hat{\theta})'\Big(R_{\hat{\theta}}(I(\hat{\theta}))^{-1}R'_{\hat{\theta}}\Big)^{-1}h(\hat{\theta})$

   (a) $h(\theta) \approx h(\hat{\theta}) + \dfrac{\partial h(\hat{\theta})}{\partial\theta'}(\theta - \hat{\theta}) \quad \Longleftarrow \quad h(\theta)$ is linearized around $\theta = \hat{\theta}$.

Under the null hypothesis $h(\theta) = 0$,

$$h(\hat{\theta}) \approx \frac{\partial h(\hat{\theta})}{\partial\theta'}(\hat{\theta} - \theta) = R_{\hat{\theta}}(\hat{\theta} - \theta)$$

(b) $\hat{\theta}$ is MLE.

From the properties of MLE,

$$\sqrt{n}(\hat{\theta} - \theta) \longrightarrow N\Big(0, \lim_{n\to\infty}\Big(\frac{I(\theta)}{n}\Big)^{-1}\Big),$$

That is, approximately, we have the following result:

$$(\hat{\theta} - \theta) \sim N\Big(0, (I(\theta))^{-1}\Big).$$

(c) The distribution of $h(\hat{\theta})$ is approximately given by:

$$h(\hat{\theta}) \sim N\Big(0, R_{\hat{\theta}}(I(\theta))^{-1}R'_{\hat{\theta}}\Big)$$

(d) Therefore, the $\chi^2(G)$ distribution is derived as follows:

$$h(\hat{\theta})\Big(R_{\hat{\theta}}(I(\theta))^{-1}R'_{\hat{\theta}}\Big)^{-1}h(\hat{\theta})' \longrightarrow \chi^2(G).$$

Furthermore, from the fact that $I(\hat{\theta}) \longrightarrow I(\theta)$ as $n \longrightarrow \infty$ (i.e., convergence in probability, 確率収束), we can replace θ by $\hat{\theta}$ as follows:

$$h(\hat{\theta})\Big(R_{\hat{\theta}}(I(\hat{\theta}))^{-1}R'_{\hat{\theta}}\Big)^{-1}h(\hat{\theta})' \longrightarrow \chi^2(G).$$

2. **Lagrange Multiplier Test (ラグランジェ乗数検定):** $\quad LM = F'_{\tilde{\theta}}(I(\tilde{\theta}))^{-1}F_{\tilde{\theta}}$

   (a) MLE with the constraint $h(\theta) = 0$:

$$\max_{\theta}\ \log L(\theta), \quad \text{subject to} \quad h(\theta) = 0$$

The Lagrangian function:

$$L = \log L(\theta) + \lambda h(\theta)$$

(b) For maximization, we have the following two equations:

$$\frac{\partial L}{\partial\theta} = \frac{\partial \log L(\theta)}{\partial\theta} + \lambda\frac{\partial h(\theta)}{\partial\theta} = 0$$

$$\frac{\partial L}{\partial\lambda} = h(\theta) = 0$$

(c) Mean and variance of $\dfrac{\partial \log L(\theta)}{\partial\theta}$ are given by:

$$\mathrm{E}\Big(\frac{\partial \log L(\theta)}{\partial\theta}\Big) = 0, \qquad \mathrm{V}\Big(\frac{\partial \log L(\theta)}{\partial\theta}\Big) = -\mathrm{E}\Big(\frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta'}\Big) = I(\theta).$$

(d) Therefore, using the central limit theorem,

$$\frac{1}{\sqrt{n}}\frac{\partial \log L(\theta)}{\partial \theta} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial \theta} \longrightarrow N\Big(0, \lim_{n\to\infty}\Big(\frac{1}{n}I(\theta)\Big)\Big)$$

(e) Therefore,

$$\frac{\partial \log L(\theta)}{\partial \theta}(I(\theta))^{-1}\frac{\partial \log L(\theta)}{\partial \theta'} \longrightarrow \chi^2(G)$$

Because MLE is consistent, i.e., $\tilde{\theta} \longrightarrow \theta$, we have the result:

$$F'_{\tilde{\theta}}(I(\tilde{\theta}))^{-1}F_{\tilde{\theta}} \longrightarrow \chi^2(G).$$

3. **Likelihood Ratio Test (尤度比検定):** $LR = -2\log\lambda \longrightarrow \chi^2(G)$

$$\lambda = \frac{L(\tilde{\theta})}{L(\hat{\theta})}$$

(a) By Taylor series expansion evaluated at $\theta = \hat{\theta}$, $\log L(\theta)$ is given by:

$$\begin{aligned}\log L(\theta) &= \log L(\hat{\theta}) + \frac{\partial \log L(\hat{\theta})}{\partial \theta}(\theta - \hat{\theta}) \\ &\quad + \frac{1}{2}(\theta-\hat{\theta})'\frac{\partial^2 \log L(\hat{\theta})}{\partial\theta\partial\theta'}(\theta-\hat{\theta}) + \cdots \\ &= \log L(\hat{\theta}) + \frac{1}{2}(\theta-\hat{\theta})'\frac{\partial^2 \log L(\hat{\theta})}{\partial\theta\partial\theta'}(\theta-\hat{\theta}) + \cdots\end{aligned}$$

Note that $\dfrac{\partial \log L(\hat{\theta})}{\partial \theta} = 0$ because $\hat{\theta}$ is MLE.

$$\begin{aligned}-2(\log L(\theta) - \log L(\hat{\theta})) &\approx -(\theta-\hat{\theta})'\Big(\frac{\partial^2 \log L(\hat{\theta})}{\partial\theta\partial\theta'}\Big)(\theta-\hat{\theta}) \\ &= \sqrt{n}(\hat{\theta}-\theta)'\Big(-\frac{1}{n}\frac{\partial^2 \log L(\hat{\theta})}{\partial\theta\partial\theta'}\Big)\sqrt{n}(\hat{\theta}-\theta) \\ &\longrightarrow \chi^2(G)\end{aligned}$$

Note:

(1) $\hat{\theta} \longrightarrow \theta$,

(2) $-\dfrac{1}{n}\dfrac{\partial^2 \log L(\hat{\theta})}{\partial\theta\partial\theta'} \longrightarrow -\lim_{n\to\infty}\Big(\dfrac{1}{n}\mathrm{E}\Big(\dfrac{\partial^2 \log L(\hat{\theta})}{\partial\theta\partial\theta'}\Big)\Big) = \lim_{n\to\infty}\Big(\dfrac{1}{n}I(\theta)\Big)$,

(3) $\sqrt{n}(\hat{\theta}-\theta) \longrightarrow N\Big(0, \lim_{n\to\infty}\Big(\dfrac{1}{n}I(\theta)\Big)\Big)$.

(b) Under $H_0:\ h(\theta) = 0$,

$$-2(\log L(\tilde{\theta}) - \log L(\hat{\theta})) \longrightarrow \chi^2(G).$$

Remember that $h(\tilde{\theta}) = 0$ is always satisfied.

For proof, see Theil (1971, p.396).

4. All of W, LM and LR are asymptotically distributed as $\chi^2(G)$ random variables under the null hypothesis $H_0 : h(\theta) = 0$.

5. Under some comditions, we have $W \geq LR \geq LM$. See Engle (1981) "Wald, Likelihood and Lagrange Multiplier Tests in Econometrics," Chap. 13 in *Handbook of Econometrics*, Vol.2, Grilliches and Intriligator eds, North-Holland.