Advanced Econometrics

Chapter 6: Asymptotic inference

In Choi

Sogang University

Maximum likelihood estimation

- Suppose that $z_{1,}z_2, ..., z_n \overset{iid}{\sim} f(z_n|\gamma)$. Then, the log-likelihood is

$$l(\gamma) = \sum_{i=1}^{n} \ln f(z_i|\gamma)$$

Let $\hat{\gamma}$ be the MLE (maximum likelihood estimator) resulting from the maximization of the log-likelihood. Under some regularity conditions,

1. $\hat{\gamma} \xrightarrow{p} \gamma$.
2. $\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, \lim(\frac{\mathbf{I}(\gamma)}{n})^{-1})$ where $\mathbf{I}(\gamma) = E\left[\frac{\partial l}{\partial \gamma}\frac{\partial l}{\partial \gamma'}\right]$ (information matrix; equal to $-E(\frac{\partial^2 l}{\partial \gamma \partial \gamma'})$ under some regularity conditions[1]).
3. $\hat{\gamma}$ is asymptotically efficient relative to all other consistent asymptotically normal estimators.

[1] The conditions are: (i) $\ln(f(z,\gamma))$ is twice differentiable with respect to γ and (ii) $\int \frac{\partial^2}{\partial \gamma^2} f(z,\gamma)dz = 0$.

Wald test

- $H_0 : R\gamma = r$
  Under H_0,

$$\sqrt{n}(R\hat{\gamma} - r) = \sqrt{n}R(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, R\lim(\frac{\mathbf{I}(\gamma)}{n})^{-1}R').$$

  Define Wald test statistic by

$$W = (R\hat{\gamma} - r)' \left[R\mathbf{I}(\gamma)^{-1}R'\right]^{-1}(R\hat{\gamma} - r).$$

  Then, under H_0,

$$W \xrightarrow{d} \chi^2(rank(R))$$

  Under H_A,

$$W \xrightarrow{p} \infty.$$

  Hence, if $W > C_\alpha$, reject the null hypothesis.

Wald test

Applications

$y = X\beta + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2 \mathcal{I}).$

1. $H_0 : R\beta = r$

   Since

$$\mathbf{I}(\theta)^{-1} = \begin{bmatrix} \sigma^2 (X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix},$$

$$W = (R\hat{\beta} - r)' \left[\hat{\sigma}^2 R (X'X)^{-1} R'\right]^{-1} (R\hat{\beta} - r),$$

   where

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2 \quad (\hat{\sigma}^2 = \frac{\hat{e}'\hat{e}}{n}, \frac{\hat{e}'\hat{e}}{n-K}).$$

   Under H_0,

$$W \xrightarrow{d} \chi^2(rank \ of \ R).$$

   Since $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 \lim_{n\to\infty}(\frac{X'X}{n})^{-1})$ under the i.i.d. assumption on $\{\varepsilon_i\}$, we do not need a normality assumption.

2. $H_0 : f(\beta) = r$, where f is continuous; there exists $\frac{\partial f}{\partial \beta}$; and $\frac{\partial f}{\partial \beta}$ is continuous.

- Perform the Taylor expansion

$$f(\hat{\beta}) = f(\beta) + \frac{\partial f}{\partial \beta'}(\hat{\beta} - \beta) + O_p(n^{-1})$$

or equivalently

$$\sqrt{n}\left(f(\hat{\beta}) - f(\beta)\right) = \frac{\partial f}{\partial \beta'}\sqrt{n}(\hat{\beta} - \beta) + O_p(n^{-\frac{1}{2}}).$$

Under the null,

$$\sqrt{n}\left(f(\hat{\beta}) - r\right) \xrightarrow{d} N\left(0, \frac{\partial f}{\partial \beta'}\sigma^2 \lim\left(\frac{X'X}{n}\right)^{-1}\frac{\partial f}{\partial \beta}\right).$$

Wald test
Applications

- Since we don't know $\frac{\partial f}{\partial \beta}$, we replace it with $\frac{\partial f}{\partial \beta}|_{\beta=\hat{\beta}}$. Since $\hat{\beta} \xrightarrow{p} \beta$ and $\frac{\partial f}{\partial \beta}$ is continuous,

$$\frac{\partial f}{\partial \beta}|_{\beta=\hat{\beta}} \xrightarrow{p} \frac{\partial f}{\partial \beta}.$$

Hence, the Wald test statistic is

$$W = (f(\hat{\beta}) - r)' \left[\frac{\partial f}{\partial \beta'}|_{\hat{\beta}} (X'X)^{-1} \frac{\partial f}{\partial \beta}|_{\hat{\beta}} \right]^{-1} (f(\hat{\beta}) - r)/\hat{\sigma}^2.$$

If $H_0 : \beta_1^2 + \beta_2^2 = 1$, we have $f(\hat{\beta}) = \hat{\beta}_1^2 + \hat{\beta}_2^2$, $\frac{\partial f}{\partial \beta}|_{\beta=\hat{\beta}} = \begin{bmatrix} 2\hat{\beta}_1 \\ 2\hat{\beta}_2 \end{bmatrix}$ and $\hat{\sigma}^2 = \frac{(y-X\hat{\beta})'(y-X\hat{\beta})}{n-K}$.

Wald test

Remarks

1. In order to use Wald tests, we need to estimate the model without the null constraints.
2. Wald tests are not invariant to alternative specifications. For example, consider two equivalent null hypotheses

$$
\begin{aligned}
H_0 &: \beta_1 \beta_2 = 1; \\
H_0' &: \beta_1 - \frac{1}{\beta_2} = 0.
\end{aligned}
$$

For H_0, $f(\hat{\beta}) = \hat{\beta}_1 \hat{\beta}_2$, $r = 1$, $\frac{\partial f}{\partial \beta}|_{\beta=\hat{\beta}} = \begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_1 \end{bmatrix}$.

For H_0', $f(\hat{\beta}) = \hat{\beta}_1 - \frac{1}{\hat{\beta}_2}$, $r = 0$, $\frac{\partial f}{\partial \beta}|_{\beta=\hat{\beta}} = \begin{bmatrix} 1 \\ \frac{1}{\hat{\beta}_2^2} \end{bmatrix}$.

The Wald tests corresponding to H_0 and H_0' are different from each other. (See Gregory and Veal (1985), "On Formulating Wald Tests of Nonlinear Restrictions", Econometrica).

LR test

- The likelihood ratio (LR) test statistic is defined by $LR = 2$[unrestricted log-likelihood $-$ restricted log-likelihood] $= 2\left[l(\hat{\gamma}) - l(\gamma^*)\right]$, say.
- If the null is true, $\hat{\gamma} \simeq \gamma^*$ and $l(\hat{\gamma}) - l(\gamma^*)$ will take a small value. If not, $l(\hat{\gamma}) - l(\gamma^*)$ will take a large value. Note that we always have $l(\hat{\gamma}) - l(\gamma^*) \geq 0$.
- Note that under some regularity conditions

$$n^{-\frac{1}{2}} \frac{\partial l(\gamma)}{\partial \gamma} \xrightarrow{d} N\left(0, \lim n^{-1}\mathbf{I}(\gamma)\right)$$

where $\frac{\partial l(\gamma)}{\partial \gamma} = d(\gamma)$ is called the score.

LR test

- The restricted MLE γ^* is obtained by solving the constrained optimization problem:

$$\max_{\gamma} l(\gamma)$$

$$s.t.\ R\gamma = r.$$

Introduce the Lagrangian

$$\Lambda(\gamma, \lambda) = l(\gamma) - \lambda'(R\gamma - r).$$

LR test

- The first order conditions are

$$\begin{aligned} d(\gamma^*) - R'\lambda^* &= 0; \\ R\gamma^* - r &= 0. \end{aligned} \tag{1}$$

Under H_0,

$$\begin{aligned} & d(\gamma) + \frac{\partial d}{\partial \gamma'}(\gamma^* - \gamma) - R'\lambda^* \\ = \;& d(\gamma) + D(\gamma)(\gamma^* - \gamma) - R'\lambda^* \\ \simeq \;& 0 \qquad \left(D(\gamma) = \frac{\partial d}{\partial \gamma'} = \frac{\partial^2 l(\gamma)}{\partial \gamma \partial \gamma'} \right) \end{aligned} \tag{2}$$

and

$$R(\gamma^* - \gamma) = 0. \tag{3}$$

- Multiply $\mathbf{I}(\gamma)^{-\frac{1}{2}}$ on both sides of (2). Then,

$$\mathbf{I}^{-\frac{1}{2}}(\gamma)d(\gamma) + \mathbf{I}^{-\frac{1}{2}}(\gamma)D(\gamma)(\gamma^* - \gamma) - \mathbf{I}^{-\frac{1}{2}}(\gamma)R'\lambda^* \simeq 0.$$

Because

$$\begin{aligned} & \mathbf{I}^{-\frac{1}{2}}(\gamma)D(\gamma)(\gamma^* - \gamma) \\ = \; & \mathbf{I}^{-\frac{1}{2}}(\gamma)D(\gamma)\mathbf{I}^{-\frac{1}{2}}(\gamma)\mathbf{I}^{\frac{1}{2}}(\gamma)(\gamma^* - \gamma) \end{aligned}$$

and

$$\mathbf{I}^{-\frac{1}{2}}(\gamma)D(\gamma)\mathbf{I}^{-\frac{1}{2}}(\gamma) \xrightarrow{p} -\mathcal{I} \text{ (Identity matrix)},$$

it follows that

$$\mathbf{I}^{-\frac{1}{2}}(\gamma)d(\gamma) - \mathbf{I}^{\frac{1}{2}}(\gamma)(\gamma^* - \gamma) - \mathbf{I}^{-\frac{1}{2}}(\gamma)R'\lambda^* \simeq 0. \qquad (4)$$

Now, define

$$A(\gamma) = \left[R\mathbf{I}(\gamma)^{-1}R'\right]^{-\frac{1}{2}} R\mathbf{I}(\gamma)^{-\frac{1}{2}}.$$

LR test

- We obtain by premultiplying (4) by $A(\gamma)$

$$
\begin{aligned}
& A(\gamma)\mathbf{I}^{-\frac{1}{2}}(\gamma)d(\gamma) - A(\gamma)\mathbf{I}^{\frac{1}{2}}(\gamma)(\gamma^* - \gamma) - A(\gamma)\mathbf{I}^{-\frac{1}{2}}(\gamma)R'\lambda^* \\
= \;& A(\gamma)\mathbf{I}^{-\frac{1}{2}}(\gamma)d(\gamma) - A(\gamma)\mathbf{I}^{-\frac{1}{2}}(\gamma)R'\lambda^* \\
= \;& A(\gamma)\mathbf{I}^{-\frac{1}{2}}(\gamma)d(\gamma) - \left[R\mathbf{I}(\gamma)^{-1}R'\right]^{-\frac{1}{2}} R\mathbf{I}(\gamma)^{-1}R'\lambda^* \\
= \;& A(\gamma)\mathbf{I}^{-\frac{1}{2}}(\gamma)d(\gamma) - \left[R\mathbf{I}(\gamma)^{-1}R'\right]^{\frac{1}{2}} \lambda^* \simeq 0.
\end{aligned}
$$

The first equality holds since

$$
\begin{aligned}
& A(\gamma)\mathbf{I}^{\frac{1}{2}}(\gamma)(\gamma^* - \gamma) \\
= \;& \left[R\mathbf{I}(\gamma)^{-1}R'\right]^{-\frac{1}{2}} R\mathbf{I}(\gamma)^{-\frac{1}{2}}\mathbf{I}(\gamma)^{\frac{1}{2}}(\gamma^* - \gamma) \\
= \;& \left[R\mathbf{I}(\gamma)^{-1}R'\right]^{-\frac{1}{2}} R(\gamma^* - \gamma) = 0
\end{aligned}
$$

by (3).

LR test

- Thus we have

$$A(\gamma)\mathbf{I}^{-\frac{1}{2}}(\gamma)d(\gamma) \simeq \left[R\mathbf{I}(\gamma)^{-1}R'\right]^{\frac{1}{2}}\lambda^*.$$

Since $A(\gamma)A(\gamma)' = \mathcal{I}$,

$$A(\gamma)\mathbf{I}^{-\frac{1}{2}}(\gamma)d(\gamma) \xrightarrow{d} N(0, \mathcal{I}),$$

which implies

$$\left[R\mathbf{I}(\gamma)^{-1}R'\right]^{\frac{1}{2}}\lambda^* \xrightarrow{d} N(0, \mathcal{I})$$

or equivalently

$$n^{-\frac{1}{2}}\lambda^* \xrightarrow{d} N\left(0, \lim\left(R\left(\frac{\mathbf{I}(\gamma)}{n}\right)^{-1}R'\right)^{-1}\right) \quad (5)$$

LR test

- By the Taylor expansion,

$$\begin{aligned} d(\gamma^*) &\simeq d(\hat{\gamma}) + D(\hat{\gamma})(\gamma^* - \hat{\gamma}) \\ &= D(\hat{\gamma})(\gamma^* - \hat{\gamma}), \end{aligned}$$

which along with the first order condition (1) gives

$$D(\hat{\gamma})(\gamma^* - \hat{\gamma}) - R'\lambda^* \simeq 0. \tag{6}$$

LR test

- Since

$$l(\gamma^*) \simeq l(\hat{\gamma}) + d(\hat{\gamma})(\gamma^* - \hat{\gamma}) + \frac{1}{2}(\gamma^* - \hat{\gamma})' D(\hat{\gamma})(\gamma^* - \hat{\gamma}),$$

we have (note that $d(\hat{\gamma}) = 0$)

$$\begin{aligned} LR &= 2\left[l(\hat{\gamma}) - l(\gamma^*)\right] \\ &\simeq (\gamma^* - \hat{\gamma})' \left(-D(\hat{\gamma})\right)(\gamma^* - \hat{\gamma}). \end{aligned}$$

By (5) & (6),

$$\begin{aligned} LR &\simeq \lambda^{*\prime} R \left(-D(\hat{\gamma})\right)^{-1} R' \lambda^* \\ &= \lambda^{*\prime} R \mathbf{I}(\hat{\gamma})^{-1} R' \lambda^* \\ &\xrightarrow{d} \chi^2(number\ of\ restrictions). \end{aligned}$$

LR test
Application

$$y = X\beta + \varepsilon; \;\; \varepsilon \sim N(0, \sigma^2 \mathcal{I})$$

- $H_0 : R\beta = r$

$$\begin{aligned} LR &= 2\left[\ln\left(\frac{\hat{\sigma}^2_{MLE}}{\sigma^{*2}_{MLE}}\right)^{-\frac{n}{2}}\right] \\ &= n\left[\ln\left(\frac{\sigma^{*2}_{MLE}}{\hat{\sigma}^2_{MLE}}\right)\right]. \end{aligned} \quad (7)$$

Remark LR tests are computationally more involved, because we need to estimate the model with and without restrictions.

LM test

- **Intuition:** If the null restriction is true, we expect that γ^* would be close to $\hat{\gamma}$ and that $d(\gamma^*) \simeq 0$. Thus, under the null $R'\lambda^* \simeq 0$. Otherwise, $R'\lambda^*$ would take a large value.
- The Lagrange multiplier (LM) test statistic is defined by

$$\begin{aligned} LM &= \lambda^{*\prime} R \mathbf{I}(\gamma^*)^{-1} R' \lambda^* \\ &= \frac{\partial l}{\partial \gamma'}|_{\gamma^*} \mathbf{I}(\gamma^*)^{-1} \frac{\partial l}{\partial \gamma}|_{\gamma^*}. \end{aligned}$$

Using (5), we find under the null

$$LM \xrightarrow{d} \chi^2(rank(R)).$$

$$y = X\beta + \varepsilon; \;\; \varepsilon \sim N(0, \sigma^2 \mathcal{I})$$

- $H_0 : R\beta = r$

$$l(\beta, \sigma^2) = const. - \frac{n}{2}\sigma^2 - \frac{(y - X\beta)'(y - X\beta)}{2\sigma^2}$$

- Note that

$$\begin{aligned}\frac{\partial l}{\partial \beta}\Big|_{\beta^*} &= -\frac{1}{2\sigma^2}\left(-2X'y + 2(X'X)\beta^*\right)\\ &= \frac{1}{\sigma^2}\left(X'y - (X'X)\beta^*\right)\\ &= \frac{1}{\sigma^2}\left(X'y - \left\{X'y + R'\left[R(X'X)^{-1}R'\right]^{-1}(r - R\hat{\beta})\right\}\right)^2\\ &= \frac{1}{\sigma^2}\left(R'\left[R(X'X)^{-1}R'\right]^{-1}(R\hat{\beta} - r)\right) \qquad (8)\end{aligned}$$

and that

$$\mathbf{I}(\gamma^*)^{-1} = \sigma^{*2}(X'X)^{-1}.$$

[2] $\beta^* = \hat{\beta} + R'\left[R(X'X)^{-1}R'\right]^{-1}(r - R\hat{\beta})$

- Replacing σ^2 in (8) with σ^{*2}, we obtain

$$\begin{aligned} LM &= \frac{1}{\sigma^{*4}}(R\hat{\beta}-r)'\left[R(X'X)^{-1}R'\right]^{-1}\left[\left[R(X'X)^{-1}R'\right]^{-1}\sigma^{*2}\right] \\ &\quad \left[R(X'X)^{-1}R'\right](R\hat{\beta}-r) \\ &= \frac{(R\hat{\beta}-r)'\left[R(X'X)^{-1}R'\right]^{-1}(R\hat{\beta}-r)}{\sigma^{*2}} \end{aligned}$$

or

$$LM = \frac{e^{*\prime}e^{*} - e'e}{\frac{e^{*\prime}e^{*}}{n}}. \tag{9}$$

In contrast,

$$W = \frac{e^{*\prime}e^{*} - e'e}{\frac{\hat{e}'\hat{e}}{n}}. \tag{10}$$

Remarks

- To compute LM tests, we need to estimate the model only under the null restrictions. This make it easy to compute LM tests.

Overview

1. $W, \; LR, \; LM \xrightarrow{d} \chi^2$.
2. $LM \leq LR \leq W$ for $R\beta = r$.

Exercises

1. Prove relation (7).
2. Prove relation (9).
3. Prove relation (10).
4. Prove that $LM \leq LR \leq W$ when the null hypothesis is $H_0 : R\beta = r$.[3]

[3] Hint: $\frac{x}{1+x} \leq \ln(1+x) \leq x \ (x \geq 0)$.