

Testing Linear Restrictions - General Types of Tests: Lecture XVII

Charles B. Moss ¹

¹University of Florida

February 16, 2012

- 1 Readings
- 2 Definitions and Intuitions - Engle
 - Tests as Random Variables
 - Size and Power of Test
 - Regions of Test
 - Three Specific Tests
 - Definition of the Test Statistics
- 3 Translog Cost Function Example

Readings

- * Greene, W.H. 2012. *Econometric Analysis* Seventh Edition (Chapter 5 Section 5.3 5.7 pp. 114 133)
- * Engle, R.F. 2004. Chapter 13: Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics. *Handbook of Econometrics* Volume 2, 775-826.
(<http://www.sciencedirect.com/science/article/pii/S15734412840200>)
- * Hansen, B.E. 2012. *Economics* Draft Textbook. (Chapter 8 pp. 157-177).

Definitions and Intuitions - Engle

- Hypothesis testing concerns the question of whether data appear to favor or disfavor a particular description of nature.
 - Testing is inherently concerned with one particular hypothesis which will be called the *null* hypothesis.
 - If the data fall into a particular region of the sample space called the *critical region* then the test is said to *reject* the null hypothesis, otherwise it *accepts*.
- Engle then recounts this statement - we can never truly "accept" the null hypothesis; only "fail to reject it."

Tests as Random Variables

- Taking a slightly different attack – Our challenge to define some measure (or more specifically a metric) based on the sample by which we can make a decision to either **reject** or **fail to reject** a null hypothesis.
- This metric is going to be defined (or maybe better computed) based on outcomes of the random variables in the sample.
 - Because the metric is a function of the sample it is also a random variable with a distribution.
 - Hence, the outcome of the test metric has a probability density function.
- Based on this probability density function we define the standard testing errors:
 - **Type I Errors** are committed when the null hypothesis is falsely rejected, and
 - **Type II Errors** occur when it is incorrectly accepted.

Size and Power of Test

- We call α the **size** of the test which is the probability of **Type I Error**.
- We call β the probability of **Type II Errors**.
- The **power** of a test is the probability of rejecting the null when it is false, which is therefore $1 - \beta$.
- Moss (pp. 118-130) presents a discussion of the tradeoff between α and β .
- **Definition 8.3 [p.119]** Let (α_1, β_1) and (α_2, β_2) be the characteristics of two tests. The first test is better (or more powerful) than the second test if $\alpha_1 \leq \alpha_2$ and $\beta_1 \leq \beta_2$ with a strict inequality holding for at least one point.

Size and Power of Test

- **Definition 8.4 [p. 120]** A test is inadmissible if there exists another test which is better in the sense of Definition 8.3. Otherwise it is called admissible.
- **Definition 8.5 [p. 120]** R is the most powerful test of size α if $\alpha|R$ *alpha* and for any test R_1 of size α , $\beta(R) \leq \beta(R_1)$.

Regions of Test

- Suppose that we have a test metric R defined based on Definitions 8.3, 8.4, and 8.5. (For example $\bar{R} = \sum_{t=1}^T y_t / T$).
 - This metric is based on calculations based on a sample y which is a $T \times 1$ vector.
 - This sample is based on a distribution function $f(y, \theta)$ where θ is a $k \times 1$ vector (i.e., in the case of a normal distribution $\theta = \{ \mu, \sigma^2 \}$ so $k = 2$).
 - The purpose of the test statistic is to determine whether the vector of parameters belongs to a set $\theta \in \Theta_0 \subset \Theta$ (the null hypothesis), or whether the vector of parameters belongs to a set $\theta \in \Theta_1 \subset \Theta$.
 - A basic assumption is that these subsets are different $\Theta_0 \cap \Theta_1 = \{ \}$.
- In general the alternative assumption implies a composite or the broadest possible set of outcomes $\Theta_1 = \Theta - \Theta_0$.

Regions of Test - Continued

- The critical region is then defined by that set of outcomes (C_R) which define the rejection of the null hypothesis.
 - For example if $\bar{R} \in C_R$ then we reject the hypothesis (i.e., concluding that $\theta \in \Theta_2$).
 - Given this structure, we can define the size of the test (level of Type I error) as

$$\alpha_T = \Pr(\bar{R} \in C_R | \theta \in \Theta_0) \quad (1)$$

- The power of the test then defined as

$$\pi_T(\theta) = \Pr(\bar{R} \in C_R | \theta), \text{ for } \theta \in \Theta_1. \quad (2)$$

Three Specific Tests

- For this course, I want to start by defining three different types of tests:
 - Wald Tests - This is the typical metric defined as a difference in parameters normalized by the standard deviation

$$t = \frac{\hat{\beta} - \beta}{\sigma} \sim N(0, 1). \quad (3)$$

- Likelihood Ratio Tests - A comparison of the two optimal functions

$$\begin{aligned} L_1 &\Rightarrow \max_{\theta} \prod_{t=1}^T f(y|\theta \in \Theta_1) \\ L_2 &\Rightarrow \max_{\theta} \prod_{t=1}^T f(y|\theta \in \Theta_2) \\ t &= 2 [\ln(L_2) - \ln(L_1)]. \end{aligned} \quad (4)$$

Three Specific Tests - Continued

- Continued
 - Lagrange Multiplier Tests - The instantaneous loss of the likelihood function at the optimal point

$$\lambda(\theta) \Leftarrow \max_{\theta} \sum_{t=1}^T \ln [f(y|\theta \in \Theta)] - \lambda(\theta \in \Theta_1) \quad (5)$$

Three Specific Tests - Continued

- Essentially, the Lagrange Multiplier approach starts at the null and asks whether movement toward the alternative would be an improvement,
- while the Wald approach starts at the alternative and considers movement towards the null.
- The Likelihood ratio method compares the two hypothesis directly on an equal basis (Engle p. 776).

Definition of the Test Statistics

- To define the test statistics we define the logarithm of the likelihood function

$$\tilde{L}(y, \theta) = \ln [L(y, \theta)] = \sum_{t=1}^T \ln [f(y_t, \theta)] \quad (6)$$

- Given that the likelihood is maximized at $\hat{\theta}$ we have

$$\nabla_{\theta'} \tilde{L}(y, \theta) |_{\theta \rightarrow \hat{\theta}} = 0. \quad (7)$$

The score (defined as $s(\theta, y) = \nabla_{\beta'} \tilde{L}(y, \theta)$) is equal to zero at the MLE (maximum likelihood estimate).

Definition of the Test Statistics - Continued

- The variance of $\hat{\theta}$ is defined using the Cramer-Rao lower bound (as stated in Engle [p.780] as the inverse of Fisher's Information Matrix)

$$V(\hat{\theta}) = I^{-1}(\theta) / T \quad (8)$$

$$I(\theta) = -E \frac{\partial^2 \tilde{L}(y, \theta)}{\partial \theta \partial \theta'} = -\nabla_{\theta \theta'}^2 \tilde{L}(y, \theta)$$

- Thinking about the variance matrix - we have been dealing with a variance matrix of

$$V(\beta) = s^2 (X'X)^{-1}. \quad (9)$$

Definition of the Test Statistics - Continued

- Suppose we are interested in the relative or normalized variance of an alternative hypothesis β_a in comparison with our ordinary least squares estimates $\hat{\beta}$. We could compute this relative or normalized variance as

$$\begin{aligned}
 t_{\beta_a} &= (\hat{\beta} - \beta_a)' [V(\hat{\beta})]^{-1} (\hat{\beta} - \beta_a) \\
 &\Rightarrow (\hat{\beta} - \beta_a)' [s^2 (X'X)^{-1}] (\hat{\beta} - \beta_a) \\
 &\Rightarrow (\hat{\beta} - \beta_a)' \left[\frac{1}{s^2} (X'X) \right] (\hat{\beta} - \beta_a)
 \end{aligned} \tag{10}$$

Definition of the Test Statistics - Continued

- As developed by Engle (p.780) the Wald test is

$$\xi_W = T \left(\hat{\theta} - \theta_0 \right)' I \left(\hat{\theta} \right) \left(\hat{\theta} - \theta_0 \right) \quad (11)$$

which is identical to our ordinary least squares tests assuming F and t statistics.

- The likelihood ratio test is then

$$\xi_{LR} = -2 \left[\tilde{L}(y, \theta_0) - \tilde{L}(y, \hat{\theta}) \right] \quad (12)$$

- The Lagrange multiplier test becomes

$$\xi_{LM} = s'(\theta_0, y) I^{-1}(\theta_0) s(\theta_0, y) / T \quad (13)$$

Translog Cost Function Example

- The dual translog cost function can be written as

$$\begin{aligned} \ln [C(y, w)] = & \alpha_0 + \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}' \begin{bmatrix} \ln [w_1] \\ \ln [w_2] \\ \ln [w_w] \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \ln [w_1] \\ \ln [w_2] \\ \ln [w_w] \end{bmatrix}' \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{12} & A_{22} & A_{23} \\ A_{13} & A_{23} & A_{33} \end{bmatrix} \begin{bmatrix} \ln [w_1] \\ \ln [w_2] \\ \ln [w_w] \end{bmatrix} \\ & + \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}' \begin{bmatrix} \ln [y_1] \\ \ln [y_2] \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \ln [y_1] \\ \ln [y_2] \end{bmatrix}' \begin{bmatrix} B_{11} & B_{12} \\ B_{12} & B_{22} \end{bmatrix} \begin{bmatrix} \ln [y_1] \\ \ln [y_2] \end{bmatrix} \\ & + \begin{bmatrix} \ln [w_1] \\ \ln [w_2] \\ \ln [w_3]' \end{bmatrix} \begin{bmatrix} \Gamma_{11} & \Gamma_{12} & \Gamma_{13} \\ \Gamma_{21} & \Gamma_{22} & \Gamma_{23} \end{bmatrix} \begin{bmatrix} \ln [y_1] \\ \ln [y_2] \end{bmatrix} \end{aligned} \quad (14)$$

Translog Cost Function Example - Continued

- Applying Shephard's lemma and taking the derivative of the cost function with respect to input price w_1 yields

$$s_1 = \alpha_1 + A_{11} \ln[w_1] + A_{12} \ln[w_2] + A_{13} \ln[w_3] + \Gamma_{11} \ln[y_1] + \Gamma_{21} \ln[y_2] \quad (15)$$

- Where s_1 is the share of cost spent on input 1 ($w_1 x_1 / \sum_{j=1}^3 w_j x_j$).
- Formulating the regression

$$s_1 = y = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}' \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{bmatrix} = \begin{bmatrix} 1 \\ \ln[w_1] \\ \ln[w_2] \\ \ln[w_3] \\ \ln[y_1] \\ \ln[y_2] \end{bmatrix}' \begin{bmatrix} \alpha_1 \\ A_{11} \\ A_{12} \\ A_{13} \\ \Gamma_{11} \\ \Gamma_{21} \end{bmatrix} \quad (16)$$

Translog Cost Function Example - Continued

- Following our standard approach, we estimate the values of the parameters using ordinary least squares yielding

$$\hat{\beta} = (X'X)^{-1} X'y = \begin{bmatrix} -0.3002 \\ 0.0608 \\ 0.1120 \\ -0.1247 \\ -0.1438 \\ 0.2042 \end{bmatrix} \quad (17)$$

- Estimating the s^2 or Mean Squared Error

$$s^2 = \frac{y' \left[I_{T \times T} - X (X'X)^{-1} X' \right] y}{n - k} = 0.0009863 \quad (18)$$

Translog Cost Function Example - Continued

- Using the s^2 we then estimate the variance matrix for the parameters

$$V(\hat{\beta}) = s^2 (X'X)^{-1}$$

$$= \begin{bmatrix} 0.072021 & 0.003030 & -0.005358 & -0.005012 & -0.001589 & -0.004737 \\ 0.003030 & 0.012226 & 0.003932 & -0.006454 & -0.001762 & 0.001401 \\ -0.005358 & 0.003932 & 0.002092 & -0.001147 & -0.000038 & 0.000477 \\ -0.005012 & -0.006454 & -0.001147 & 0.007121 & 0.002319 & -0.001817 \\ -0.001589 & -0.001762 & -0.000038 & 0.002319 & 0.003167 & -0.002938 \\ -0.004737 & 0.001401 & 0.000477 & -0.001817 & -0.002938 & 0.003268 \end{bmatrix} \quad (19)$$

- As a starting point, consider the t-Test for $\beta_2 = 0$

$$t^* = \frac{\hat{\beta}_2 - 0}{\hat{\sigma}_{\beta_2}} = \frac{0.0608 - 0}{\sqrt{0.012226}} = 0.5501 \sim t_{\alpha, n-k} = t_{0.05, 46-6} \quad (20)$$

Translog Cost Function Example - Continued

Parameter	Estimate	Std. Dev.	t-Ratio	p-Value
β_1	-0.3002	0.2684	-1.1186	0.2700
β_2	0.0608	0.1106	0.5501	0.5853
β_3	0.1120	0.0457	2.4496	0.0188
β_4	-0.1247	0.0844	-1.4774	0.1474
β_5	-0.1438	0.0563	-2.5555	0.0145
β_6	0.2042	0.0572	3.5722	0.0009

Translog Cost Function Example - Continued

Parameter	Estimate
β_1	-0.3002 (0.2684)
β_2	0.0608 (0.1106)
β_3	0.1120** (0.0457)
β_4	-0.1247 (0.0844)
β_5	-0.1438** (0.0563)
β_6	0.2042*** (0.0572)

Translog Cost Function Example - Continued

- Like the other model, we need to impose homogeneity

$$A_{11} + A_{12} + A_{13} = 0 \quad (21)$$

- The restriction we want to test is then $\beta_2 + \beta_3 + \beta_4 = 0$ or a restriction matrix of

$$R = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}, r = [0] \quad (22)$$

Translog Cost Function Example - Continued

- We can directly test this restriction as

$$F = \frac{(r - R\beta)' \left[R (X'X)^{-1} R' \right]^{-1} (r - R\beta) / (q\{ = 1 \})}{\epsilon(\beta)' \epsilon(\beta) / (n - k)} = 0.1646 \quad (23)$$

- Given that this value is distributed $F(1, 41)$, the p-value of this statistic is 0.6871.