EC708 Discussion 10
Trinity of Tests

Yan Liu

Department of Economics
Boston University

April 16, 2021

Outline

Review of Trinity of Tests

Suppose a model has log-likelihood function $L(\theta)$ where $\dim(\theta) = k$. We are interested in the hypothesis $h(\theta) = 0$ where $h : \mathbb{R}^k \to \mathbb{R}^q$ is a differentiable vector function with $q < k$.

- $\hat{\theta}$: unrestricted MLE
- $\tilde{\theta}$: restricted MLE

Consider three ways of testing $H_0 : h(\theta) = 0$

1. Wald Test: use $\hat{\theta}$
   - Won't reject if $h(\hat{\theta})$ is close to 0
2. Likelihood Ratio Test: use $\hat{\theta}$ and $\tilde{\theta}$
   - Distance between two maximum likelihood functions
3. Lagrange Multiplier Test: use $\tilde{\theta}$
   - If null is true, constraints not binding and Lagrange multiplier λ is zero. Test how much λ is from zero.

Trinity Test Statistics

- **Wald Test**:

$$W = \sqrt{T}h(\hat{\theta})' \left(\frac{\partial h}{\partial \theta'}\Big|_{\theta=\hat{\theta}} \hat{I}^{-1}(\hat{\theta}) \frac{\partial h}{\partial \theta}\Big|_{\theta=\hat{\theta}} \right)^{-1} \sqrt{T}h(\hat{\theta})$$

- **Likelihood Ratio Test**:

$$LR = 2(L(\hat{\theta}) - L(\tilde{\theta}))$$

- **Lagrange Multiplier Test/Score Test**:

$$LM = \frac{1}{\sqrt{T}} \frac{\partial L}{\partial \theta'}\Big|_{\theta=\tilde{\theta}} \hat{I}^{-1}(\tilde{\theta}) \frac{1}{\sqrt{T}} \frac{\partial L}{\partial \theta}\Big|_{\theta=\tilde{\theta}}$$

  - An alternative test statistic is $LM = \frac{\tilde{\lambda}'}{\sqrt{T}} \frac{\partial h}{\partial \theta'}\Big|_{\theta=\tilde{\theta}} \hat{I}^{-1}(\tilde{\theta}) \frac{\partial h}{\partial \theta}\Big|_{\theta=\tilde{\theta}} \frac{\tilde{\lambda}}{\sqrt{T}}$ where $\tilde{\lambda}$ is Lagrange multiplier for $\max_\theta L(\theta) - \lambda' h(\theta)$.

In practice, there are multiple ways to estimate information matrix $\hat{I}(\theta)$.

The Alternative is Important for Power

- Two concepts in hypothesis testing: **size** and power
  - Size: given H_0 is true, probability that the test rejects H_0
  - Power: given H_1 is true, probability that the test rejects H_0
- LM test has good size because the model is estimated under H_0.
- The alternative is $h(\theta) \neq 0$, rather broad
  - Power depends on direction and magnitude of deviation from the null, e.g. local power analysis using Pitman drift.
  - Wald test has good power for specific alternatives[1].
- Among test with well-controlled size, the optimal test should have the highest power, which depends on the alernatives.

[1] However, one drawback of Wald test is that it is not invariant to parametrization of $h(\theta)$ except in linear case.

Classical Linear Model

Consider the linear model

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

where $\dim(\beta) = k$. Recall

$$\sqrt{T}(\hat{\beta}_{OLS} - \beta) \xrightarrow{d} N(0, \sigma^2 Q_{XX}^{-1})$$

where $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$ and $Q_{XX} = E[X_t X_t']$.

- We are interested in testing linear restrictions

  $$h(\beta) = R\beta - r$$

  where R is $q \times k$ and r is $q \times 1$.

- Under $H_0 : R\beta - r = 0$, $\sqrt{T}(R\hat{\beta}_{OLS} - r) \xrightarrow{d} N(0, \sigma^2 R Q_{XX}^{-1} R')$. Recall $\hat{\beta}_{OLS} = \hat{\beta}_{MLE}$ and let $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - x_t'\hat{\beta}_{OLS})^2$. Then,

  $$W = (R\hat{\beta}_{OLS} - r)'[\hat{\sigma}^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta}_{OLS} - r) \xrightarrow{d} \chi_q^2.$$

LR Test in Classical Linear Model

- Under the alternative,

$$L(\hat{\beta}_{OLS}, \hat{\sigma}^2) = \text{cons} - \frac{T}{2}\ln\hat{\sigma}^2 - \frac{1}{2\hat{\sigma}^2}(y - X\hat{\beta}_{OLS})'(y - X\hat{\beta}_{OLS}),$$

where $\hat{\sigma}^2 = \frac{1}{T}(y - X\hat{\beta}_{OLS})'(y - X\hat{\beta}_{OLS}) = \frac{1}{T}SSR$. Therefore,

$$L(\hat{\beta}_{OLS}, \hat{\sigma}^2) = \text{cons} - \frac{T}{2}\ln\hat{\sigma}^2.$$

- OLS estimator under restriction: Derivations

$$\hat{\beta}_c = \hat{\beta}_{OLS} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}_{OLS} - r),$$

$$\hat{\sigma}_c^2 = \frac{1}{T}(y - X\hat{\beta}_c)'(y - X\hat{\beta}_c) = \frac{1}{T}SSR_c.$$

Hence, $L(\hat{\beta}_c, \hat{\sigma}_c^2) = \text{cons} - \frac{T}{2}\ln\hat{\sigma}_c^2$.

- Form the LR statistic:

$$LR = 2(L(\hat{\beta}_{OLS}, \hat{\sigma}^2) - L(\hat{\beta}_c, \hat{\sigma}_c^2)) = T\ln\left(\frac{SSR}{SSR_c}\right) \xrightarrow{d} \chi_q^2.$$

LM Test in Classical Linear Model

- Scale Lagrange multiplier by $2\hat{\sigma}_c^2$:

$$\hat{\lambda} = \frac{1}{\hat{\sigma}_c^2}[R(X'X)^{-1}R']^{-1}(r - R\hat{\beta}_{OLS}).$$

- $\frac{1}{\sqrt{T}}\hat{\lambda} \xrightarrow{d} N(0, (\sigma^2 R Q_{XX}^{-1} R')^{-1})$
- Hence, the LM statistic is

$$LM = (R\hat{\beta}_{OLS} - r)'[\hat{\sigma}_c^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta}_{OLS} - r) \xrightarrow{d} \chi_q^2.$$

Remark:

- Recall Wald test:

$$W = (R\hat{\beta}_{OLS} - r)'[\hat{\sigma}^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta}_{OLS} - r) \xrightarrow{d} \chi_q^2.$$

Differ in terms of estimator of σ^2.

Inequality $LM \leq LR \leq W$ in Classical Linear Model

- For simplicity, normalize $R\beta = 0$.
- We start with two lemmas:
  1. Define $\hat{u}_c = y - X\hat{\beta}_c$ and $\hat{u} = y - X\hat{\beta}_{OLS}$, then
  $$W = \frac{\hat{u}_c'\hat{u}_c - \hat{u}'\hat{u}}{\hat{\sigma}^2}, \quad LM = \frac{\hat{u}_c'\hat{u}_c - \hat{u}'\hat{u}}{\hat{\sigma}_c^2}.$$
  2. The Wald, LR, and LM satisfy the following relations:
     - $W = 2[L(\hat{\beta}_{OLS}, \hat{\sigma}^2) - L(\hat{\beta}_c - \hat{\sigma}^2)]$.
     - $LR = 2[L(\hat{\beta}_{OLS}, \hat{\sigma}^2) - L(\hat{\beta}_c - \hat{\sigma}_c^2)]$.
     - $W = 2[L(\hat{\beta}_{OLS}, \hat{\sigma}_c^2) - L(\hat{\beta}_c - \hat{\sigma}_c^2)]$.
- **Theorem:** $LM \leq LR \leq W$.
- Result applies to the general framework:
$$y|X \sim N(X\beta, \sigma\Omega), \quad \Omega = \Omega(\omega),$$
where ω is a finite estimable parameter vector.

Proof of Lemma 1 Proof of Lemma 2 Proof of Inequality

When Are LM, LR, and Wald the Same?

- Suppose σ^2 is known. Then by Lemma 2, the three are identical.
- More generally, if the log-likelihood has the following form

$$L(\theta) = b - \frac{1}{2}(\theta - \hat{\theta})' A(\theta - \hat{\theta})$$

where A is a symmetric positive definite matrix and $\hat{\theta}$ is a function of data, then $LM = LR = W$.

Sketch of proof:

1. $\partial L/\partial\theta = -(\theta - \hat{\theta})'A$
2. $\partial^2 L/\partial\theta\partial\theta' = -A.$
3. $LM = LR = W = (\theta^0 - \hat{\theta})'A(\theta^0 - \hat{\theta}).$

Remarks:

Log-likelihood function in the neighborhood of θ is approximately quadratic, hence asymptotic equivalence of the three tests.

LM Test as A Diagnostic

- We can use hypothesis testing for specification search: null is a specification in favor and alternative is a more general specification.
- Test for this purpose is **diagnostic**: check if data are well represented by the specification.
- LM test is based on parameter fit under the null, usually expressed as residuals from the estimates under the null.
- Each alternative considered indicates a particular type of non-randomness.

Testing for Heteroskedasticity: Breush-Pagan LM

Consider the following model:

$$y_t = x_t'\beta + u_t, \quad u_t \sim N(0, \sigma_t^2), \quad t = 1, \ldots, T.$$

Null and alternative:

$$H_0 : \sigma_t^2 = \sigma^2 \;\forall t, \quad H_1 : \sigma_t^2 = h(z_t'\alpha),$$

where $z_t = (1, z_{1t}, \ldots, z_{qt})' \in \mathbb{R}^{q+1}$ and $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_q)'$. The null can be rewritten as

$$H_0 : \alpha_1 = \cdots = \alpha_q = 0.$$

- z_t can be a vector function of x_t, e.g. $z_t'\alpha = \alpha_0 + \alpha_1 x_t'\beta$.
- Under both the null and alternative, assume no serial correlation: $E[u_t u_s] = 0$ for $t \neq s$.

Breush-Pagan Test Statistic

- Log-likelihood function

$$L(\beta, \alpha) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log(h(z_t'\alpha)) - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - x_t'\beta)^2}{h(z_t'\alpha)}.$$

- Note that $\frac{\partial L}{\partial \beta}\big|_{\alpha=\tilde{\alpha},\beta=\tilde{\beta}} = 0$ and information matrix is block diagonal. Hence, the LM test statistic is

$$LM = \left(\frac{\partial L}{\partial \alpha}\right)' \left(-\frac{\partial^2 L}{\partial \alpha \partial \alpha'}\right)^{-1} \left(\frac{\partial L}{\partial \alpha}\right)\Big|_{\alpha=\tilde{\alpha},\beta=\tilde{\beta}}.$$

- Evaluated under H_0, it simplifies to ($h'(\alpha_0)$ is cancelled out)

$$LM = \frac{1}{2}f'Z(Z'Z)^{-1}Z'f,$$

where $f = (f_1, \ldots, f_T)'$ with $f_t = \frac{\tilde{u}_t^2}{\tilde{\sigma}^2} - 1 \equiv g_t - 1$. $\tilde{u}_t$ and $\tilde{\sigma}^2$ are residuals and variance estimates of the restricted MLE $(\tilde{\alpha}, \tilde{\beta})$.

Breush-Pagan Test Statistic: Compact Form

- $LM = \frac{1}{2}(g - 1_T)' P_Z (g - 1_T)$, where $g = (g_1, \ldots, g_T)'$, 1_T is a $T \times 1$ vector of ones, and $P_Z = Z(Z'Z)^{-1}Z'$.
- Since $P_Z 1_T = 1_T$ and $g' 1_T = 1_T' 1_T = T$, simplify as

$$LM = \frac{1}{2}(g' P_Z g - T).$$

- In the regression of g on Z, explained sum of squares (ESS) is

$$ESS = g' P_Z g - T\bar{g}^2.$$

- Since $\bar{g} = 1$,

$$LM = \frac{1}{2}(g' P_Z g - T\bar{g}^2) = \frac{1}{2} ESS.$$

Breush-Pagan Test Statistic: Procedures

1. Apply OLS to $y = X\beta + u$ and obtain residuals $\hat{u}$.
2. Compute $g_t = \frac{\hat{u}_t^2}{\hat{\sigma}^2}$ where $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2$.
3. Run OLS of g on Z (including a constant) and compute the LM statistic

$$LM = \frac{1}{2}(g' P_Z g - T\bar{g}^2) = \frac{1}{2}ESS.$$

Under the null, $LM \xrightarrow{d} \chi_q^2$.

LM Test for $AR(p)$ Errors

Consider the model with k regressors in X and p lags in error:

$$
\begin{aligned}
y &= X\beta + u \\
u_t &= \psi_1 u_{t-1} + \cdots + \psi_p u_{t-p} + \varepsilon_t \\
\varepsilon_t &\sim \text{i.i.d. } N(0, \sigma^2)
\end{aligned}
$$

The null is $H_0 : \psi = (\psi_1, \ldots, \psi_p) = \mathbf{0}$.

- Ignore the first p observations and rewrite the model as

$$
y_t = x_t'\beta + \sum_{j=1}^{p} \psi_j (y_{t-j} - x_{t-j}'\beta) + \varepsilon_t.
$$

- Consider it as a nonlinear model in β and ψ: $y_t = f_t(\beta, \psi) + \varepsilon_t$.

LM Test for $AR(p)$ Errors

- Denote $\theta = (\beta, \psi)$. Take first-order Taylor expansion around $\psi = \mathbf{0}$ and restricted MLE $\tilde{\beta}$:

$$y_t \approx f_t(\tilde{\beta}, \mathbf{0}) + \sum_{j=1}^{k+p} \frac{\partial f_t(\beta, \psi)}{\partial \theta_j}\Big|_{\tilde{\beta}, \mathbf{0}} (\theta_j - \tilde{\theta}_j) + \varepsilon_t.$$

- Since $\frac{\partial f_t}{\partial \beta_j}\big|_{\tilde{\beta}, \mathbf{0}} = x_{tj}, j = 1, \ldots, k$, $\frac{\partial f_t}{\partial \psi_l}\big|_{\tilde{\beta}, \mathbf{0}} = \tilde{u}_{t-l}, l = 1, \ldots, p$,

$$\underbrace{y_t - f_t(\tilde{\beta}, \mathbf{0})}_{\tilde{u}_t} \approx \sum_{j=1}^{k} x_{tj}(\beta_j - \tilde{\beta}_j) + \sum_{l=1}^{p} \tilde{u}_{t-l}\psi_l + \varepsilon_t. \qquad (*)$$

- The matrix form is

$$\tilde{u} = F \begin{pmatrix} \beta - \tilde{\beta} \\ \psi \end{pmatrix} + \varepsilon$$

where $F = (X, U)$ with U having rows $U_t = (\tilde{u}_{t-1}, \ldots, \tilde{u}_{t-p})$.

LM Test for $AR(p)$ Errors

- The log-likelihood after concentrating out σ^2 is

$$L(\tilde{\theta}) = \text{cons} - \frac{T-p}{2} \ln \tilde{\sigma}^2$$

  where $(T-p)\tilde{\sigma}^2$ is SSR under the null, i.e., OLS residuals.

- The score and estimated information matrix are

$$\frac{\partial L(\tilde{\theta})}{\partial \theta} = \frac{F'\tilde{u}}{\tilde{\sigma}^2}, \quad \hat{I}(\tilde{\theta}) = \frac{1}{T-p} \frac{F'F}{\tilde{\sigma}^2}.$$

- Construct the LM statistic

$$LM = \frac{1}{T-p} \frac{\partial L(\tilde{\theta})}{\partial \theta'} \hat{I}(\tilde{\theta})^{-1} \frac{\partial L(\tilde{\theta})}{\partial \theta} = \frac{1}{\tilde{\sigma}^2} \tilde{u}' F (F'F)^{-1} F' \tilde{u}.$$

- $LM = (T-p)\frac{ESS}{TSS} = (T-p)R^2$ where ESS, TSS, and R^2 are calculated from $(*)$.

LM Test for $AR(p)$ Errors: Procudures

1. Run OLS on y against X and get $\hat{u}$.
2. Run OLS on the auxiliary regression

$$\hat{u} = X\tau + U\delta + v.$$

3. Compute R^2 from the auxiliary regression and construct LM statistic

$$LM = (T - p)R^2.$$

Under the null, $LM \xrightarrow{d} \chi^2_p$.

Remarks:

- If X includes no lagged dependent variables, then $\text{plim}_{T\to\infty} \frac{X'U}{T} = 0$ and the auxiliary regression will be unaffected by leaving out the X's.
- For $p = 1$, this test is asymptotically equivalent to the Durbin-Watson statistic.

Restricted OLS Estimator

Constrained optimization problem

$$\min_{\beta,\lambda} S(\beta,\lambda) = (y - X\beta)'(y - X\beta) + \lambda'(R\beta - r).$$

- Two FOCs:

$$\frac{\partial S}{\partial \beta} = -2X'y + 2X'X\hat{\beta}_c + R'\hat{\lambda} = 0, \quad \frac{\partial S}{\partial \lambda} = -r + R\hat{\beta}_c = 0.$$

- Combine the two FOCs and rearrange:

$$\hat{\lambda} = -2[R(X'X)^{-1}R']^{-1}[r - R(X'X)^{-1}X'y].$$

- Plug in $\hat{\lambda}$ to obtain

$$\hat{\beta}_c = \hat{\beta}_{OLS} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}_{OLS} - r).$$

Remarks:

- If constraints are exactly satisfied, $\hat{\beta}_c = \hat{\beta}_{OLS}$.
- $\text{Var}(\hat{\beta}_c|X) \leq \text{Var}(\hat{\beta}_{OLS}|X)$. Back

Proof of Lemma 1

Recall that $\hat{\beta}_c = \hat{\beta}_{OLS} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}_{OLS} - r)$, $X'\hat{u} = 0$, and $r = 0$,

$$\begin{aligned}\hat{u}_c'\hat{u}_c - \hat{u}'\hat{u} &= [\hat{u} - X(\hat{\beta}_c - \hat{\beta}_{OLS})]'[\hat{u} - X(\hat{\beta}_c - \hat{\beta}_{OLS})] - \hat{u}'\hat{u} \\ &= (\hat{\beta}_c - \hat{\beta}_{OLS})'X'X(\hat{\beta}_c - \hat{\beta}_{OLS}) \\ &= \hat{\beta}_{OLS}'R[R(X'X)^{-1}R']^{-1}R\hat{\beta}_{OLS}.\end{aligned}$$

Plug this identity into the expressions of Wald and LM statistics. Back

Proof of Lemma 2

1. For any $\tilde{\sigma}^2$,

$$L(\hat{\beta}_{OLS}, \tilde{\sigma}^2) = -\frac{T}{2}\ln\tilde{\sigma}^2 - \frac{T}{2}\ln 2\pi - \frac{1}{2}\frac{\hat{u}'\hat{u}}{\tilde{\sigma}^2},$$
$$L(\hat{\beta}_c, \tilde{\sigma}^2) = -\frac{T}{2}\ln\tilde{\sigma}^2 - \frac{T}{2}\ln 2\pi - \frac{1}{2}\frac{\hat{u}_c'\hat{u}_c}{\tilde{\sigma}^2}.$$

2. By Lemma 1,

$$W = \frac{\hat{u}_c'\hat{u}_c - \hat{u}'\hat{u}}{\hat{\sigma}^2} = 2[L(\hat{\beta}_{OLS}, \hat{\sigma}^2) - L(\hat{\beta}_c - \hat{\sigma}^2)],$$
$$LM = \frac{\hat{u}_c'\hat{u}_c - \hat{u}'\hat{u}}{\hat{\sigma}_c^2} = 2[L(\hat{\beta}_{OLS}, \hat{\sigma}_c^2) - L(\hat{\beta}_c - \hat{\sigma}_c^2)].$$

Back

Proof of Inequality

- $LR \geq LM$ if

$$L(\hat{\beta}_{OLS}, \hat{\sigma}^2) \geq L(\hat{\beta}_{OLS}, \hat{\sigma}_c^2)$$

This inequality holds since $\hat{\beta}_{OLS}$ and $\hat{\sigma}^2$ maximizes the log-likelihood.

- $W \geq LR$ if

$$L(\hat{\beta}_c, \hat{\sigma}^2) \leq L(\hat{\beta}_c, \hat{\sigma}_c^2)$$

This inequality holds since $\hat{\beta}_c$ and $\hat{\sigma}_c^2$ are the restricted MLE.

Back