

J-Tests: To Nest Or Not To Nest, That Is The Question

by

Dale S. Bremmer

Professor of Economics, Rose-Hulman Institute of Technology

**To be presented at the “Quantitative Methods” Session of the
79th Annual Conference of the Western Economics Association International
in Vancouver, British Columbia, Canada, on July 3, 2003**

J-Tests: To Nest Or Not To Nest, That Is The Question

I. INTRODUCTION

Allegations of specification errors can be a kiss of death for applied econometricians.

Accusations of specification errors are easily made and innocence is often hard to prove, especially when economic theory is inconclusive. Specification errors are serious because if a regression model omits relevant explanatory variables or if it uses the wrong functional form, its estimates are biased and inconsistent. Equally troubling, if a misspecified model is used for hypotheses testing and statistical inference, misleading results may occur. Statistical specification tests may detect a specification error or they may show that fears of misspecification are groundless. This paper looks at two specification tests - - Davidson and MacKinnon's (1981) non-nested J test and the nested, hybrid model specification test.

The J test is commonly used in the literature. According to the Social Science Citation Index, the Davidson and MacKinnon paper that introduced the J test has been cited in 497 separate articles between 1984 and 2004.¹ Another sign of the increased acceptance of using J tests in econometric practice is the number of textbooks that discuss this specification test.² Some textbooks cite reasons why the J test is superior to the nested specification test. However, this paper shows there is a special case when the non-nested J test and the nested test give identical answers.

Choosing Between Two Competing Specifications

Both the non-nested J test and the nested specification test can be used when there is a choice between two competing model specifications. Suppose that both Model A and Model B are theoretically plausible in describing the behavior of an endogenous variable. Model A states that

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ji} + \sum_{j=1}^m \delta_j w_{ji} + \varepsilon_i, i = 1, \dots, N \quad (1)$$

¹ Citation of Davidson and MacKinnon's 1981 paper has increased over time. In his 1995 article, McAleer noted that the initial paper about the J test had only been cited 71 times between 1981 and 1995. Now almost ten years later, the number of citations has increased sevenfold.

² See Gujarati (2003, 531-536), Johnston and DiNardo (1997, 280-282), Judge, et al (1985, 884-885), Kennedy (1998, 79, 89-90, 92-93), Kmenta (1986, 593-598), Maddala (1988, 443-447), Thomas (1997, 359-360), and Wooldridge (2003, 294-295).

while the model specification Model B, which is also theoretically possible, is

$$y_i = \alpha_0 + \sum_{j=1}^k \alpha_j x_{ji} + \sum_{j=1}^n \gamma_j z_{ji} + \mu_i, i = 1, \dots, N. \quad (2)$$

In models specified in equations (1) and (2), there are N observations, and both Models A and B have the same dependent variable, y . Referring to Model A, each observation i ($i = 1, \dots, N$) has $k + m$ explanatory variables ($x_{ji}, j = 1, \dots, k$, and $w_{ji}, j = 1, \dots, m$) and there are $m + k + 1$ regression parameters ($\beta_0, \dots, \beta_k, \delta_1, \dots, \delta_m$) including the intercept (β_0). On the other hand, Model B has $k + n$ explanatory variables ($x_{ji}, j = 1, \dots, k$, and $z_{ji}, j = 1, \dots, n$). Including the intercept (α_0), Model B has $k + n + 1$ unknown regression parameters ($\alpha_i, i = 0, \dots, k$, and $\gamma_i, i = 1, \dots, n$).

Both models can have the same subset of explanatory variables. For every observation, both Models A and B in the above equations include the same k explanatory variables, $x_j, j = 1, \dots, k$. However, because theory is inconclusive on what other explanatory variables belong in the model, Model A includes m more explanatory variables, w_j ($j = 1, \dots, m$), while Model B includes a different set of n explanatory variables, $z_j, j = 1, \dots, n$. While each of these different subsets of explanatory variables are unique, each set is theoretically plausible, resulting in uncertainty over which model is the correct model specification.

The First Specification Test: The Non-Nested J Test

Ways to empirically determine the correct model specification include performing a nested specification test and a non-nested J test. Table 1 reviews the necessary steps to perform the non-nested J-Test. As Table 1 shows, the J test requires estimating four different regression equations. In Step 1, Model A is estimated and the predicted values of the dependent variable, $\hat{y}_i^A, i = 1, \dots, N$, are obtained. Step 2 is analogous to Step 1 in that Model B is estimated and the predicted values from this model, $\hat{y}_i^B, i = 1, \dots, N$, are also derived. In Step 3, the predicted values from Model B are included as an explanatory variable in Model A, while the predicted values from Model A are included in Model B's

specification in Step 4. The critical idea is whether the predicted values from one model add significant explanatory power to the other model.

As Table 1 indicates, the non-nested J test has four possible outcomes. Two of the outcomes give definitive results regarding the best specification of the model, but the other two outcomes yield inconclusive results. First, the non-nested J test may fail to reject Model A while simultaneously rejecting Model B. Second, the converse may happen where Model B is not rejected but Model A is rejected. In these cases, the predicted values of the superior model are statistically significant when added as an additional explanatory variable to the specification of the rejected model. However, when the predicted values from the rejected model are included as another explanatory variable in the superior model, the regression slope coefficient associated with the predicted values is not statistically different from zero. The inconclusive results occur when either the J test simultaneously rejects or fails to reject both models.

The Other Specification Test: The Nested Test

While the non-nested J test requires estimation of four different regressions, the nested specification test requires the estimation of two regressions. The nested specification test is essentially an F test which determines whether several slope parameters are simultaneously equal to zero. Table 2 lists the steps for this nested test.

As indicated in Table 2, this statistical test requires the estimation of two regressions. First, Model A is estimated, and then a hybrid or nested model that includes all the explanatory variables in Models A and B is estimated. Regarding the nested model, attention is directed towards the slope coefficients of those explanatory variables that are in Model B and not included in Model A. The null hypothesis is that all these slope coefficients are simultaneously equal to zero. If this nested F test fails to reject this null hypothesis, then the exogenous variables that are unique to Model B offer no additional explanatory power, and Model A is chosen over Model B. If the null hypothesis is rejected,

then the variables that are unique to Model B add important explanatory power to Model A, indicating that Model A might be incorrectly specified.

Which Specification Test Is Best?

In terms of application, the nested, hybrid specification test might be preferred to the non-nested J test because it requires estimation of only two rather than four regressions. Another problem is the possibility that the J test might yield inconclusive results. Finally, the J test is a large sample test where the critical test statistic has the proper statistical distribution only asymptotically. Gujarati (2003, p. 534) notes that in small samples “the J test may not be very powerful . . . because it tends to reject the true hypothesis or model more frequently than it ought to.”

However, the nested, hybrid specification test also has its problems. With the inclusion of every possible explanatory variable, Fomby, Hill, and Johnson (1984, p.416) note that “multicollinearity problems are a distinct possibility with the coefficients being imprecisely estimated.” Gujarati (2003, p. 531) states that “the artificially nested model . . . may not have any economic meaning.” There is a chance that the hybrid model generates asymmetric results depending on the choice of the reference model. Model A was the reference model in Table 2 because the statistical test investigated whether the variables unique to Model B added any explanatory power to Model A. Conversely, Model B could have been chosen as the reference model and then the hybrid model could have been used to test whether those exogenous variables that are unique to Model A add any statistically significant explanatory power to Model B. Fomby, Hill, and Johnson (1984, p. 416) conclude that

The choice of the reference model could determine the outcome of the choice of the model. Thus, though the mechanical nesting procedure has great appeal, it does incur statistical problems.

However, this paper shows that there is a special case when the non-nested J test and the nested, hybrid specification test give equivalent results. This case occurs when there is debate over the inclusion of only one independent variable. Assume that both Model A and Model B agree on including the same subset of explanatory variables. Both models assume there are k explanatory variables and the

first $k - 1$ explanatory variables in both models are identical. However, disagreement occurs over the k^{th} exogenous variable, the last exogenous variable included in the model. Model A assumes the last explanatory variable is variable w , while Model B assumes the last explanatory variable is z . This paper shows that when there is only debate over one exogenous variable, either w or z , the non-nested J test and the nested, hybrid model test give identical results. In this case, the nested, hybrid specification test may be preferred as it reveals all the same information while running three less regressions.

The next section of the paper provides a simple, empirical example showing a case where the two specification tests give equivalent results. The paper's third section introduces the two models and the notation. The paper's fourth section lists three propositions about the statistical properties of this special case and uses matrix algebra to prove the first proposition, which is the main result. A summary and concluding thoughts are in the fifth and final section of the paper.

II. A NUMERICAL EXAMPLE

To illustrate a case where both the non-nested J test and the nested, hybrid specification test give the same answers, the following section uses a simple, empirical model of the Phillip's curve to motivate the discussion. The empirical results presented here are not novel or path-breaking. Though the model and its empirical results add nothing to the literature, the intent of this section is to show an empirical example of a case where both specification tests yield test statistics with the same absolute value.

The Correct Functional Form for the Phillip's Curve

The simple Phillip's curve examines the short-run trade-off between inflation (π) and unemployment (U). Usually there is an inverse relationship between these two variables, and in the past, policy makers have taken steps to reduce unemployment while accepting the short-run cost of higher inflation. Suppose two different, but both theoretically plausible, empirical models of the Phillip's curve are proposed. While the correct functional form is uncertain, assume that current inflation in time t (π_t) is a function of two variables: inflation in the previous period (π_{t-1}) and current unemployment (U_t).

The specification for the first model, Model A, is

$$\pi_t = \beta_0 + \beta_1\pi_{t-1} + \beta_2U_t + \varepsilon_t \quad (3)$$

and Model B's specification is

$$\pi_t = \alpha_0 + \alpha_1\pi_{t-1} + \alpha_2\frac{1}{U_t} + \mu_t \quad (4)$$

The difference in the specification of the two models is easy to spot. In Model A, current unemployment (U_t) enters the specification linearly while the same variable enters the specification in Model B in a reciprocal fashion ($\frac{1}{U_t}$). Since inflation is a time series exhibiting some inertia, both β_1 and α_1 are expected to be positive. If there is an inverse, short-run relationship between current inflation and current unemployment, then β_2 is expected to be negative while α_2 should be positive.

The Two Specification Tests

To determine the best model specification with the correct functional form, the J-test requires the estimation of the models listed in equations (3) and (4). Let the predicted values of inflation from Models A and B be denoted as $\hat{\pi}_t^A$ and $\hat{\pi}_t^B$, respectively. Performing the J test requires including the predicted values from Model A as an additional explanatory variable in Model B and the predicted values from Model B are also included as an additional right-side variable in the specification of Model A. After estimating the two models in equations (3) and (4), the J test requires the estimation of still two more additional equations:

$$\pi_t = \delta_0 + \delta_1\pi_{t-1} + \delta_2U_t + \delta_3\hat{\pi}_t^B + e_{1t} \quad (5)$$

and

$$\pi_t = \gamma_0 + \gamma_1\pi_{t-1} + \gamma_2\frac{1}{U_t} + \gamma_3\hat{\pi}_t^A + e_{2t} \quad (6)$$

The outcome of the J test rests on the statistical significance of the ordinary least squares (OLS) estimates of δ_3 and γ_3 . In the search for the proper functional form, hopefully only one of these

estimates is statistically different from zero while the other estimate is not. The J test will be inconclusive if both of these estimates are either simultaneously statistically different from zero or not.

To perform the other specification test using the nested, hybrid model, the following regression is estimated

$$\pi_t = \theta_0 + \theta_1\pi_{t-1} + \theta_2U_t + \theta_3\frac{1}{U_t} + e_{3t} . \quad (7)$$

Here the choice between Model A and Model B depends on the statistical significance of the estimates of θ_2 and θ_3 . A definitive choice between the two models can be made when one of these estimates is statistically different from zero and the other estimate is statistically insignificant. Like the J test, choice of the proper functional form becomes more complicated when the nested specification test yields inconclusive results when θ_2 and θ_3 are both simultaneously statistically different from zero or both simultaneously statistically insignificant.

The Data and the Empirical Results

The estimation results for all the regressions described in equations (3) through (7) are listed in Table 3. Both the consumer price index (CPI) data and the unemployment rate data are from the Bureau of Labor Statistics. The annual inflation rate was derived from the CPI data for all urban consumers (1967 = 100) and the unemployment rate is the seasonally unadjusted value reported for those unemployed workers who are 16 years of age or older.³ Annual observations of these variables are available from 1949 to 2003. Since the Chow test indicates there is parameter instability at the breakpoint of 1974, the regressions in Table 3 are based on the 30 annual observations between 1974 and 2003.⁴

Looking at the estimation results for Models A and B in Table 3, all the estimated slope

³ The data is easily obtained from the Bureau of Labor Statistics web site at <http://www.bls.gov>. An Excel spreadsheet of the data is available upon request.

⁴ The F-statistic of the Chow test was 3.51 with a p-value of 0.022, implying the null hypothesis of parameter stability could be rejected at the 5 percent level.

coefficients are statistically significant at the 10 percent level or better and the coefficients have the expected signs. The models have adequate explanatory power as the R^2 is either 71 or 75 percent. Even though the data set consists of annual time series, neither regression appears to be plagued by serial correlation. In the case of Model A, the Breusch-Godfrey serial correlation test indicates there is no first-order autocorrelation with an insignificant F statistic of 1.25 and a p-value of 0.27. Nor is first-order autocorrelation present in Model B as the Breusch-Godfrey autocorrelation test has an insignificant F statistic of 2.34 with a p-value of 0.14.

Determining The Correct Functional Form

While both models have significant slope coefficients with correct sign and other desirable statistical results, the specification tests reported in Table 3 indicate that Model A is the superior model. First refer to the J test results reported in column labeled “Test 1.” The null hypothesis that the coefficient associated with the predicted values from Model B is equal to zero cannot be rejected at the one percent level of significance. Therefore, the predicted values of inflation from Model B fail to add any meaningful explanatory power to Model A, indicating the superiority of Model A’s specification. On the other hand, referring to the column labeled “Test 2” in Table 3, the hypothesis that the predicted values of inflation from Model A fail to add any explanatory power to Model B is rejected at the 1 percent level. Since the predicted values of Model A improve the performance of Model B, and the converse is not true, given a relatively strict level of significance set at one percent, the J tests indicate that Model B should be rejected in favor of Model A.

The nested specification test agrees with the results of the J test. Referring to the column labeled “Test 3,” the coefficient associated with the unemployment rate equals -2.32 and it is statistically different from zero at the 1 percent level. But the coefficient associated with the reciprocal of the unemployment rate equals -58.63, which is statistically different from zero only at the 10 percent level. Given a stringent significance level equal to one percent, then the conclusion is that the unemployment rate enters the model specification in a linear fashion rather than as a reciprocal.

The t statistics from both specification tests have the same absolute value

The purpose of the empirical example reported in Table 3 is to show how the J test and the nested, hybrid model specification test yield the same results when the two models in question differ by only one variable. Referring to Table 3, examine the three columns labeled “Test 1,” “Test 2” and “Test 3.” Looking at the “Test 1” column, when the predicted values of Model B are included as another explanatory variable in Model A, the t statistic associated with the coefficient is equal to -1.98. This is exactly the same t statistic associated with the reciprocal of the unemployment rate in the nested specification test reported in the column labeled “Test 3.” Likewise, the absolute value of the t statistic associated with the predicted values of Model A in the “Test 2” column and the absolute value of the t statistic accompanying the linear unemployment rate in the “Test 3” column are exactly the same value, 2.92.

Based on this example, the generalization becomes when two model specifications differ by one variable, the nested specification test and the J test produce the same results as the test statistics have the same absolute values. In the nested, hybrid model, the absolute value of the t statistic of one of the exogenous variables in question equals the absolute value of the t statistic of the predicted values from that variable’s model in the J test. This result will be proven in the fourth section of the paper. When two models differ only by one variable, one can save time performing specification tests. The single nested, hybrid model, obtained by running only one regression, will obtain the exact same results of the J test without having to estimate the four regressions needed to perform such a test.

The explanatory variables in both tests have the same slope coefficients

The estimated slope coefficient for the linear unemployment rate in the regression performing the J test in the “Test 1” column has the same value as the slope coefficient of the linear unemployment rate in the hybrid model in the “Test 3” column. Both of these coefficients equal -2.32. Analogously, the slope coefficient associated with the reciprocal of the unemployment rate equals -58.63 in the columns labeled as “Test 2” and “Test 3.” Since these coefficients have the same standard errors, the

simple t statistic testing whether the coefficients are equal to zero are also identical. The generalization is that when comparing the non-nested J test with the nested specification test, the coefficients, standard errors, and t statistics associated with a given explanatory variable are the same in both regressions.

Both tests have identical regression diagnostic statistics

It is also noteworthy that other diagnostic regression statistics of the two J test regressions and the nested, hybrid model are also identical. For example, the R²s, the adjusted R²s, the F test statistics, the Durbin-Watson tests, and the rest of the other statistics reported in the bottom half of Table 3 are the same in the columns labeled “Test 1,” “Test 2” and “Test 3.” Table 3 also hints why these results occur. Examining the last three columns in Table 3, all these three regressions have the identical residual sum of squares, 65.36. Based on this result, one can infer that the regressions reported in the last three columns of Table 3 have identical residuals for a given observation.

III. MODEL SPECIFICATION: NULL AND ALTERNATIVE HYPOTHESES

Suppose a researcher has to choose between two competing model specifications and economic theory is inconclusive on which model specification is correct. The key assumption is that the two specifications differ only in the choice of one explanatory variable. The first model specification, the null hypothesis, is

$$H_0: \mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{e}_0 = \mathbf{X}\boldsymbol{\psi} + \delta\mathbf{w} + \mathbf{e}_0 \tag{8}$$

while the competing model specification or the alternative hypothesis is

$$H_A: \mathbf{y} = \mathbf{X}_A\boldsymbol{\beta}_A + \mathbf{e}_A = \mathbf{X}\boldsymbol{\alpha} + \gamma\mathbf{z} + \mathbf{e}_A \tag{9}$$

Given N observations, \mathbf{y} is the N x 1 vector of observations on the dependent variable. The N x 1 vector of residuals under the null hypothesis is \mathbf{e}_0 , while \mathbf{e}_A is the corresponding N x 1 vector of residuals under the alternative hypothesis.

Referring to equation (8), the explanatory variables under the null hypothesis are found in matrix \mathbf{X}_0 with a dimension of N x (k+1). Likewise, the N x (k+1) data matrix under the alternative hypothesis

is \mathbf{X}_A . The first column of both \mathbf{X}_0 and \mathbf{X}_A is a vector of ones because both models include an intercept. The vector of unknown regression parameters under the null hypothesis is $\boldsymbol{\beta}_0$ and, under the alternative hypothesis, $\boldsymbol{\beta}_A$ is the vector of unknown parameters to be estimated. Both $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_A$ are dimensioned $(k+1) \times 1$, and the first element in each vector is the intercept.

As equations (8) and (9) indicate, if both \mathbf{X}_0 and \mathbf{X}_A are partitioned into a submatrix consisting of the first k columns and a vector consisting of the last column, they both have a common submatrix, \mathbf{X} , which is dimensioned $N \times k$. This follows from the assumptions that, under the null or the alternative hypotheses, the first $k - 1$ explanatory variables in both models are the same. The first column of \mathbf{X} consists of a vector of ones because an intercept is included in both models and the remaining $k - 1$ columns of \mathbf{X} are vectors of the observations on the explanatory variables that both models have in common.

Finally, equations (8) and (9) highlight the critical difference between the model's specification under the null and alternative hypotheses. While the first $k - 1$ explanatory variables in both model specifications are the same, the models differ over the selection of the last, k^{th} explanatory variable. Under the null hypothesis, the last explanatory variable consists of the observations in vector \mathbf{w} , while vector \mathbf{z} captures a different explanatory variable under the alternative hypothesis. Vector \mathbf{w} has a dimension of $N \times 1$, and it is the last column of \mathbf{X}_0 , the data matrix under the null hypothesis. However, the last column in \mathbf{X}_A , the data matrix associated with the alternative hypothesis, consists of the vector \mathbf{z} which is also $N \times 1$.

Given the null hypothesis, the vector of the OLS estimates is $\hat{\boldsymbol{\beta}}_0 = (\mathbf{X}_0^T \mathbf{X}_0)^{-1} \mathbf{X}_0^T \mathbf{y}$ where $\hat{\boldsymbol{\beta}}_0$ is a $(k + 1) \times 1$ column vector. Likewise, under the alternative hypothesis, the $(k + 1) \times 1$ vector of OLS estimates is $\hat{\boldsymbol{\beta}}_A = (\mathbf{X}_A^T \mathbf{X}_A)^{-1} \mathbf{X}_A^T \mathbf{y}$.

Model Specification Tests

To help determine the correct model specification, two statistical tests can be performed. The

first statistical test consists of estimating a nested model that includes both of the variables being considered, \mathbf{w} and \mathbf{z} , as explanatory variables. Performing the nested test involves estimating a single regression equation. The second procedure, Davidson and MacKinnon's J test, is a non-nested statistical test that involves estimating four different regression models.

The nested hypothesis test

The regression testing the nested hypothesis is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\lambda} + \theta_1\mathbf{w} + \theta_2\mathbf{z} + \mathbf{e}_N \quad (10)$$

where $\boldsymbol{\lambda}$ is a $k \times 1$ vector of regression parameters to be estimated. This vector includes the intercept and the slope coefficients for the $k - 1$ explanatory variables that are common in both models. Additional regression slope parameters include both θ_1 and θ_2 which are 1×1 scalars that need to be estimated. Finally, \mathbf{e}_N is the $N \times 1$ vector of unobserved residuals associated with the model specified in equation (10).

The non-nested J test

While the nested test requires the estimation of only one regression model, the non-nested, J test requires estimating four regressions. First, returning to null hypothesis in equation (8), the regression model is estimated and the $N \times 1$ vector of the predicted values of \mathbf{y} , $\hat{\mathbf{y}}_0$ is obtained. Second, the regression model under the alternative hypothesis in equation (9) is estimated and the $N \times 1$ vector of predicted values from this regression, $\hat{\mathbf{y}}_A$, is also retrieved.

Third, the predicted values under the alternative hypothesis are added as another explanatory variable in the model specification under the null hypothesis and the following model is estimated:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \phi_1\mathbf{w} + \phi_2\hat{\mathbf{y}}_A + \mathbf{e}_1 \quad (11)$$

The vector $\boldsymbol{\tau}$ is a $k \times 1$ vector of the regression parameters corresponding to the intercept and the $k - 1$ explanatory variables that both the specifications have in common. The regression parameters ϕ_1 and ϕ_2 are 1×1 scalars with ϕ_1 denoting the slope coefficient associated with the explanatory variable \mathbf{w} in

equation (11) and φ_2 is the slope coefficient corresponding to the predicted values from the model based on the alternative hypothesis. As usual, \mathbf{e}_1 is the $N \times 1$ residual vector of white noise.

Fourth and finally, predicted values found from estimating the model under the null hypothesis, $\hat{\mathbf{y}}_0$, are added as an additional explanatory variable to the model specified under the alternative hypothesis. This requires estimating the regression

$$\mathbf{y} = \mathbf{X}\boldsymbol{\kappa} + \eta_1\mathbf{z} + \eta_2\hat{\mathbf{y}}_0 + \mathbf{e}_2 \quad (12)$$

where $\boldsymbol{\kappa}$ is the $k \times 1$ vector of unknown regression parameters associated with the model's intercept and the slope coefficients of the $k - 1$ explanatory variables that are common in both models. The $N \times 1$ vector of regression residuals is \mathbf{e}_2 . Both η_1 and η_2 are unknown 1×1 scalars. The parameter η_1 is the unknown slope coefficient associated with the vector \mathbf{w} and η_2 is the unknown slope parameter associated with the predicted values obtained from estimating the model under the null hypothesis, $\hat{\mathbf{y}}_0$.

IV. WHEN THE NESTED AND NON-NESTED TESTS ARE EQUIVALENT

The two econometric models presented in the previous section of the paper are assumed to be identical except for one explanatory variable. Given this assumption, both the non-nested and the nested specification tests give equivalent results with the same test statistics, at least in absolute value. Three propositions about equations (10), (11) and (12) can be proven. These equations are replicated below for the reader's convenience.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\lambda} + \theta_1\mathbf{w} + \theta_2\mathbf{z} + \mathbf{e}_N \quad (10)$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \varphi_1\mathbf{w} + \varphi_2\hat{\mathbf{y}}_A + \mathbf{e}_1 \quad (11)$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\kappa} + \eta_1\mathbf{z} + \eta_2\hat{\mathbf{y}}_0 + \mathbf{e}_2 \quad (12)$$

In the propositions that follow, the following notation is used. Let $\hat{\theta}_2$ denote the OLS estimate for θ_2 in equation (10). If $se(\hat{\theta}_2)$ is the standard error of $\hat{\theta}_2$, then the simple t test used to determine whether $\hat{\theta}_2$ is statistically different from zero has a test statistic equal to $t_{\hat{\theta}_2}$ where

$$t_{\hat{\theta}_2} = \frac{\hat{\theta}_2}{\text{se}(\hat{\theta}_2)} .$$

Similar notation is used for $\hat{\theta}_1$, $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\eta}_1$, and $\hat{\eta}_2$. The three propositions are listed below.

Proposition #1: Given the assumptions of the null and alternative hypotheses, then $|t_{\hat{\theta}_2}| = |t_{\hat{\phi}_2}|$ and $|t_{\hat{\theta}_1}| = |t_{\hat{\eta}_2}|$. Thus, the nested and non-nested specification tests yield the same test statistics in absolute value.

Proposition #2: Given the assumptions of the null and alternative hypotheses, $\hat{\theta}_1 = \hat{\phi}_1$, $\text{se}(\hat{\theta}_1) = \text{se}(\hat{\phi}_1)$ and $t_{\hat{\theta}_1} = t_{\hat{\phi}_1}$. Likewise, $\hat{\eta}_1 = \hat{\phi}_2$, $\text{se}(\hat{\eta}_1) = \text{se}(\hat{\phi}_2)$ and $t_{\hat{\eta}_1} = t_{\hat{\phi}_2}$. Thus, the coefficients of the debated explanatory variables in both the non-nested and the nested regression have the same estimates, standard errors, and t statistics.

Proposition #3: Given the assumptions of the null and alternative hypotheses, each observation of equations (10), (11) and (12) have the same residual. In other words, $\hat{\mathbf{e}}_N = \hat{\mathbf{e}}_1 = \hat{\mathbf{e}}_2$. Thus, the non-nested and nested specification tests have the same set of regression diagnostic statistics.

The first proposition, the key result of the paper, is proven in the next section.⁵ To prove the second proposition, mimic the steps and techniques used to prove the first proposition. The third proposition follows from the first two results.

V. PROVING PROPOSITION #1

The discussion that follows uses simple matrix algebra to prove the first proposition. The key assumption is that two theoretically plausible model specifications differ only by one explanatory variable. In this special case, the test statistic derived by the non-nested J test and the test statistic derived by the nest specification test have the same absolute value.

First: The Nested Specification in Equation (10)

To prove Proposition #1, the formulas for the OLS estimates of θ_2 and ϕ_2 are derived, as are formulas for the standard errors of these estimates. Given the formula of the estimates and standard

⁵ A note describing the proof of the other two propositions is available upon request.

errors of these parameters, proving the first proposition rests on demonstrating that the t statistics associated with each of the estimates have the same absolute value.

The strategy to prove these results will be mimicked at several steps. First, the OLS estimates for equations (10) and (11) are derived by partitioning the matrices involved, taking the inverse of these partitioned matrices, performing the necessary matrix multiplication, and then simplifying the expressions. Showing that the t statistics of two different equations have the same absolute value also relies on the properties of idempotent and symmetric matrices. Attention is first directed towards the nested model specification test in equation (11).

The OLS estimate for θ_2 , the coefficient of z in equation (10)

Examining the regression with the nested hypothesis first, the data matrix and the vector of parameters can be partitioned so that

$$y = \mathbf{X}\lambda + \theta_1\mathbf{w} + \theta_2\mathbf{z} + \mathbf{e}_N = [\mathbf{X} \quad \mathbf{w} \quad \mathbf{z}] \begin{bmatrix} \lambda \\ \theta_1 \\ \theta_2 \end{bmatrix} + \mathbf{e}_N = [\mathbf{X}_1 \quad \mathbf{z}] \begin{bmatrix} \beta_1 \\ \theta_2 \end{bmatrix} + \mathbf{e}_N \quad (13)$$

where \mathbf{X}_1 is a partitioned matrix and β_1 is a partitioned vector so that

$$\mathbf{X}_1 = [\mathbf{X} \quad \mathbf{w}] \text{ and } \beta_1 = \begin{bmatrix} \lambda \\ \theta_1 \end{bmatrix}. \quad (14)$$

The OLS estimates of the intercept and slope coefficients are equal to

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \left\{ \begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{z}^T \end{bmatrix} [\mathbf{X}_1 \quad \mathbf{z}] \right\}^{-1} \begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{z}^T \end{bmatrix} \mathbf{y} = \begin{bmatrix} \mathbf{X}_1^T \mathbf{X}_1 & \mathbf{X}_1^T \mathbf{z} \\ \mathbf{z}^T \mathbf{X}_1 & \mathbf{z}^T \mathbf{z} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}_1^T \mathbf{y} \\ \mathbf{z}^T \mathbf{y} \end{bmatrix}.$$

Using the well-known formula for the inverse of a partitioned matrix⁶, the above equation can be simplified to obtain

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \begin{bmatrix} (\mathbf{X}_1^T \mathbf{X}_1)^{-1} + b_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{z} \mathbf{z}^T \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} & -b_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{z} \\ -b_1 \mathbf{z}^T \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} & b_1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_1^T \mathbf{y} \\ \mathbf{z}^T \mathbf{y} \end{bmatrix} \quad (15)$$

⁶ See, for example, Johnston and DiNardo (1997, 472 – 473).

where b_1 is a 1×1 scalar such that

$$b_1 = \left\{ \mathbf{z}^T \left[\mathbf{I} - \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \right] \mathbf{z} \right\}^{-1} = (\mathbf{z}^T \mathbf{M}_1 \mathbf{z})^{-1} . \quad (16)$$

Referring to equation (16), $\mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T$ where \mathbf{I} is the identity matrix and \mathbf{M}_1 is a symmetric, idempotent matrix. After performing the matrix multiplication shown in equation (15) and simplifying, it can be shown that the OLS estimate of θ_2 equals

$$\hat{\theta}_2 = (\mathbf{z}^T \mathbf{M}_1 \mathbf{z})^{-1} \mathbf{z}^T \mathbf{M}_1 \mathbf{y} . \quad (17)$$

Since \mathbf{M}_1 is symmetric and idempotent, and equivalent expression for equation (17) is

$$\hat{\theta}_2 = [(\mathbf{M}_1 \mathbf{z})^T \mathbf{M}_1 \mathbf{z}]^{-1} (\mathbf{M}_1 \mathbf{z})^T \mathbf{M}_1 \mathbf{y} . \quad (18)$$

The estimate for $\hat{\sigma}_N^2$ the variance of the residuals in equation (10)

If $\hat{\mathbf{e}}_N$ is the vector of residuals obtained from estimating equation (10), then the variance of these residuals, $\hat{\sigma}_N^2$, is equal to

$$\hat{\sigma}_N^2 = \frac{\hat{\mathbf{e}}_N^T \hat{\mathbf{e}}_N}{N - k - 2} = \frac{\mathbf{y}^T \mathbf{M}_N \mathbf{y}}{N - k - 2} . \quad (19)$$

In this case, \mathbf{M}_N is the well-known and familiar idempotent and symmetric matrix associated with the regression in equation (10) and

$$\mathbf{M}_N = \mathbf{I} - \mathbf{X}_0 (\mathbf{X}_0^T \mathbf{X}_0)^{-1} \mathbf{X}_0^T = \mathbf{I} - \begin{bmatrix} \mathbf{X}_1 & \mathbf{z} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1^T \mathbf{X}_1 & \mathbf{X}_1^T \mathbf{z} \\ \mathbf{z}^T \mathbf{X}_1 & \mathbf{z}^T \mathbf{z} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{z}^T \end{bmatrix} . \quad (20)$$

Again using the formula for the inverse of a partitioned matrix, \mathbf{M}_N in equation (20) can be written as

$$\mathbf{M}_N = \mathbf{I} - \begin{bmatrix} \mathbf{X}_1 & \mathbf{z} \end{bmatrix} \begin{bmatrix} (\mathbf{X}_1^T \mathbf{X}_1)^{-1} + b_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{z} \mathbf{z}^T \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} & -b_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{z} \\ -b_1 \mathbf{z}^T \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} & b_1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{z}^T \end{bmatrix} \quad (21)$$

where b_1 is still defined as it was in equation (16) or $b_1 = (\mathbf{z}^T \mathbf{M}_1 \mathbf{z})^{-1}$. After performing the indicated matrix multiplication in equation (21) and simplifying the results, the relationship between \mathbf{M}_N and \mathbf{M}_1 becomes

$$\mathbf{M}_N = \mathbf{M}_1 - \mathbf{b}_1 \mathbf{M}_1 \mathbf{z} \mathbf{z}^T \mathbf{M}_1 \quad . \quad (22)$$

Therefore, equation (19) can be expanded and the OLS estimate of the variance of the errors terms in equation (10) becomes

$$\hat{\sigma}_N^2 = \frac{\mathbf{y}^T [\mathbf{M}_1 - (\mathbf{z}^T \mathbf{M}_1 \mathbf{z})^{-1} \mathbf{M}_1 \mathbf{z} \mathbf{z}^T \mathbf{M}_1] \mathbf{y}}{N - k - 2}$$

or, using the fact that \mathbf{M}_1 is idempotent,

$$\hat{\sigma}_N^2 = \frac{\mathbf{y}^T \left\{ \mathbf{M}_1 - [(\mathbf{M}_1 \mathbf{z})^T (\mathbf{M}_1 \mathbf{z})]^{-1} (\mathbf{M}_1 \mathbf{z})(\mathbf{M}_1 \mathbf{z})^T \right\} \mathbf{y}}{N - k - 2} \quad . \quad (23)$$

The standard error of $\hat{\theta}_2$ or $se(\hat{\theta}_2)$

To construct the simple t statistic for $\hat{\theta}_2$, the standard error of $\hat{\theta}_2$ is need. The variance of $\hat{\theta}_2$ equals $\hat{\sigma}_N^2$ multiplied by $b_1 = (\mathbf{z}^T \mathbf{M}_1 \mathbf{z})^{-1}$. This follows because b_1 is the element appearing in row $k + 2$ and column $k + 2$ of the first matrix on the right-hand side of the equal sign in equation (15).

Therefore, the standard error of $\hat{\theta}_2$, denoted by $se(\hat{\theta}_2)$, is equal to the square root of its variance or

$$se(\hat{\theta}_2) = \hat{\sigma}_N \sqrt{(\mathbf{z}^T \mathbf{M}_1 \mathbf{z})^{-1}} = \hat{\sigma}_N \sqrt{[(\mathbf{M}_1 \mathbf{z})^T \mathbf{M}_1 \mathbf{z}]^{-1}} \quad . \quad (24)$$

Again, the last equality on the right-hand side of equation (24) follows from the properties of idempotent matrix \mathbf{M}_1 .

Second: The Non-Nested J Test Specification in Equation (11)

The next major step in proving Proposition #1 is deriving the formula for the OLS estimate of φ_2 in equation (11). In addition, formulas for the standard error of $\hat{\varphi}_2$ and its corresponding t statistic must also be derived. Just as equation (10) was partitioned in to several submatrices, equation (11) can be written as

$$\mathbf{y} = \begin{bmatrix} \mathbf{X} & \mathbf{w} & \hat{\mathbf{y}}_A \end{bmatrix} \begin{bmatrix} \boldsymbol{\tau} \\ \varphi_1 \\ \varphi_2 \end{bmatrix} + \mathbf{e}_1 = \begin{bmatrix} \mathbf{X}_1 & \hat{\mathbf{y}}_A \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_2 \\ \varphi_2 \end{bmatrix} + \mathbf{e}_1 . \quad (25)$$

Recall from equation (14) that \mathbf{X}_1 is a partitioned matrix where $\mathbf{X}_1 = [\mathbf{X} \ \mathbf{w}]$ and similar to $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ is also a partitioned vector in that

$$\boldsymbol{\beta}_2 = \begin{bmatrix} \boldsymbol{\tau} \\ \varphi_1 \end{bmatrix} . \quad (26)$$

The vector of OLS estimates equals

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_2 \\ \hat{\varphi}_2 \end{bmatrix} = \begin{bmatrix} (\mathbf{X}_1^T \mathbf{X}_1)^{-1} + \mathbf{b}_2 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \hat{\mathbf{y}}_A \hat{\mathbf{y}}_A^T \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} & -\mathbf{b}_2 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \hat{\mathbf{y}}_A \\ -\mathbf{b}_2 \hat{\mathbf{y}}_A^T \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} & \mathbf{b}_2 \end{bmatrix} \begin{bmatrix} \mathbf{X}_1^T \mathbf{y} \\ \hat{\mathbf{y}}_A^T \mathbf{y} \end{bmatrix} \quad (27)$$

where $\mathbf{b}_2 = (\hat{\mathbf{y}}_A^T \mathbf{M}_1 \hat{\mathbf{y}}_A)^{-1}$ and \mathbf{M}_1 is as defined before, $\mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T$.

Mimicking the steps performed on equation (10), the OLS estimate for φ_2 equals

$$\hat{\varphi}_2 = [(\mathbf{M}_1 \hat{\mathbf{y}}_A)^T \mathbf{M}_1 \hat{\mathbf{y}}_A]^{-1} (\mathbf{M}_1 \hat{\mathbf{y}}_A)^T \mathbf{M}_1 \mathbf{y} . \quad (28)$$

The result for $\hat{\varphi}_2$ in equation (26) is similar to the formula for $\hat{\theta}_2$ in equation (18). If $\hat{\mathbf{e}}_1$ is the $N \times 1$ vector of residuals for equation (11), then it follows from the previous derivations that the OLS estimate of the variance of the random errors in equation (11) equals

$$\hat{\sigma}_1^2 = \frac{\hat{\mathbf{e}}_1^T \hat{\mathbf{e}}_1}{N - k - 2} = \frac{\mathbf{y}^T [\mathbf{M}_1 - (\hat{\mathbf{y}}_A^T \mathbf{M}_1 \hat{\mathbf{y}}_A)^{-1} \mathbf{M}_1 \hat{\mathbf{y}}_A \hat{\mathbf{y}}_A^T \mathbf{M}_1] \mathbf{y}}{N - k - 2} . \quad (29)$$

Since \mathbf{M}_1 is symmetric and idempotent, this implies $\mathbf{M}_1^2 = \mathbf{M}_1 \mathbf{M}_1 = \mathbf{M}_1$, $\mathbf{M}_1 = \mathbf{M}_1^T$, and equation (29) may be written as

$$\hat{\sigma}_1^2 = \frac{\hat{\mathbf{e}}_1^T \hat{\mathbf{e}}_1}{N - k - 2} = \frac{\mathbf{y}^T \left\{ \mathbf{M}_1 - [(\mathbf{M}_1 \hat{\mathbf{y}}_A)^T (\mathbf{M}_1 \hat{\mathbf{y}}_A)]^{-1} (\mathbf{M}_1 \hat{\mathbf{y}}_A) (\mathbf{M}_1 \hat{\mathbf{y}}_A)^T \right\} \mathbf{y}}{N - k - 2} . \quad (30)$$

Like the standard error of $\hat{\theta}_2$ discussed previously, the standard error of $\hat{\varphi}_2$, $\text{se}(\hat{\varphi}_2)$, equals

$$\text{se}(\hat{\varphi}_2) = \hat{\sigma}_1 \sqrt{\mathbf{b}_2} = \hat{\sigma}_1 \sqrt{(\hat{\mathbf{y}}_A^T \mathbf{M}_1 \hat{\mathbf{y}}_A)^{-1}} = \hat{\sigma}_1 \sqrt{[(\mathbf{M}_1 \hat{\mathbf{y}}_A)^T \mathbf{M}_1 \hat{\mathbf{y}}_A]^{-1}} . \quad (31)$$

Digression: The Properties of the Idempotent Matrix \mathbf{M}_1

To show that $|\hat{\theta}_2 / \text{se}(\hat{\theta}_2)| = |\hat{\phi}_2 / \text{se}(\hat{\phi}_2)|$, the properties of $\mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1(\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T$ must be determined. Since $\mathbf{X}_1 = [\mathbf{X} \ \mathbf{w}]$, then \mathbf{M}_1 is equal to

$$\mathbf{M}_1 = \mathbf{I} - [\mathbf{X} \ \mathbf{w}] \begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{w} \\ \mathbf{w}^T \mathbf{X} & \mathbf{w}^T \mathbf{w} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^T \\ \mathbf{w}^T \end{bmatrix} . \quad (32)$$

Using the formula for the inverse of a partitioned matrix, performing the matrix multiplication and simplifying the results gives the following expression

$$\mathbf{M}_1 = \mathbf{M} - b_3 \mathbf{M} \mathbf{w} \mathbf{w}^T \mathbf{M} . \quad (33)$$

Referring to equation (31), $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ and \mathbf{M} is another idempotent, symmetric matrix. Also in equation (31), b_3 is the bottom, right-hand term of the inverse of the partitioned matrix in equation (32) and $b_3 = (\mathbf{w}^T \mathbf{M} \mathbf{w})^{-1}$.

Proving Proposition #1 rests on showing the relationship between two vectors: $\mathbf{M}_1 \hat{\mathbf{y}}_A$ and $\mathbf{M}_1 \mathbf{z}$.

$\mathbf{M}_1 \mathbf{z}$ and its transpose appears frequently in the equations (18) and (24), the equations for $\hat{\theta}_2$ and $\text{se}(\hat{\theta}_2)$, while $\mathbf{M}_1 \hat{\mathbf{y}}_A$ and its transpose are seen in equations (27) and (29), the formulas for $\hat{\phi}_2$ and $\text{se}(\hat{\phi}_2)$. The product of \mathbf{M}_1 and \mathbf{z} equals

$$\mathbf{M}_1 \mathbf{z} = (\mathbf{M} - b_3 \mathbf{M} \mathbf{w} \mathbf{w}^T \mathbf{M}) \mathbf{z} . \quad (34)$$

Since $\hat{\mathbf{y}}_A = \mathbf{X} \hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\gamma}} \mathbf{z}$, $\mathbf{M}_1 \hat{\mathbf{y}}_A$ may be written as

$$\mathbf{M}_1 \hat{\mathbf{y}}_A = \mathbf{M} \mathbf{X} \hat{\boldsymbol{\alpha}} - b_3 \mathbf{M} \mathbf{w} \mathbf{w}^T \mathbf{M} \mathbf{X} \hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\gamma}} (\mathbf{M} - b_3 \mathbf{M} \mathbf{w} \mathbf{w}^T \mathbf{M}) \mathbf{z} = \hat{\boldsymbol{\gamma}} \mathbf{M}_1 \mathbf{z} . \quad (35)$$

$\mathbf{M}_1 \hat{\mathbf{y}}_A = \hat{\boldsymbol{\gamma}} \mathbf{M}_1 \mathbf{z}$ because \mathbf{M} is idempotent and $\mathbf{M} \mathbf{X} = [\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \mathbf{X} = \mathbf{X} - \mathbf{X} = \mathbf{0}$.

The first key result to demonstrate: $\hat{\sigma}_N^2 = \hat{\sigma}_1^2$

Equation (35) gives the key result needed to prove Proposition #1. The first thing to demonstrate is that $\hat{\sigma}_N^2 = \hat{\sigma}_1^2$. This result is obtained through repeated substitution starting with equation

(30) and the steps are shown below.

$$\begin{aligned}
\hat{\sigma}_1^2 &= \frac{\mathbf{y}^T \left\{ \mathbf{M}_1 - \left[(\mathbf{M}_1 \hat{\mathbf{y}}_A)^T (\mathbf{M}_1 \hat{\mathbf{y}}_A) \right]^{-1} (\mathbf{M}_1 \hat{\mathbf{y}}_A) (\mathbf{M}_1 \hat{\mathbf{y}}_A)^T \right\} \mathbf{y}}{N - k - 2} \\
\hat{\sigma}_1^2 &= \frac{\mathbf{y}^T \left\{ \mathbf{M}_1 - \left[(\hat{\gamma} \mathbf{M}_1 \mathbf{z})^T (\hat{\gamma} \mathbf{M}_1 \mathbf{z}) \right]^{-1} (\hat{\gamma} \mathbf{M}_1 \mathbf{z}) (\hat{\gamma} \mathbf{M}_1 \mathbf{z})^T \right\} \mathbf{y}}{N - k - 2} \\
\hat{\sigma}_1^2 &= \frac{\mathbf{y}^T \left\{ \mathbf{M}_1 - \frac{1}{\hat{\gamma}^2} \hat{\gamma}^2 \left[(\mathbf{M}_1 \mathbf{z})^T (\mathbf{M}_1 \mathbf{z}) \right]^{-1} (\mathbf{M}_1 \mathbf{z}) (\mathbf{M}_1 \mathbf{z})^T \right\} \mathbf{y}}{N - k - 2} \\
\hat{\sigma}_1^2 &= \frac{\mathbf{y}^T \left\{ \mathbf{M}_1 - \left[(\mathbf{M}_1 \mathbf{z})^T (\mathbf{M}_1 \mathbf{z}) \right]^{-1} (\mathbf{M}_1 \mathbf{z}) (\mathbf{M}_1 \mathbf{z})^T \right\} \mathbf{y}}{N - k - 2} = \hat{\sigma}_N^2 \tag{36}
\end{aligned}$$

Equation (36) follows from the equations directly above it because the transpose of the 1 x 1 scalar, $\hat{\gamma}$, equals itself and the inverse of the 1 x 1 scalar $\hat{\gamma}^2$ is its reciprocal or $\frac{1}{\hat{\gamma}^2}$.

Thus, the OLS estimate of the variance of the residuals in equations (10) and (11) is the same or $\hat{\sigma}_N^2 = \hat{\sigma}_1^2$. Let the OLS estimate of the variance of the residuals in equation (12) equal $\hat{\sigma}_2^2$. Since choice of the null hypothesis and the alternative hypothesis is arbitrary (the original null hypothesis can become the new alternative hypothesis while the initial alternative hypothesis is now denoted as the new null hypothesis), variables can be relabeled, and therefore $\hat{\sigma}_N^2 = \hat{\sigma}_1^2 = \hat{\sigma}_2^2$.

The second key result to demonstrate: $\text{se}(\hat{\phi}_2) = \text{se}(\hat{\theta}_2)$

Given that $\hat{\sigma}_N^2 = \hat{\sigma}_1^2$ and $\mathbf{M}_1 \hat{\mathbf{y}}_A = \hat{\gamma} \mathbf{M}_1 \mathbf{z}$, it is now easy to show the relationship between $\text{se}(\hat{\theta}_2)$ and $\text{se}(\hat{\phi}_2)$. From equation (31), $\text{se}(\hat{\phi}_2)$ equals

$$\text{se}(\hat{\phi}_2) = \hat{\sigma}_1 \sqrt{[(\mathbf{M}_1 \hat{\mathbf{y}}_A)^T (\mathbf{M}_1 \hat{\mathbf{y}}_A)]^{-1}} = \hat{\sigma}_N \sqrt{[(\hat{\gamma} \mathbf{M}_1 \mathbf{z})^T \hat{\gamma} \mathbf{M}_1 \mathbf{z}]^{-1}}$$

or

$$\text{se}(\hat{\phi}_2) = \hat{\sigma}_N \sqrt{\frac{1}{\hat{\gamma}^2} [(\mathbf{M}_1 \mathbf{z})^\top \mathbf{M}_1 \mathbf{z}]^{-1}} = \frac{1}{\hat{\gamma}} \hat{\sigma}_N \sqrt{[(\mathbf{M}_1 \mathbf{z})^\top \mathbf{M}_1 \mathbf{z}]^{-1}} = \frac{1}{\hat{\gamma}} \text{se}(\hat{\theta}). \quad (37)$$

Equation (37) follows from the mathematical relationships used to obtain the results in equation (36).

Therefore, $\text{se}(\hat{\phi}_2) = \frac{1}{\hat{\gamma}} \text{se}(\hat{\theta})$ and, as shown below, the same mathematical relationships used above will be used to demonstrate a similar relationship between $\hat{\theta}_2$ and $\hat{\phi}_2$.

The third key result to demonstrate: $\hat{\phi}_2 = \hat{\theta}_2$

The formula for the OLS estimate of ϕ_2 is in equation (28). Beginning with that equation, repeated substitution and simplifying will show that $\hat{\phi}_2 = \frac{1}{\hat{\gamma}} \hat{\theta}_2$. The steps of this derivation follow.

$$\begin{aligned} \hat{\phi}_2 &= [(\mathbf{M}_1 \hat{\mathbf{y}}_A)^\top (\mathbf{M}_1 \hat{\mathbf{y}}_A)]^{-1} (\mathbf{M}_1 \hat{\mathbf{y}}_A)^\top (\mathbf{M}_1 \mathbf{y}) = [(\hat{\gamma} \mathbf{M}_1 \mathbf{z})^\top (\hat{\gamma} \mathbf{M}_1 \mathbf{z})]^{-1} (\hat{\gamma} \mathbf{M}_1 \mathbf{z})^\top (\mathbf{M}_1 \mathbf{y}) \\ \hat{\phi}_2 &= [\hat{\gamma}^2 (\mathbf{M}_1 \mathbf{z})^\top (\mathbf{M}_1 \mathbf{z})]^{-1} \hat{\gamma} (\mathbf{M}_1 \mathbf{z})^\top (\mathbf{M}_1 \mathbf{y}) \\ \hat{\phi}_2 &= \frac{\hat{\gamma}}{\hat{\gamma}^2} [(\mathbf{M}_1 \mathbf{z})^\top (\mathbf{M}_1 \mathbf{z})]^{-1} (\mathbf{M}_1 \mathbf{z})^\top (\mathbf{M}_1 \mathbf{y}) = \frac{1}{\hat{\gamma}} \hat{\theta}_2 \end{aligned} \quad (38)$$

Using previous key results to prove Proposition #1

Given $\hat{\phi}_2 = \frac{1}{\hat{\gamma}} \hat{\theta}_2$ and $\text{se}(\hat{\phi}_2) = \frac{1}{\hat{\gamma}} \text{se}(\hat{\theta})$, Proposition #1 can finally be proven. Proposition #1 states that if two competing model specifications differ by only one explanatory variable, the test statistic derived by the non-nested J test and the test statistic derived by the nested, hybrid model have the same absolute value. This result follows from the linear relationship between $\hat{\phi}_2$ and $\hat{\theta}_2$ and the linear relationship between $\text{se}(\hat{\phi}_2)$ and $\text{se}(\hat{\theta}_2)$ because

$$|t_{\hat{\phi}_2}| = \left| \frac{\hat{\phi}_2}{\text{se}(\hat{\phi}_2)} \right| = \left| \frac{\hat{\theta}_2 / \hat{\gamma}}{\text{se}(\hat{\theta}_2) / \hat{\gamma}} \right| = \left| \frac{\hat{\theta}_2}{\text{se}(\hat{\theta}_2)} \right| = |t_{\hat{\theta}_2}|. \quad (39)$$

Thus, in this special case where the two model specifications differ by one explanatory variable, both the non-nested J test and the nested, hybrid model give the same results.

V. CONCLUSION

Even though the J test involves running four regressions and the nested, hybrid model specification test only requires the estimation of two regressions, some argue that the J test is a superior test. There is increased probability that multicollinearity will adversely affect the precision at which the parameters are estimated in the nested, hybrid model. The nested model may lack economic meaning and the results of the nested model may not be invariant to which specification is chosen as the reference model.

However, this paper shows that in the special case when two competing models differ by only one exogenous variable, the nested model specification test results in test statistics that have the same absolute value as the non-nested J test. In this case, obtaining statistical evidence on the correct model specification is quicker with the nested test than with the non-nested test. Estimation of one, and only one, nested, hybrid model results generates the same test statistics in absolute value that the non-nested J test generates after estimating four separate regressions. When two models differ by only one explanatory variable, use of the nested test offers some time savings in determining the correct model over using the non-nested J test.

REFERENCES

- Breusch, T. S. "Testing for Autocorrelation in Dynamic Linear Models." *Australian Economic Papers*, 17(31), 1978, 334 – 355.
- Chow, Gregory. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica*, 28(3), July 1960, 591 – 605.
- Davidson, R. and J. MacKinnon. "Several Tests for Model Specification in the Presence of Alternative Hypotheses." *Econometrica*, 49(3), May 1981, 781-793.
- Fomby, T., R. Carter, and S. Johnson. *Advanced Econometric Methods*. New York: Springer-Verlag, 1984.
- Godfrey, L. G. "Testing Against General Autoregressive and Moving Average Error Models When the Regressor include Lagged Dependent Variable." *Econometrica*, 46(6), November 1978, 1293 – 1302.
- Gujarati, D. *Basic Econometrics*. 4th. ed. Boston: McGraw Hill, 2003.
- Johnston, J. and J. DiNardo. *Econometric Methods*. 4th. ed. New York: McGraw Hill, 1997.
- Judge, G., W. Griffiths, R. Hill, H. Lütkepohl, and T. Lee. *The Theory and Practice of Econometrics*. 2nd. ed. New York: Wiley, 1985.
- Kennedy, P. *A Guide to Econometrics*. 4th. ed. Cambridge, Massachusetts: MIT Press, 1998.
- Kmenta, J. *Elements of Econometrics*. 2nd. ed. New York, Macmillan, 1986.
- Maddala, G. *Introduction to Econometrics*. New York: Macmillan, 1988.
- McAleer, Michael. "The Significance of Testing Empirical Non-Nested Models." *Journal of Econometrics*, 67(1), May 1995, 149-171.
- Phillips, A. W. "The Relationship Between Unemployment and the Rate of Change of Money Wages in the United Kingdom, 1861-1957." *Economica*, 25(2), November 1958, 293 - 299.
- Thomas, R. *Modern Econometrics: An Introduction*. Harlow, England: Addison-Wesley, 1997.

Wooldridge, J. *Introductory Econometrics: A Modern Approach*. 2nd. Ed. Mason, Ohio: South-Western, 2003.

TABLE 1
Steps and Decision Rule for Non-Nested J Test

Steps to Implement Non-Nested J-Test

Step 1: Estimate Model A, $y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ji} + \sum_{j=1}^m \delta_j w_{ji} + \varepsilon_i$. Obtain predicted values for Model A, \hat{y}_i^A , $i = 1, \dots, N$.

Step 2: Estimate Model B, $y_i = \alpha_0 + \sum_{j=1}^k \alpha_j x_{ji} + \sum_{j=1}^n \gamma_j z_{ji} + \mu_i$. Obtain predicted values for Model B, \hat{y}_i^B , $i = 1, \dots, N$.

Step 3: Add predicted values from Model B as an additional explanatory variable in Model A. Estimate $y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ji} + \sum_{j=1}^m \delta_j w_{ji} + \phi \hat{y}_i^B + \varepsilon_i$. Obtain $t_{\hat{\phi}}$, the t-statistic associated with $\hat{\phi}$, the OLS estimate for ϕ , (the test statistic for the null hypothesis that $\phi = 0$).

Step 4: Add predicted values from Model A as an additional explanatory variable in Model B. Estimate $y_i = \alpha_0 + \sum_{j=1}^k \alpha_j x_{ji} + \sum_{j=1}^n \gamma_j z_{ji} + \theta \hat{y}_i^A + \mu_i$. Obtain $t_{\hat{\theta}}$, the t-statistic associated with $\hat{\theta}$, the OLS estimate for θ , (the test statistic for the null hypothesis that $\theta = 0$).

Decision Rule

Pick significance level α (usually 1%, 5%, or 10%). Using t-table, find critical value corresponding to $\frac{\alpha}{2}$ and the appropriate degrees of freedom (df), or $t_{\alpha/2, df}$. There are four possible outcomes.

Outcome 1: If $t_{\hat{\phi}} \geq t_{\alpha/2, df}$ and $t_{\hat{\theta}} < t_{\alpha/2, df}$, fail to reject Model B and reject Model A.

Outcome 2: If $t_{\hat{\phi}} < t_{\alpha/2, df}$ and $t_{\hat{\theta}} \geq t_{\alpha/2, df}$, fail to reject Model A and reject Model B.

Outcome 3: If $t_{\hat{\phi}} < t_{\alpha/2, df}$ and $t_{\hat{\theta}} < t_{\alpha/2, df}$, reject both Models A and B.

Outcome 4: If $t_{\hat{\phi}} \geq t_{\alpha/2, df}$ and $t_{\hat{\theta}} \geq t_{\alpha/2, df}$, fail to reject both Models A and B.

TABLE 2
Steps and Decision Rules for Nested Specification F Test

Steps to Implement the Nested Specification F Test

Step 1: Estimate Model A, $y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ji} + \sum_{j=1}^m \delta_j w_{ji} + \varepsilon_i$, and obtain the R^2 associated with this regression, R_A^2 .

Step 2: Estimate a nested or hybrid model that includes all explanatory variables used in both Model A and in Model B, $y_i = \varphi_0 + \sum_{j=1}^k \varphi_j x_{ji} + \sum_{j=1}^m \psi_j w_{ji} + \sum_{j=1}^n \theta_j z_{ji} + \eta_i$, and obtain the R^2 associated with this nested or hybrid regression, R_N^2 .

Step 3: To determine whether Model B has more explanatory power than Model A, test whether all the θ_i , $i = 1, \dots, n$, are simultaneously equal to zero. The null hypothesis is $H_0: \theta_1 = \theta_2 = \dots = \theta_n = 0$, while the alternative hypotheses is H_A : Not all the θ_i are

simultaneously equal to zero. Derive the test statistic, $F = \frac{(R_N^2 - R_A^2)/n}{(1 - R_N^2)/(N - k - m - n - 1)}$.

Under the null hypothesis, this test statistic is distributed F with n degrees of freedom in the numerator and $N - k - m - n - 1$ degrees of freedom in the denominator. i.e., $F_{n, N-k-m-n-1}$.

Decision Rule

Pick significance level α (usually 1%, 5%, or 10%). Using F table, find critical value corresponding to α with n degrees of freedom in numerator and $N - k - m - n - 1$ degrees of freedom in the denominator, or $F_{\alpha n, N-k-m-n-1}$. There are two possible outcomes.

Outcome 1: If $F > F_{\alpha n, N-k-m-n-1}$, reject the null hypothesis. This indicates that the different exogenous variables in Model B add significant explanatory power to Model A. Thus, Model B is chosen over Model A.

Outcome 2: If $F < F_{\alpha n, N-k-m-n-1}$, fail to reject the null hypothesis. This indicates that the different exogenous variables in Model B add no significant explanatory power to Model A. Thus, Model A is chosen over Model B.

TABLE 3
Empirical Example Results: The Phillip's Curve, Annual Data, 1974-2003
Dependent Variable: Current Inflation Rate (π_t)

<i>Right-Hand-Side Variables</i>	The Two Different Model Specification Tests				
	Model A Equation (3)	Model B Equation (4)	Non-Nested J Test		Nested Test
			Test 1 Equation (5)	Test 2 Equation (6)	Test 3 Equation (7)
Intercept	4.91* (3.27)	-3.75 (-1.64)	14.28* (2.90)	10.26*** (1.97)	23.94** (2.47)
Lagged Inflation (π_{t-1})	1.06* (8.47)	0.98* (7.55)	3.59* (2.80)	-1.87*** (-1.90)	1.07* (9.00)
Current Unemployment Rate (U_t)	-0.83* (3.00)		-2.32* (-2.92)		-2.32* (-2.92)
Reciprocal of Current Unemployment Rate ($\frac{1}{U_t}$)		22.76*** (2.05)		-58.63*** (1.98)	-58.63*** (-1.98)
Predicted Value of Inflation Rate from Model A ($\hat{\pi}_t^A$)				2.78* (2.92)	
Predicted Value of Inflation Rate from Model B ($\hat{\pi}_t^B$)			-2.58*** (-1.98)		
<i>Other Regression Diagnostic Statistics</i>					
Breusch-Godfrey Serial Correlation Test	1.25	2.34			
R ²	0.75	0.71	0.78	0.78	0.78
Adjusted R ²	0.73	0.69	0.76	0.76	0.76
Sum of Squared Residuals	75.26	86.81	65.36	65.36	65.36
Log Likelihood	-53.36	-58.51	-54.25	-54.25	-54.25
Durbin-Watson Statistic	1.41	1.30	1.67	1.67	1.67
Akaike Information Criterion (AIC)	3.96	4.10	3.88	3.88	3.88
Schwarz Criterion	4.10	4.24	4.07	4.07	4.07
F Statistic	40.15 [†]	33.01 [†]	30.99 [†]	30.99 [†]	30.99 [†]

* , ** , and *** indicate the estimated coefficient is statistically different from zero at the one percent, 5 percent, and 10 percent level, respectively, using a two-tail t test. † indicates the F statistic has a level of significance equal to one percent.