Non-Nested Tests

Shelby Brewer and Enrique Salonga

ECON 5350: Advanced Econometric Theory I

December 8, 2022

Background

Types of Model Misspecification Tests:

1. **Nested Models:** One model that can be written as a special case of the other

   (a) $Y_i = \beta_1 + \beta_2 X_i + \epsilon_i$
   (b) $Y_i = \beta_1 + \epsilon_i$

   Model B is nested on A since $\beta_2 = 0$.

Background

Types of Model Misspecification Tests:

2. **Non-Nested Models:** One model cannot be written as a special case of the other – the tested models are completely independent
   (a) $Y_i = \beta_1 + \beta_2 X_i + \epsilon_i$
   (b) $Y_i = \alpha_1 + \alpha_2 Z_i + \mu_i$

Roadmap

1. Approach 1: J-Test
2. MATLAB Example
3. Approach 2: Encompassing Model

Testing Non-Nested Alternatives

Example: Deciding whether an independent variable should appear in linear or log form

$$Y_i = \beta_1 + \beta_2 X_{1i} + \beta_3 X_{2i} + \epsilon_i$$

vs.

$$Y_i = \alpha_1 + \alpha_2 \log(X_{1i}) + \alpha_3 \log(X_{2i}) + \mu_i$$

Two Testing Approaches for Non-Nested Alternatives:

1. J-Test
2. Encompassing Approach

J-Test

Consider two non-nested models:

$$\text{(a)} \quad Y_i = \alpha_1 + \alpha_2 X_{1i} + \alpha_3 X_{2i} + \mu_i$$
$$\text{(b)} \quad Y_i = \beta_1 + \beta_2 X_{3i} + \beta_3 X_{4i} + \epsilon_i$$

We want to know which model is best to use, so we use the J-test.

J-Test: Reference Model (a)

$$\text{(a)} \quad Y_i = \alpha_1 + \alpha_2 X_{1i} + \alpha_3 X_{2i} + \mu_i$$
$$\text{(b)} \quad Y_i = \beta_1 + \beta_2 X_{3i} + \beta_3 X_{4i} + \epsilon_i$$

J-test steps when Model (a) is reference model:

- Assume Model (a) is H_0 and H_a is not Model (a)(i.e., $H_0 : \alpha_4 = 0$ and $H_a : \alpha_4 \neq 0$)
- Use OLS to estimate Model (b)
- Use estimated Y values ($\hat{Y}_b$) from (b) as additional regressor on Model (a) giving Model (c)

$$\text{(c)} \quad \hat{Y} = \hat{\alpha}_1 + \hat{\alpha}_2 X_{1i} + \hat{\alpha}_3 X_{2i} + \hat{\alpha}_4 \hat{Y}_b + \mu_i$$

J-Test: Reference Model (a)

$$\text{(c)} \quad \hat{Y} = \hat{\alpha}_1 + \hat{\alpha}_2 X_{1i} + \hat{\alpha}_3 X_{2i} + \hat{\alpha}_4 \hat{Y}_b + \mu_i$$

- Estimate Model (c) to check whether $\hat{\alpha}_4$ is statistically significant

Decision Criteria:

1. If $\hat{\alpha}_4$ is significant, reject H_0, Model (b) is better than Model (a)
2. If $\hat{\alpha}_4$ is not significant, fail to reject H_0, accept Model (a) and reject Model (b)

J-Test: Reference Model (b)

Recall the models we're testing:

$$(a)\quad Y_i = \alpha_1 + \alpha_2 X_{1i} + \alpha_3 X_{2i} + \mu_i$$
$$(b)\quad Y_i = \beta_1 + \beta_2 X_{3i} + \beta_3 X_{4i} + \epsilon_i$$

This time use J-test with Model (b) as the reference model:

- Assume Model (b) is H_0 (i.e., $H_0 : \beta_4 = 0$ and $H_a : \beta_4 \neq 0$)
- Use OLS to estimate Model (a)
- Use estimated Y values ($\hat{Y}_a$) from (a) as additional regressor on Model (b) giving Model (d)

$$(d)\quad \hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_{3i} + \hat{\beta}_3 X_{4i} + \hat{\beta}_4 \hat{Y}_a + \epsilon_i$$

J-Test: Reference Model (b)

(d) $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_{3i} + \hat{\beta}_3 X_{4i} + \hat{\beta}_4 \hat{Y}_a + \epsilon_i$

- Estimate Model (d) to check whether $\hat{\beta}_4$ is statistically significant

Decision Criteria:

1. If $\hat{\beta}_4$ is significant, reject H_0, Model (a) is better than Model (b)
2. If $\hat{\beta}_4$ is not significant, fail to reject H_0, accept Model (b) and reject Model (a)

Possible Outcomes

- The J-test using each Model as a reference yields four possible outcomes:

		Model (a)	
		Fail to Reject	**Reject**
Model (b)	**Fail to Reject**	Accept Both	Accept B, Reject A
	Reject	Accept A, Reject B	Reject Both

- Clearly, two of these outcomes are not ideal!

Limitations of Non-Nested Testing

1. Clear winners may not emerge among models
   - Both models could be rejected
   - Neither model could be rejected (both could be accepted)
2. Even if the null is rejected, we're still unsure if the chosen model is actually correct
3. Non-Nested tests do not apply if the models have different dependent variables (Y vs. $\log(Y)$)

Matlab J-Test: Which Model is Better?

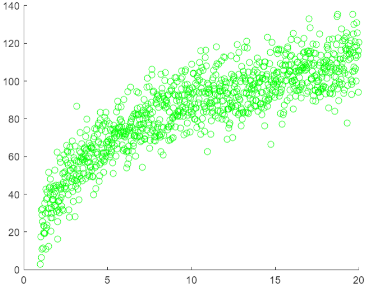

Figure: Model A: $Y = a_1 + a_2 \log(X_1) + u_i$

Matlab J-Test: Which Model is Better?

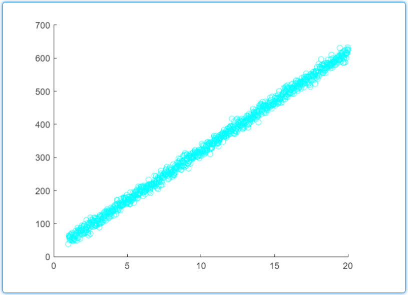

Figure: Model B: $Y = b_1 + b_2 X_1 + e_i$

MATLAB J-Test: Reference Model (a)

Assume:

- $Y \equiv$ annual salary
- $X_1 \equiv$ years of education

With potential models:

$$\text{(a)} \quad Y = a_1 + a_2 \log(X_1) + u_i$$

$$\text{(b)} \quad Y = b_1 + b_2 X_1 + e_i$$

Using Model (a) as the reference model

$$\text{(c)} \quad \hat{Y} = \hat{a}_1 + \hat{a}_2 \log(X_1) + \hat{a}_3 \hat{Y}_b + u_i$$

- H_0: Model (a) ($a_3 = 0$)
- H_a: Not Model (a) ($a_3 \neq 0$)

If a_4 is significant, we prefer Model (b). Reject Model (a).

MATLAB J-Test: Reference Model (b)

Following the same assumptions from before and the same Models (a) and (b):

$$\text{(a)} \quad Y = a_1 + a_2 \log(X_1) + u_i$$

$$\text{(b)} \quad Y = b_1 + b_2 X_1 + e_i$$

Use Model (b) as the reference model.

$$\text{(d)} \quad \hat{Y} = \hat{b}_1 + \hat{b}_2 X_1 + \hat{b}_3 \hat{Y}_a + e_i$$

- H_0: Model (b) ($b_3 = 0$)
- H_a: Not Model (b) ($b_3 \neq 0$)

If b_4 is statistically significant, we prefer Model (a). Reject Model (b).

Encompassing Model

Consider the following two models:

1. $Y_i = \beta_1 + \beta_2 X_{1i} + \beta_3 X_{2i} + \epsilon_i$
2. $Y_i = \alpha_1 + \alpha_2 \log(X_{1i}) + \alpha_3 \log(X_{2i}) + \mu_i$

Method: combine two non-nested models into one testable hybrid model (Model 3).

3. $\hat{Y} = \hat{\gamma}_1 + \hat{\gamma}_2 X_{1i} + \hat{\gamma}_3 X_{2i} + \hat{\gamma}_4 \log(X_{1i}) + \hat{\gamma}_5 \log(X_{2i}) + \epsilon_i$

Where H_0 : Model 1 (i.e., $\gamma_4 = \gamma_5 = 0$) and H_a : Hybrid Model

- If F-test is significant on γ_4 and γ_5, then we reject Model 1, and conclude that keeping the log form in the model is important.