TA Session 8: Trinity of Tests

Shuowen Chen[1]

EC708: PhD Econometrics I (Spring 2020)

[1]Parts of the materials are borrowed from Robert Engle (1984) and teaching slides by Pierre Perron.

Outline (Future Nostalgia)

- Review of Trinity of Tests
  - Wald Test (Wald, 1943)
  - Likelihood Ratio Test (Wilks, 1938)
  - Lagrange Multiplier/Rao/Score Test (Rao, 1948; Silvey, 1958, 59)
- The Inequality $LM \leq LR \leq W$ in Classical Linear Model
- LM Test for Model Misspecification
  - Testing for Heteroskedasticity (Breusch–Pagan, 1980)
  - Testing for Serial Correlation
    1. Durbin–Watson (1950, 1951)
    2. Durbin's "h" statistic (1970)
    3. Godfrey (1978)

Trinity of Tests

Denote $y = (y_1, \ldots, y_T)$. Suppose a model has likelihood function $L(y; \theta)$, where $dim(\theta) = k$, and we are interested in the hypothesis $h(\theta) = 0$, where $h(\cdot)$ is a q–vector of differentiable function with $q < k$

- $\widehat{\theta}$: unrestricted MLE
- $\widetilde{\theta}$: restricted MLE

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

1. Wald Test
    - Won't reject if $\widehat{\theta}$ close to $\widetilde{\theta}$
    - Estimate model under H_1 and treat hypothesis as the true value
2. Likelihood Ratio Test
    - Distance between two values of maximum likelihood function
    - Estimate model under both H_0 and H_1
3. Lagrange Multiplier Test
    - If null is true, constraints not binding, Lagrange multiplier (λ) is zero, test how from it is from zero
    - Estimate model under H_0

The Alternative is Important for Power

- Two concepts in hypothesis testing: **size** and **power**
  - Size: given that H_0 is true, probability that the test rejects H_0
  - Power: given that H_1 is true, probability that the test rejects H_0
- For size properties, examine the distribution of the test under H_0
  - Implication? LM test has good size
  - In Monte Carlo, generate data under H_0 and compare empirical critical values with theoretical ones
- The alternative is $h(\theta) \neq 0$, rather broad
  - Depends on direction and magnitude of deviation from the null
    1. Direction: Tests are typically not omnibus
    2. Magnitude: local power analysis
  - Implication? Wald test has good power for specific alternatives[2]
  - In Monte Carlo, generate data under specific H_1
- Takeaway: Among tests that have well–controled size, the optimal test should have the highest power, which depends on the alternatives. Researchers should be careful with the alternatives to investigate the power properties.

[2] However, one drawback of Wald is that it is not invariant to the way the hypothesis is written unless it is linear.

Trinity Test Statistics

- **Wald**:

$$\sqrt{T}h(\widehat{\theta})'\left[\frac{\partial h}{\partial \theta'}\bigg|_{\theta=\widehat{\theta}} \mathcal{I}^{-1}(\widehat{\theta})\frac{\partial h'}{\partial \theta}\bigg|_{\theta=\widehat{\theta}}\right]^{-1}\sqrt{T}h(\widehat{\theta})$$

  If restriction is true $h(\widehat{\theta})$ close to 0, a metric of deviation

- **LR**:

$$2[\log L(\widehat{\theta}) - \log L(\widetilde{\theta})]$$

- **LM**:

$$\frac{1}{\sqrt{T}}\frac{\partial \log L}{\partial \theta'}\bigg|_{\theta=\widetilde{\theta}} \mathcal{I}^{-1}(\widetilde{\theta})\frac{1}{\sqrt{T}}\frac{\partial \log L}{\partial \theta}\bigg|_{\theta=\widetilde{\theta}}$$

  Score $\partial \log L/\partial\theta$ depicts the slope; at maximum close to 0; a metric of deviation.

In practice need to estimate information matrix ▸ Three Options

Alternative Test Stastistic of LM

Consider the constrained optimization problem

$$\max_{\theta} \log L - \lambda' h(\theta)$$

- FOCs imply $\frac{\partial \log L}{\partial \theta}\big|_{\theta=\widetilde{\theta}} = \frac{\partial h'}{\partial \theta}\big|_{\theta=\widetilde{\theta}}\widetilde{\lambda}$
- Substitute in LM expression and get

$$LM = \frac{1}{T}\widetilde{\lambda}' \frac{\partial h}{\partial \theta'}\bigg|_{\theta=\widetilde{\theta}} \mathcal{I}^{-1}(\widetilde{\theta}) \frac{\partial h'}{\partial \theta}\bigg|_{\theta=\widetilde{\theta}} \widetilde{\lambda}$$

- Blue term is a consistent estimate of the asymptotic covariance of $\frac{\widetilde{\lambda}}{\sqrt{T}}$, hence LM assesses whether the Lagrange multipliers are significantly different from 0
- Indeed, if null is true, by KKT constraints not binding, multiplier close to 0

Classic Linear Model

Assume conditional homoskedasticity, recall

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

where $Q_{XX} = \mathbb{E}(X'X)$ and $dim(\beta) = k$, $\widehat{\beta}_{OLS} = (X'X)^{-1}X'y$

- Suppose interested in testing the q linear restrictions:

  $$R\beta = r,$$

  where R is $q \times k$ and r is $q \times 1$
- Under the null ($R\beta - r = 0$), $\sqrt{T}(R\widehat{\beta}_{OLS} - r) \xrightarrow{d} \mathcal{N}(0, \sigma^2 R Q_{XX}^{-1} R')$, and hence

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

Remarks on Wald Test in Classical Linear Model

- Recall in linear model, $\widehat{\beta}_{OLS} = \widehat{\beta}_{MLE}$, $\widehat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - x_t'\widehat{\beta}_{OLS})^2$
- Recall Wald test, now $h(\beta) = R\beta - r$
- Given $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$, by Slutsky

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

- If we have q nonlinear restrictions $\psi(\beta) = 0$, by Delta method

$$W := \psi(\widehat{\beta}_{OLS})' \frac{\left[\frac{\partial \psi(\widehat{\beta}_{OLS})}{\partial \beta}(X'X)^{-1}\frac{\partial \psi'(\widehat{\beta}_{OLS})}{\partial \beta}\right]^{-1}}{\widehat{\sigma}^2}\psi(\widehat{\beta}_{OLS}) \xrightarrow{d} \chi_q^2$$

LR Test in Classical Linear Model

- Under the alternative:

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

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

$$\log L = \text{cons2} - \frac{T}{2}\log\widehat{\sigma}^2$$

- Recall regression estimator under restriction: ▸ Derivations

$$\widehat{\beta}_c = \widehat{\beta}_{OLS} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\widehat{\beta}_{OLS} - r)$$
$$\widehat{\sigma}_c^2 = \frac{1}{T}(y - X\widehat{\beta}_c)'(y - X\widehat{\beta}_c) = SSR_c$$

Hence $\log L_c = \text{cons2} - \frac{T}{2}\log\widehat{\sigma}_c^2$

- Form the LR statistics:

$$LR = 2(\log L - \log L_c) = T\log\left(\frac{SSR_c}{SSR}\right) = T\log\left(1 + \frac{qF}{T-k}\right)$$

LM Test in Classical Linear Model

- Scale lagrange multiplier by $2\widehat{\sigma}_c^2$:

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

- $\frac{1}{\sqrt{T}}\widehat{\lambda} \xrightarrow{d} \frac{1}{\sigma^2}[RQ_{XX}^{-1}R']^{-1}\mathcal{N}(0, \sigma^2 RQ_{XX}^{-1}R') \equiv \mathcal{N}(0, [\sigma^2 RQ_{XX}^{-1}R']^{-1})$
- Hence the LM test is

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

Remark:

- Recall Wald test:

$$W := (R\widehat{\beta}_{OLS} - r)' \frac{[R(X'X)^{-1}R']^{-1}}{\widehat{\sigma}^2}(R\widehat{\beta}_{OLS} - r)$$

Differs in terms of estimator of σ^2

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

- For simplicity, assume $R\beta = 0$. Proof applicable to $R\beta = r$ as well
- We start with two lemmas
  1. Define $\widehat{u}_c = y - X\widehat{\beta}_c$ and $\widehat{u} = y - X\widehat{\beta}_{OLS}$, then

$$W = \frac{\widehat{u}_c'\widehat{u}_c - \widehat{u}'\widehat{u}}{\widehat{\sigma}^2}, \quad LM = \frac{\widehat{u}_c'\widehat{u}_c - \widehat{u}'\widehat{u}}{\widehat{\sigma}_c^2}$$

  2. The Wald, LM and LR satisfy the following relations:
     - $W = 2[\log L(\widehat{\beta}_{OLS}, \widehat{\sigma}^2) - \log L(\widehat{\beta}_c, \widehat{\sigma}^2)]$
     - $LM = 2[\log L(\widehat{\beta}_{OLS}, \widehat{\sigma}_c^2) - \log L(\widehat{\beta}_c, \widehat{\sigma}_c^2)]$
     - $LR = 2[\log L(\widehat{\beta}_{OLS}, \widehat{\sigma}^2) - \log L(\widehat{\beta}_c, \widehat{\sigma}_c^2)]$
- **Theorem:** $LM \leq LR \leq W$
- Result applies to the general framework:

$$y|X \sim \mathcal{N}(X\beta, \sigma^2\Omega), \quad \Omega = \Omega(\omega),$$

where ω is a finite estimable parameter vector

Lemma 1 Proof | Lemma 2 Proof | Inequality Proof

When are LM, LR and W the same?

- Suppose σ^2 is known, by Lemma 2 the three have the same form
- More generally, if the log likelihood has the following form

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

  where A is symmetric and positive definite, then $LM = LR = W$

Sketch proof:

1. $\partial \log L / \partial \theta = -A'(\theta - \widehat{\theta})$
2. $\partial^2 \log L / \partial \theta \partial \theta' = -A = -T \mathcal{I}_T(\theta)$
3. Plug them into respective forms of the three tests

Remarks:

1. Key: $\partial^2 \log L / \partial \theta \partial \theta'$ is the same when evaluated at $\widehat{\theta}$ or θ_0
2. Asymptotically, if $\widehat{\theta}$ close to θ_0, likelihood function in the neighborhood of θ_0 approximately quadratic; hence asymptotic equivalence of the three tests ▸ Math

LM Test as a Diagnostic

- Researchers usually don't know the exact variables, functional forms and distribution implicit in a particular theory, thus requires a specification search
- For hypothesis testing, null is a specification in favor, alternative is a more general specification
- Test for this purpose is a **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 teh estimates under the null
- Each alternative considered indicates a particular type of non−randomness

Testing for Heteroskedasticity: Breush–Pagan LM

Consider the following model:

$$y = X\beta + u, \quad u_t \sim \mathcal{N}(0, \sigma_t^2), \quad t = 1, \cdots, T$$

- 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 is a q–vector of variables with $z_{1t} = 1$, hence null can be rewritten as

  $$H_0 : \alpha_2 = \cdots = \alpha_q = 0$$

- Under both bull and alternative, assume no serial correlation: $\mathbb{E}(u_t u_s) = 0$ for $t \neq s$

Breush–Pagan Test Statistic

- Log likelihood function

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

- Hessian is block diagonal ▸ Details
- Under null $\partial\mathcal{L}/\partial\beta|_{\alpha=0,\widehat{\beta}} = 0$: OLS residuals orthogonal to regressors and $h(0)$ is a constant
- LM test statstic:

$$LM = \left(\frac{\partial\mathcal{L}}{\partial\alpha}\right)'\left[-\mathbb{E}\left(\frac{\partial^2\mathcal{L}}{\partial\alpha\partial\alpha'}\right)\right]^{-1}\left(\frac{\partial\mathcal{L}}{\partial\alpha}\right)\Big|_{\alpha=0,\widehat{\beta}}$$

- Evaluated under the null, simplified to

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

where $f' = [f_1, \ldots, f_T]$, $f_t = \frac{\widehat{u}_t}{\widehat{\sigma}^2} - 1 \equiv g_t - 1$ and $Z = [z_1, \ldots, z_T]$

Breush–Pagan LM Test: Compact Form

- Denote $f = g - i$, where $g' = [g_1, \ldots, g_T]$ and i is a column of vector of ones
- $LM = \frac{1}{2}(g - i)' P_Z (g - i)$
- As $P_Z i = i$ and $g' i = i' i = T$, simplify as $LM = \frac{1}{2}(g' P_Z g - T)$
- In the regression of g on Z, explained sum of squares (ESS) is

$$ESS := \widehat{g}'\widehat{g} - T\overline{g}^2 = g' P_Z g - T\overline{g}^2$$

- Since $g' i = T$, $\overline{g} = 1$, hence

$$LM = \frac{1}{2}[g' P_Z g - T\overline{g}^2] = \frac{1}{2} ESS$$

Breush–Pagan LM Test: Procedures

1. Apply OLS to $y = X\beta + u$ and obtain residuals $\widehat{u}$
2. Compute $g_t = \frac{\widehat{u}_t}{\widehat{\sigma}^2}$, where $\widehat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2$
3. Run OLS on $g = (g_1, \ldots, g_T)$ on Z (including a constant) and compute LM statistic

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

   where ESS denotes the explained sum of squares

Under the null, $LM \xrightarrow{d} \chi^2_{q-1}$

Testing for Serial Correlation: Durbin–Watson

- First consider the case without lagged dependent variables
- Consider the model with k regressors

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

- Consider the following test statstic

$$
d = \frac{\sum_{t=2}^{T} (\widehat{u}_t - \widehat{u}_{t-1})^2}{\sum_{t=1}^{T} \widehat{u}_t^2},
$$

where $\{\widehat{u}_t\}$ are OLS residuals

- Why this statistic? Note that

$$
d \approx \frac{2\sum_{t=2}^{T} (\widehat{u}_t^2 - \widehat{u}_t \widehat{u}_{t-1})}{\sum_{t=2}^{T} \widehat{u}_{t-1}^2} = 2 - 2\frac{\sum_{t=2}^{T} \widehat{u}_t \widehat{u}_{t-1}}{\sum_{t=2}^{T} \widehat{u}_{t-1}^2} = 2(1 - \widehat{\rho})
$$

- If $\rho = 0$, $d \approx 2$; if $\rho = 1$, $d \approx 0$. Test if d much smaller than 2

Using Durbin–Watson Test

- H_0: errors not serially correlated; H_1: correlated at order 1
- Exact distribution depends on X ▸ Math Intuition
- Can bound this dependence as a function of k and T
- Hence critical values depend on k, T, size α and whether we consider upper or lower bound

Testing procedures for positive correlation:

1. Run OLS on y against X and obtain $\{\widehat{u}_t\}$
2. Compute d
3. If $d < d_L(k, T, \alpha)$ reject H_0; if $d > d_u(k, T, \alpha)$ do not reject; if $d_L \leq d \leq d_u$ test is inconclusive

Durbin's "h" Statistic

- Adding lagged dependent variables in X biases $\widehat{\rho}$ downward, therefore needs a correction
- Denote $\widehat{\alpha}_1$ as OLS coefficient on the lagged dependent variable, consider the following statstic:

$$h = \widehat{\rho}\sqrt{\frac{T}{1 - T\widehat{V}(\widehat{\alpha}_1)}} \approx \left(1 - \frac{d}{2}\right)\sqrt{\frac{T}{1 - T\widehat{V}(\widehat{\alpha}_1)}},$$

  where $\widehat{V}(\widehat{\alpha}_1)$ is the variance estimator of $\widehat{\alpha}_1$ in the regression $y_t = \alpha_1 y_{t-1} + \cdots + \alpha_s y_{t-s} + X_t'\beta + u_t$ and $\widehat{\rho} = \frac{\sum_{t=2}^{T} \widehat{u}_t \widehat{u}_{t-1}}{\sum_{t=2}^{T} \widehat{u}_{t-1}^2}$ using the residual of the previous regression
- Under $H_0 : h \xrightarrow{d} \mathcal{N}(0, 1)$
- **Caveat**: can have $T\widehat{V}(\widehat{\alpha}_1) \geq 1$, test statstic undefined

LM Test for $AR(p)$ Errors (Godfrey, 1978)

Consider the model with k regressors in X and p lags in errors

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

- Ignoring the first p obs 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: $y_t = f_t(\beta, \psi) + \varepsilon_t$
- Denote $\theta := (\beta, \psi)$. Take first-order Taylor expansion around $\psi = \mathbf{0}$ evaluated at the restricted estimate $\widetilde{\beta}$ (i.e., under the null)

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

- Under the null, $u_t = \varepsilon_t$

Godfrey Test Derivations

- Since $\left(\frac{\partial f_t}{\partial \beta_j}\right)|_{\widetilde{\beta},\mathbf{0}} = x_{tj}$, $j = 1, \ldots, k$; $\left(\frac{\partial f_t}{\partial \psi_j}\right)|_{\widetilde{\beta},\mathbf{0}} = \widetilde{u}_{t-j}$, $j = 1, \ldots, p$

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

- The matrix form of this $AR(p)$ process is

$$\underbrace{y - f}_{(T-p)\times 1} = \underbrace{F}_{(T-p)\times(k+p)} \begin{pmatrix} \beta_1 - \widetilde{\beta}_1 \\ \vdots \\ \beta_k - \widetilde{\beta}_k \\ \psi_1 \\ \vdots \\ \psi_p \end{pmatrix} + \varepsilon \equiv F\theta + \varepsilon$$

- The log likelihood of $AR(p)$ after concentration of σ^2 is

$$\mathcal{L} = \text{const} - \frac{T-p}{2} \log(\varepsilon'\varepsilon),$$

Godfrey Test: Compact Form

- $\frac{\partial \mathcal{L}(\widetilde{\theta})}{\partial \theta} = \frac{F'(y-f)}{\widetilde{\sigma}^2}$ and $V(\widetilde{\theta}) = \widetilde{\sigma}^2 (F'F)^{-1}$, where $\widetilde{\sigma}^2$ is the sum of squared residuals (SSR) of $(*)$ under the null, i.e., OLS residuals
- Construct LM statistic

$$LM = \Big(\frac{\partial \mathcal{L}(\widetilde{\theta})}{\partial \theta}\Big)' V^{-1}(\widetilde{\theta}) \Big(\frac{\partial \mathcal{L}(\widetilde{\theta})}{\partial \theta}\Big)$$
$$= \frac{1}{\widetilde{\sigma}^2}\big[(y-f)'F(F'F)^{-1}F'(y-f)\big]$$

- Under the null: $LM \xrightarrow{d} \chi^2_p$

Remarks:

- $\widetilde{\sigma}^2 = (y-f)'(y-f)/(T-p) \equiv TSS/(T-p)$, where TSS denotes total sum of squares from $(*)$
- $(y-f)'P_F(y-f)$ is the explained sum of squares (ESS) of $(*)$
- Therefore $LM = (T-p)\frac{ESS}{TSS} = (T-p)R^2$, where R^2 is calculated from $(*)$

Godfrey Test Procedures

1. Run OLS[3] on y against X and p lagged y and get $\widehat{u}$
2. Run OLS on the auxiliary regression

$$\widehat{u} = X\tau + \underline{\widehat{u}}\delta + v,$$

where $\underline{\widehat{u}} := (\widehat{u}_{-1}, \ldots, \widehat{u}_{-p})$ is $(T-p) \times p$. The j–th column is $[\widehat{u}_{1-j}, \ldots, \widehat{u}_{T-j}]$; if $t \leq j$ the entry $\widehat{u}_{t-j} = 0$

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

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

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

[3]Note: The notation is still X but keep in mind that we run regression on lagged dependent variables and hence need to drop the first p observations of X. This applies to the next step as well.

Consistent Estimates of Information Matrix

$$\widehat{\mathcal{I}}_{1,T}(\theta^*) = -\frac{1}{T}\sum_{t=1}^{T}\frac{\log L_t(y;\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta^*}$$

$$\widehat{\mathcal{I}}_{2,T}(\theta^*) = \frac{1}{T}\sum_{t=1}^{T} s_t(y;\theta^*)s_t'(y;\theta^*)$$

$$\widehat{\mathcal{I}}_{3,T}(\theta^*) = -\frac{1}{T}\mathbb{E}\left[\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right]\Big|_{\theta=\theta^*}$$

- θ^* either denotes restricted or unrestricted estimate
- s_t denotes the score function
- Performance might differ in finite sample simulations

◂ Back

Restricted OLS Derivations

Constrained optimziation problem

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

- Two first order conditions:

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

- Combine the two FOCs and rearrange:

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

- Plug in $\widehat{\lambda}$ to obtain:

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

Remarks:

- If constraints are exactly satisfied, $\widehat{\beta}_c = \widehat{\beta}_{OLS}$
- $SSR_r := (y - X\widehat{\beta}_c)'(y - X\widehat{\beta}_c) \geq SSR_u := (y - X\widehat{\beta}_{OLS})'(y - X\widehat{\beta}_{OLS})$
- $Var(\widehat{\beta}_c|X) \leq Var(\widehat{\beta}_{OLS}|X)$ whether or not constraints are true
- $MSE(\widehat{\beta}_c|X)$

Back

Proof of Lemma 1

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

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

Plug this identity into the formulae of Wald and LM test statstics ◂ Back

Proof of Lemma 2

- For LR test this is the definition. For Wald test:

$$\log L(\widehat{\beta}_{OLS}, \widehat{\sigma}^2) = -\frac{T}{2}\log\widehat{\sigma}^2 - \frac{T}{2}\log 2\pi - \frac{1}{2}\frac{\widehat{u}'\widehat{u}}{\widehat{\sigma}^2}$$
$$\log L(\widehat{\beta}_c, \widehat{\sigma}^2) = -\frac{T}{2}\log\widehat{\sigma}^2 - \frac{T}{2}\log 2\pi - \frac{1}{2}\frac{\widehat{u}_c'\widehat{u}_c}{\widehat{\sigma}^2}$$

By Lemma 1, $W = \frac{\widehat{u}_c'\widehat{u}_c - \widehat{u}'\widehat{u}}{\widehat{\sigma}^2} = 2[\log L(\widehat{\beta}_{OLS}, \widehat{\sigma}^2) - \log L(\widehat{\beta}_c, \widehat{\sigma}^2)]$

- For LM test,

$$\log L(\widehat{\beta}_{OLS}, \widehat{\sigma}_c^2) = -\frac{T}{2}\log\widehat{\sigma}_c^2 - \frac{T}{2}\log 2\pi - \frac{1}{2}\frac{\widehat{u}'\widehat{u}}{\widehat{\sigma}_c^2}$$
$$\log L(\widehat{\beta}_c, \widehat{\sigma}_c^2) = -\frac{T}{2}\log\widehat{\sigma}_c^2 - \frac{T}{2}\log 2\pi - \frac{1}{2}\frac{\widehat{u}_c'\widehat{u}_c}{\widehat{\sigma}_c^2}$$

By Lemma 1, $LM = \frac{\widehat{u}_c'\widehat{u}_c - \widehat{u}'\widehat{u}}{\widehat{\sigma}_c^2} = 2[\log L(\widehat{\beta}_{OLS}, \widehat{\sigma}_c^2) - \log L(\widehat{\beta}_c, \widehat{\sigma}_c^2)]$

◂ Back

Proof of Inequality

- $LR \geq LM$ if
$$\log L(\widehat{\beta}_{OLS}, \widehat{\sigma}^2) \geq \log L(\widehat{\beta}_{OLS}, \widehat{\sigma}_c^2)$$
The inequality holds since $\widehat{\beta}_{OLS}$ and $\widehat{\sigma}^2$ maximizes the likelihood[4]
- $W \geq LR$ if
$$\log L(\widehat{\beta}_c, \widehat{\sigma}^2) \leq \log L(\widehat{\beta}_c, \widehat{\sigma}_c^2)$$
This indeed holds. Let B_0 denote the space of β such that $R\beta = 0$, then
$$L(\widehat{\beta}_c, \widehat{\sigma}^2) = \sup_{\beta \in B_0} L(\beta, \widehat{\sigma}^2) \leq \sup_{\beta \in B_0, \sigma^2 \in B_\sigma} L(\beta, \sigma^2),$$
given $\widehat{\beta}_c$ and $\widehat{\sigma}_c^2$ maximizes the likelihood.
- **Remark:** This result may not hold for more complex models

◂ Back

[4] Recall PS3 Q3.

Asymptotic Equivalence

- Taylor expansion of the likelihood function

$$L(y;\theta) = L(y;\widehat{\theta}) - \frac{T}{2}(\widehat{\theta}-\theta)'\mathcal{I}_T(\theta_0)(\widehat{\theta}-\theta) + o_p(1)$$

- Asymptotically $\widehat{\theta} \xrightarrow{p} \theta_0$ under the null and

$$\mathcal{I}_T(\theta_0) \to \lim_{T\to\infty} -\frac{1}{T}\mathbb{E}\Big[\frac{\partial^2 \log L(\theta_0)}{\partial\theta\partial\theta'}\Big]$$

  Hence we have a quadratic form
- The matrix in the quadratic form is the same in samples with either $\widehat{\theta}$ or θ_0 as $\widehat{\theta} \xrightarrow{p} \theta_0$
- Also applies to composite null in which only a subset of parameters are fixed (use partitioned matrix)

◂ Back

Hessian in Breush–Pagan Log Likelihood

- $\frac{\partial \mathcal{L}}{\partial \beta} = \sum_{t=1}^{T} (y_t - x_t' \beta) x_t / h(z_t' \alpha)$
- $\frac{\partial^2 \mathcal{L}}{\partial \beta_j \partial \alpha_i} = -\frac{[\sum_{t=1}^{T} (y_t - x_t' \beta) x_{j,t}] h'(z_t' \alpha) z_{i,t}}{[h(z_t' \alpha)^2]}$
- Since $x_{i,t}$ and $z_{i,t}$ are fixed,

$$\mathbb{E}\left[\frac{\partial^2 \mathcal{L}}{\partial \beta_j \partial \alpha_i}\right] = 0$$

Hence the Hessian is block–diagonal

◂ Back

Durbin–Watson Distribution Depends on X

Can rewrite d to be

$$d = \frac{\widehat{u}' A \widehat{u}}{\widehat{u}' \widehat{u}} = \frac{u'(I - P_X)A(I - P_X)u}{u'(I - P_X)u},$$

where

$$A = \begin{pmatrix} 1 & -1 & & & & \\ -1 & 2 & \ddots & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & 2 & -1 \\ & & & & -1 & 1 \end{pmatrix}$$

blank space means 0 ◂ Back