TA Session for Econometrics II 2025
1

Jukina HATAKEYAMA

The University of Osaka, Department of Economics

October 9, 2025

Introduction to Large Sample Test

$$y = X\beta + u, \quad u \sim N_{\mathbb{R}^N}(0, \sigma^2 I_N)$$

Objective function:

$$g(\theta|X) = (y - X\beta)^\top (y - X\beta)$$

- In the linear regression model, finite-sample inference cannot rely solely on the Gaussian assumption.
- Instead, we use the t-distribution or perform an F-test.
- But how can we proceed in the case of **non-linear models**?

- For non-linear models, the exact finite-sample distribution of estimators and test statistics is usually unknown.
- Maximum likelihood estimators satisfy

$$\hat{\theta} \xrightarrow[n \to \infty]{d} N_{\mathbb{R}^k}\left(\theta_0,\ I(\theta_0)^{-1}\right),$$

  where $I(\theta_0)$ is the Fisher information matrix.
- This asymptotic normality enables us to construct tests even in non-linear settings.
- Based on this idea, three major tests are developed:
  - Likelihood Ratio Test (LRT)
  - Wald Test
  - Score Test (Lagrange Multiplier Test)

Making doubly sure

- Linear regression:
  - Errors are Gaussian.
  - OLS estimator is a linear transformation of errors.
  - $\Rightarrow$ Exact normal distribution for $\hat{\beta}$ (finite n).
- Non-linear models:
  - Parameters are defined via non-linear transformations of data.
  - Finite-sample distribution of $\hat{\beta}$ cannot be derived.
  - Only asymptotic normality (as $n \to \infty$) is available.
- $\Rightarrow$ We rely on **large sample tests**.

Framework

$$y = X\beta + u, \quad u \sim N_{\mathbb{R}^N}(0, \sigma^2 I_N)$$

Let $f_u(u)$ denote the probability density function of u. Since $u = y - X\beta$ is a transformation of y, the likelihood function can be expressed as:

$$L(\theta \mid y, X) = f_u(y - X\beta) \left|\det(\nabla_y u)\right|,$$

where $\nabla_y u = \frac{\partial u}{\partial y}$ is the Jacobian of the transformation from y to u.
Thus, the likelihood becomes:

$$L(\beta, \sigma^2 \mid y, X) = (2\pi\sigma^2)^{-N/2} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)^\top (y - X\beta)\right].$$

Taking the logarithm, we obtain the log-likelihood function:

$$\ell(\beta, \sigma^2 \mid y, X) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)^\top (y - X\beta)$$

The statistical criterion is:

$$\hat{\theta} = \arg\max_{\theta} M_n(\theta),$$

where $M_n(\theta)$ is:

$$M_n(\theta) = \frac{1}{N}\sum_{i=1}^{N} \ell_i(\theta), \quad \ell_i(\theta) = \log f(y_i \mid X_i, \theta).$$

Let $\beta = (\beta_1^\top, \beta_2^\top)^\top$ and null hyposesis $H_0 : \beta_2 = \beta_{c,2} = 0$.

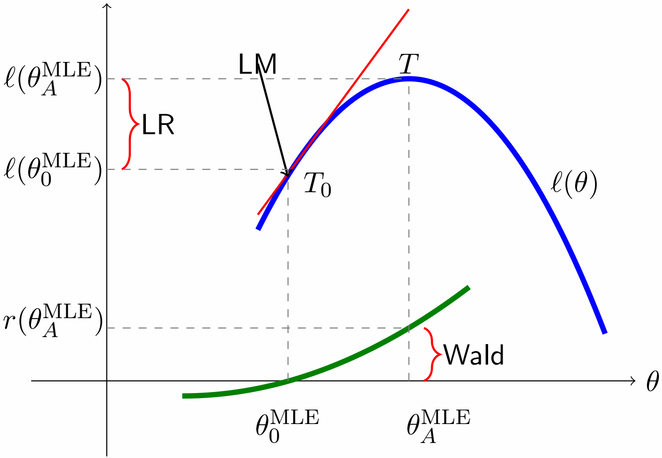
For instance, consider 2 estimators, an unconstrained one:

$$\hat{\theta} = \begin{pmatrix} \hat{\theta_1} \\ \hat{\theta_2} \\ \hat{\sigma^2} \end{pmatrix} = \arg\max_{\theta} M_n(\theta)$$

and a constrained one:

$$\hat{\theta}^0 = \begin{pmatrix} \hat{\beta_1} \\ \hat{\beta_{2,c}} \\ \hat{\sigma^2} \end{pmatrix} = \begin{pmatrix} \hat{\beta_1} \\ \hat{\beta_{2,c}} \\ \hat{\sigma^2} \end{pmatrix} = \arg\max_{\theta} M_n((\beta_1^\top, \beta_{2,c}^\top, \sigma^2))$$

Which estimator provides a better fit under the null?



Restriction function $r(\theta) = 0$. $M_N(\hat{\theta})$ is the log-likelihood evaluated at the MLE of the restricted model (:under H_0) which corresponds to $\ell(\theta_0^{\mathrm{MLE}})$; $M_N(\hat{\theta})$ is the log-likelihood function evaluated at the MLE of the unrestricted model (:under H_1) corresponds to $\ell(\theta_A^{\mathrm{MLE}})$

Comparison of Large Sample Tests

- Consider the null hypothesis in the linear model:

$$H_0 : \beta_2 = \beta_{2,c}.$$

Here $\theta = (\beta, \sigma^2)$, with $\beta = (\beta_1^\top, \beta_2^\top)^\top \in \mathbb{R}^{k_1+k_2}$.

1. **Likelihood Ratio (LR) Test**

$$LR = 2\left[\ell(\hat{\beta}, \hat{\sigma}^2) - \ell(\hat{\beta}^0, \hat{\sigma}^{2,0})\right] \quad \xrightarrow[n\to\infty]{d} \chi^2(k_2),$$

where $(\hat{\beta}, \hat{\sigma}^2)$ is the unrestricted MLE, and $(\hat{\beta}^0, \hat{\sigma}^{2,0})$ the restricted MLE under H_0.

2. **Wald Test**

$$W = (\hat{\beta}_2 - \beta_{2,c})^\top \left[\widehat{\mathrm{Var}}(\hat{\beta}_2)\right]^{-1} (\hat{\beta}_2 - \beta_{2,c}) \quad \xrightarrow[n\to\infty]{d} \chi^2(k_2).$$

3. **Score (Lagrange Multiplier) Test**

$$S = s_{\beta_2}(\hat{\beta}^0, \hat{\sigma}^2)^\top \left[\widehat{\mathrm{Var}}(s_{\beta_2}(\hat{\beta}^0, \hat{\sigma}^2))\right]^{-1} s_{\beta_2}(\hat{\beta}^0, \hat{\sigma}^2) \quad \xrightarrow[n\to\infty]{d} \chi^2(k_2),$$

where $s_{\beta_2}(\beta, \sigma^2) = \partial\ell(\beta, \sigma^2)/\partial\beta_2$ is the score function with respect to β_2.

- **LR Test:**
  - Compares the maximized log-likelihood under unrestricted vs. restricted models
  - Symmetric and invariant to parameterization
- **Wald Test:**
  - Uses the unrestricted estimator $\hat{\theta}$ to test constraints
  - Can be unstable if estimator is near the boundary or sample size is small
- **Score Test:**
  - Uses only the restricted estimator $\hat{\theta}^0$
  - Useful when the unrestricted MLE is difficult to compute
- All three tests are asymptotically equivalent as $n \to \infty$

- **Unconstrained estimates only** Wald test
  - Tests the validity of the null hypothesis using the unconstrained estimator
- **Constrained estimates only** Score (Lagrange Multiplier) test
  - Tests whether relaxing the restriction improves the fit, based on the constrained estimator
- **Both constrained and unconstrained estimates** Likelihood Ratio test
  - Compares the fit of the model with and without the restriction

All the following asymptotic distributions are based on:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow[n\to\infty]{d} N_{\mathbb{R}^N}(0, H(\theta_0)^{-1}J(\theta_0)H(\theta_0)^{-1}).$$

These tests hold for any M-estimator.

1. Introduction to Large Sample Test

2. Framework

3. 3 tests
   - Wald test
   - Score test
   - Likelihood Ration test
   - Test statistics
   - Just in case... LR

4. Example

5. Simulation Experiment

Wald test

$H_0 : r(\theta_0) = 0$ with $r : \mathbb{R}^d \to \mathbb{R}^L$ where $L \leq d$.
$\Rightarrow$ we want to check whether $r(\hat{\theta})$ is close to zero or not.
$\hat{\theta} = \arg\max_{\theta} M_n(\theta) = \arg\max_{\theta} \frac{1}{n} \sum_{i=1}^{n} \ell(X_i, \theta)$

Theorem

Under H_0, the statistic of the Wald test, $\zeta_n^W = nr(\hat{\theta})\Sigma_W^{-1}(\hat{\theta})r(\hat{\theta})$, is asymptotically distributed as $\xi^s(L)$ with:

$$\Sigma_W(\hat{\theta}) = (\nabla_\theta r(\hat{\theta}))^\top \hat{H}^{-1}(\hat{\theta})\hat{J}(\hat{\theta})\hat{H}^{-1}(\hat{\theta})(\nabla_\theta r(\hat{\theta})),$$

$$\Sigma_W(\theta_0) = (\nabla_\theta r(\theta_0))^\top \hat{H}^{-1}(\theta_0)\hat{J}(\theta_0)\hat{H}^{-1}(\theta_0)(\nabla_\theta r(\theta_0)),$$

$H(\theta) = \mathbb{E}[\nabla^2_{\theta\theta^\top}\ell(X_i, \theta)]$, and $J(\theta) = \mathbb{E}[\nabla_\theta \ell(X_i, \theta)\nabla_{\theta^\top}\ell(X_i, \theta)]$ with $\hat{H}(\hat{\theta})$ and $\hat{J}(\hat{\theta})$ aonsistent estimators of $H(\theta)$ and $J(\theta)$. The test with rejection area $RA = \{\zeta_n^W : \zeta_n^W \geq q_{1-\alpha}(\xi^2(L))\}$ has an asymptotic level α.

$\nabla_\theta r(\theta))^\top$ is the Jacobian matrix of the size $L \times d$.

Score test (Lagrange Multiplier test)

$\widehat{\theta}^0 = \arg\max_\theta M_n(\theta)$ subject to the constraints.

$\Rightarrow$ we want to check whether $\nabla_\theta M_n(\widehat{\theta}^0)$ is close to zero or not. Assume $J(\theta) = H(\theta)$.

Theorem

Under H_0, the statistic of score test, $\zeta_n^S = n(\nabla_\theta M_n(\widehat{\theta}^0))^\top \hat{H}^{-1}(\widehat{\theta}^0)\nabla_\theta M_n(\widehat{\theta}^0)$, satisfies below:

$$\zeta_n^S = \zeta_n^W + o_p(1).$$

The Score test does not involve the unconstrained estimator.

- Define the Hessian-based and outer-product versions:

$$H(\theta) = -\mathbb{E}\big[\nabla^2_{\theta\theta^\top}\ell(X_i,\theta)\big], \qquad J(\theta) = \mathbb{E}\big[\nabla_\theta \ell(X_i,\theta)\,\nabla_\theta \ell(X_i,\theta)^\top\big].$$

- **Information Matrix Equality**: Under correct model specification,

$$J(\theta) = H(\theta).$$

- Consequences:
  1. The Score test statistic simplifies to

$$\zeta_n^S = n\, s(\hat\theta^0)^\top \big[\hat H(\hat\theta^0)\big]^{-1} s(\hat\theta^0).$$

  2. This ensures the asymptotic equivalence of the Wald, Score, and LR tests.
- If the model is misspecified: $J(\theta) \neq H(\theta)$, and robust (sandwich) variance estimation is required.

Likelihood Ration test

$\Rightarrow$ We want to test whether $M_n(\hat{\theta}) - M_n(\widehat{\theta^0})$ is close to zero or not. Assume $J(\theta) = H(\theta)$.

Theorem

Under H_0, the test statistic, $\zeta_n^R = 2n(M_n(\hat{\theta} - M_n(\widehat{\theta^0})))$, satisfies:

$$\zeta_n^R = \zeta_n^W + o_p(1) = \zeta_n^S + o_p(1)$$

Test statistics

Difine:

- $M_n(\hat{\theta})$: the value of the maximized UNRESTRICTED objective function.
- $M_n(\widehat{\theta^0})$: the value of the maximized RESTRICTED objective function.
- $H(\theta)$: Hessian matrix, $H(\theta) = -\mathbb{E}\left[\nabla^2_{\theta\theta^\top}\ell(X_i,\theta)\right]$.
- $J(\theta)$: outer-product of the score, $J(\theta) = \mathbb{E}[\nabla_\theta \ell(X_i,\theta)\nabla_\theta \ell(X_i,\theta)^\top]$.
- $r(\cdot)$: restriction function, $r(\theta) = 0$ under H_0.
- n: number of observations.

- Likelihood Ratio:

$$\zeta_n^R = 2n\big(M_n(\hat{\theta}) - M_n(\widehat{\theta^0})\big) \xrightarrow[n\to\infty]{d} \chi^2(L)$$

- Score:

$$\zeta_n^S = n\, s(\widehat{\theta^0})^\top \hat{H}^{-1}(\widehat{\theta^0})\, s(\widehat{\theta^0}) \xrightarrow[n\to\infty]{d} \chi^2(L),$$

  where $s(\theta) = \nabla_\theta M_n(\theta)$ is the score vector.
- Wald:

$$\zeta_n^W = n\, r(\hat{\theta})^\top \Sigma_W^{-1}(\hat{\theta}) r(\hat{\theta}) \xrightarrow[n\to\infty]{d} \chi^2(L),$$

$$\Sigma_W(\hat{\theta}) = (\nabla_\theta r(\hat{\theta}))^\top \hat{H}^{-1}(\hat{\theta})\hat{J}(\hat{\theta})\hat{H}^{-1}(\hat{\theta})(\nabla_\theta r(\hat{\theta})).$$

Why the Likelihood Ratio Test is χ^2-distributed

- Consider the Likelihood Ratio (LR) statistic:

$$\zeta_n^R = 2\left[\ell(\hat{\theta}) - \ell(\hat{\theta}^0)\right],$$

where $\hat{\theta}$ is the unrestricted MLE and $\hat{\theta}^0$ the restricted MLE.

- **Taylor expansion around the restricted MLE:**

$$\ell(\hat{\theta}) \approx \ell(\hat{\theta}^0) + (\hat{\theta} - \hat{\theta}^0)^\top s(\hat{\theta}^0) + \frac{1}{2}(\hat{\theta} - \hat{\theta}^0)^\top H(\hat{\theta}^0)(\hat{\theta} - \hat{\theta}^0),$$

where $s(\hat{\theta}^0) = \nabla_\theta \ell(\hat{\theta}^0)$ and $H(\hat{\theta}^0)$ is the Hessian.

- At the restricted MLE, $s(\hat{\theta}^0) = 0$, so

$$\ell(\hat{\theta}) - \ell(\hat{\theta}^0) \approx \frac{1}{2}(\hat{\theta} - \hat{\theta}^0)^\top H(\hat{\theta}^0)(\hat{\theta} - \hat{\theta}^0).$$

- Multiplying by 2 cancels the $\frac{1}{2}$ in the quadratic form, yielding

$$\zeta_n^R \approx (\hat{\theta} - \hat{\theta}^0)^\top H(\hat{\theta}^0)(\hat{\theta} - \hat{\theta}^0),$$

which is asymptotically a sum of squares of normal variables $\Rightarrow \chi^2$.

Wald Test Example 1

Let $\theta \in \mathbb{R}^d$ and consider the null hypothesis:

$$H_0 : \theta_j = 0 \quad \text{(the } j\text{th element of the parameter vector is 0).}$$

The restriction function is

$$r : \mathbb{R}^d \to \mathbb{R}, \quad r(\theta) = \theta_j.$$

For the Wald test, the Jacobian is

$$(\nabla_\theta r(\theta))^\top = (0, \ldots, 0, 1, 0, \ldots, 0),$$

with 1 in the jth position.
The matrix $\Sigma_W(\theta)$ reduces to the scalar $\sigma_j^2(\theta)$, which is the jth diagonal element of

$$\Sigma(\theta) = H^{-1}(\theta) J(\theta) H^{-1}(\theta).$$

Hence, the Wald statistic is

$$\zeta_n^W = n \frac{\widehat{\theta}_j^2}{\sigma_j^2(\theta)},$$

Wald Test Example 2

$$\ell(y_i, X_i, \theta) = (Y_i - \alpha - \mu \exp(X_i\beta))^2.$$

$\beta \in \mathbb{R}^{d-2}, \theta = (\alpha, \mu, \beta^\top)^\top \in \mathbb{R}^d.$
$H_0 : \beta_{0,1} = \beta_{0,2}^3,\ \beta_{0,4} = 5\beta_{0,5}^2.$

The restriction function is $r : \mathbb{R}^d \to \mathbb{R}^2$ with:

$$r(\theta_0) = \begin{pmatrix} r_1(\theta_0) \\ r_2(\theta_0) \end{pmatrix} = \begin{pmatrix} \beta_{0,1} - \beta_{0,2}^3 \\ \beta_{0,4} - 5\beta_{0,5}^2 \end{pmatrix}$$

The Jacobian becomes:

$$(\nabla_t hetar(theta))^\top = \begin{pmatrix} 0 & 0 & 1 & -3\beta_2^2 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -10\beta_5 & 0 & \cdots & 0 \end{pmatrix}$$

$$\zeta_n^W = n r(\hat{\theta})^\top \Sigma_W^{-1}(\hat{\theta}) r(\hat{\theta})$$

where $\Sigma_W(\hat{\theta}) = (\nabla_\theta r(\hat{\theta}))^\top \hat{H}^{-1}(\theta)\hat{J}(\theta)\hat{H}^{-1}(\theta)(\nabla_\theta r(\hat{\theta}))$.

Simulation Experiment

Data generating process:

$$X_t = \omega + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t, \qquad t = 3, \ldots, T,$$

where $u_t \sim N(0, \sigma_u^2)$.
Set the true parameters as:

$$\omega = 0.001, \quad \beta_1 = 0.7, \quad \beta_2 = 0.2, \quad \sigma_u^2 = 0.05.$$

The initial values are drawn as $X_1, X_2 \sim N(0, \sigma_u^2)$, and we set $T = 10000$.
The (negative log-likelihood) loss function is:

$$M_n(\theta) = \frac{1}{2n} \sum_{t=1}^{T} \left\{ \log(2\pi\sigma_u^2) + \frac{\left(X_t - \omega - \beta_1 X_{t-1} - \beta_2 X_{t-2}\right)^2}{\sigma_u^2} \right\}.$$

$$\theta = (\omega, \beta_1, \beta_2, \sigma_u^2)$$

$$H_0 : \beta_2 = 0$$

Simulation Experiment: R Code and Results

Output

```
=== 自作計算 ===
LR test statistic:    461.5066
Wald test statistic:  472.1822
Score test statistic: 451.017

=== パッケージ版 ===
Likelihood ratio test
Model 1: Y ~ X_lag1
Model 2: Y ~ X_lag1 + X_lag2
  #Df LogLik Df  Chisq Pr(>Chisq)
1   3 573.23
2   4 803.98  1 461.51  < 2.2e-16 ***
```

Output

```
Wald test
Model 1: Y ~ X_lag1
Model 2: Y ~ X_lag1 + X_lag2
  Res.Df Df  Chisq Pr(>Chisq)
1   9996
2   9995  1 472.18  < 2.2e-16 ***

Linear hypothesis test:
X_lag2 = 0
Model 1: restricted model
Model 2: Y ~ X_lag1 + X_lag2
  Res.Df    RSS Df Sum of Sq  Chisq Pr(>Chisq)
1   9996 521.96
2   9995 498.42  1    23.546 472.18  < 2.2e-16 ***
```

$T-2$ がデータの数なのに留意。