EMPIRICAL POWER COMPARISON OF NON-NESTED TESTS FOR THE EVM: SOME MONTE CARLO EVIDENCE

BHATTI, M.Ishaq*
BODLA, Mahmud, A.

Abstract
Recently, Bodla and Bhatti (2007) revisited Davidson and MacKinnon's (2002) well-known J test and noted that thought the test is simple to compute but lack small sample exact test computation properties. This paper is one of the attempts to compute a new version of the J test and compare its power performance with the various existing tests to see the relative strength of our test to be called as an approximately most powerful test. The main objective of this paper is to study Monte Carlo evidence on finite sample performance of the now modified non-nested tests of mismeasured regression models in EVM, Errors in Variables Models, setting to see if the power performance of the new test.

Key words: Nonnested models, power & size of a test, Monte Carlo Simulation.

1. Introduction

Two models are called nested if one model is a special case of the other. Alternatively, of the two models, if one can be reduced to the other by imposing restrictions on certain parameters then they are called nested models. For example, $H_0 = \beta_0 + \beta_1 X_1 + e_0$ and $H_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_1$ are nested models because by imposing the restriction $\beta_2 = 0$, H_1 becomes H_0. In fact, H_1 encompasses H_0. If we wish to discriminate these two models, we just need to test the restriction on β_2. This is generally done by a t-

* M. Ishaq Bhatti, Department of Economics and Finance, La Trobe University, Melbourne, Australia, E-Mail: i.bhatti@latrobe.edu.au and Mahmood A. Bodla, Department of Management Sciences, M.A. Jinnah Campus, Defense Road off Raiwind Road, Lahore, Pakistan, E-Mail: MABodla@ciitlahore.edu.pk

test under ordinary least squares (OLS). On the other hand, two models are said to be nonnested (also called separate) if one model cannot be reduced to the other model by imposing restrictions on certain parameters. For example, $H_0 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_0$ and $H_1 = \beta_3 + \beta_4 X_4 + \beta_5 X_5 + e_1$ are nonnested models because H_1 cannot be H_0 in any way. That means, in this particular setting, one model is not a special case of the other.

The literature on nonnested hypothesis testing stems from a pioneering seminal work of Cox (1961, 1962) and then followed by Davidson and MacKinnon (1981), Doubleman et. el. (1978), Ericsson (1986), Hostag and Hein (9190), Bernanke et al. (1986), McAleer (1987), Pesaran and Pesaran (1993), Huth et al (1993), Silvapulle and King (1993), McAleer (1995), Fan and Li (1995), Godfrey (1998), Berger and Mortera (1999), Davidson and MacKinnon (1981, 2001, 2002), King (1989, 1998), Camilli (2006), Huang, et. al. (2006) and Bodla and Bhatti (2007), among others. Much of the emphasis in this literature has been on testing non-nested regression.

Davidson and MacKinnon's (2002) well-known J test is perhaps the most widely used procedure for testing nonnested regression models. This test is conceptually simple and easy to compute but not an exact test in finite sample. As a result, a number of attempts have been made to improve the finite sample properties of the J test. This paper is one of the attempts to assess the power performance of the new modified J tests to see the relative strength of our test to be called as the most powerful test.

Theoretical properties of non-nested tests in general and the modified non-nested tests for errors-in-variables models (EVM) in particular, (see, Shumway and Gottrel (1991), Camilli (2006) and Huang, et. al. (2006) are limited to asymptotic results, whereas real world econometric analysis deals with finite samples. Earlier studies by Godfrey and Pesaran (1983), Davidson and MacKinnon (1982, 2001) and provide some evidence on the small sample behavior of various non-nested tests. Their design specificity however, limits the scope of their findings to non-nested linear regression models when all regressors are non-stochastic and/or exogenous. The main objective of this paper is to study Monte Carlo evidence on finite sample performance of the now modified non-nested tests of mismeasured regression models in EVM setting. The structure of the rest of this paper is as follows. In the subsequent sections the EVM model specification is given, the description and the design of the Monte Carlo experiment is presented in section three. This includes, data generating process of mismeasured models, measurement of EVM specification with instrumental variable (IV) and the alternative choices of IV parameter selection. Section four demonstrates the details of computational aspects of the Monte Carlo study. The results of various experiments are presented in section five. The final section contains some concluding remarks.

2. The EVM - Model

Following Chan et al (2005), Camilli (2006) and Bodla and Bhatti (2007) non-nested errors-in-variables models - EVM is expressed as:

$$H_0:\ y = X\beta + u_0, \text{ and } H_1:\ y = Z\gamma + u_1 \tag{1}$$

where in (1) above the nxk_1 matrices of explanatory variables, X and Z, contain non-overlapping variables and consist of some or all mismeasured variables; i.e., $X = X_T + V_0$ and $Z = Z_T + V_1$. X_T and Z_T represent the correctly measured components of X and Z whereas V_0 and V_1 are the matrices of measurement errors. The number of mismeasured regressors in X and Z are denoted by m_0 and m_1, respectively, so that, m_0 and m_1 must also be the number of non-zero columns in V_0 and V_1. Each non-zero column in V_0 and V_1, denoted v_{0i} and v_{1j} for i = l,...,m_0 and j = l,...,m_1, is assumed to be N(0, $\sigma^2_{voi} I_n$) and N(o, $\sigma^2_{v1j} I_n$) where we assume that σ^2_{voi} and σ^2_{v1J} are constants (i.e., homocedastic). Godfrey and Pesaran (1983), considered a similar model with fixed and stochastic regressors and do not considered the issue of mismeasure regressors. One may compare their power studies with ours by conducting a similar procedure of conducting Monte Carlo experiment with their non-stochastic regressor models and the EVM model considered in this studied here.

3. **Design of the Experiments**

The Monte Carlo experiments aimed at performance evaluation of the IV-based or modified tests for H_0 and H_1 above are described in the following subsections. The first subsection defines the 'true' data generating process (DGP) and its 'fixed' alternative followed by a discussion of the mismeasured models. Since these EVM specifications are to be estimated with instrumental variables, the third section discusses alternative choices for these IV matrices and the fourth subsection deals with parameter selection.

Description of the True' DGP and 'Fixed' Alternative

The experimental design described below is a reproduction of the models in and Godfrey and Pesaran (1983), with the exception of the introduction of mis-measured regressors. By replicating their Monte Carlo design, a direct comparison may be made between their non-stochastic regressor models and the EVM tests studied here. The true, albeit unobservable, DGP for each replication of the Monte Carlo experiment is specified by the following model:

$$H_T:\ y = X_T\beta + u_T. \tag{2}$$

The n observations for each of the k_0 columns in X_T are generated as iid standard normal variates, while β is chosen as a unit vector of order k_0x1. The error term in (2) is generated as $u_T \sim N(0,\ \sigma_T^2 I_n)$, whose variance, σ_T^2, is given by

$$\sigma_T^2 = \frac{\beta'\beta(1 - R^2)}{R^2} = \frac{k_0(l - R^2)}{R^2}$$

where R^2 in σ_T^2 above is the coefficient of multiple determination for H_T given in (2).

The 'false' non-nested alternative model is given by:

$$H_F: \; y = Z_T\gamma + u_F, \tag{3}$$

where y represents the true DGP in (2), Z_T is an nxk_1 matrix of explanatory variables, and u_F represents an nxl vector of unexplained discrepancies between y and $Z_{T\,r}$. The condition of non-orthogonality between X_T and Z_T is ensured by generating the ith column in Z_T as follows:

$$Z_{Ti} = \lambda i X_{Ti} + e_{Ti}, \text{ for all values of } i = 1,...,k_I; \tag{4}$$

where $e_{Ti} \sim N(0, I_n)$, $\lambda i = p_i / \sqrt{1 - p_i^2}$ and p_i is the simple correlation coefficient between X_{Ti} and Z_{Ti}. It is assumed for simplicity that p_i = p, a constant for all i, and that X_T and Z_T contain no common regressors; that is, $p \neq 1$ for any i and p_i is non-zero for all i. Assuming that the value of *p* is fixed as n increases, H_F may be viewed as the 'fixed ' alternative of Davidson and MacKinnon (2001).

Although Godfrey and Pesaran (1983) consider both equal and unequal numbers of regressors, the case of $k_0 = k_1 = k$ will be undertaken here.

Description of the EVM

The mismeasured non-nested models, H_0 and H_1, are obtained by transforming the correctly measured regressor matrices X_T and Z_T into the mismeasured regressor matrices X and Z. The m_0 independent columns of measurement errors for V_0 were generated as $v_{oi} \sim N(0, \sigma^2_{voi} I_n)$ for i = l,...,m_0: similarly, $v_{1j} \sim N(0, \sigma^2_{v1j} I_n)$ for j = 1,...,m_1. By adding V_0 to the first m_0 columns of X_T and V_1 to the last m_1 columns of Z_T, the mismeasured regressors are defined as $X = [X_m \mid X_f]$ and $Z = [Z_f \mid Z_m]$ where X_m and Z_m denote mismeasured columns in X and Z while X_f and Z_f represent the correctly measured regressors. The number of mismeasured regressors and the variances of the measurement errors are expected to influence the performance

of the modified non-nested tests. However, to keep the number of experiments within limits, it is assumed that $m_0 = m_1 = m$ and
$\sigma^2_{voi} = \sigma^2_{v1j} = \sigma^2_m$ for all i and j.

Description of the IV

As discussed in earlier section, appropriate instrumental variables for the non-nested tests of errors-in-variables models may, under certain conditions, be chosen as regressors of one model acting as instrumental variables for the other model. In the experiment detailed below, the null and alternative hypotheses will each contain the same number of correctly measured and mismeasured regressors. (For example, when $k = 2$, $m_0 = m_1 = 1$.) In this instance, it is appropriate to choose the correctly measured regressor(s) in H_0 (H_1) as the instrument(s) for the mismeasured regressor(s) in H_0 (H_1). More formally, the IV matrix for these experiments is given by $W_0 = [Z_f \mid X_f] = W_1$. Thus, $W_0 = W_1$ is obtained by combining the correctly measured columns in X and Z.

Parameter Selection

The IV selection from the regressors of competing non-nested models provides a novel and specific approach to the general problem of instrumental variable specification. To fully examine the properties of the EVM non-nested tests, the chosen experimental design involves the 320 design points defined by the following chosen parametric values.

(k, m) = (2, 1), (4, 2)

R^2 = (0.99, 0.95, 0.80, 0.50, 0.30)

ρ^2 = (0.30, 0.50, 0.80, 0.95)

σ^2_m = (1.0, 0.25)

n = (20, 40, 60, 100)

It is expected that, *ceteris paribus,* the performance of the non-nested EVM tests will worsen with decreases in the sample size (n) and R^2, and with increases in ρ^2 and σ_m^2. For instance, since σ_m^2 is the mismeasurement error variance, as σ_m^2 approaches zero, this experiment collapses to that of Godfrey and Pesaran (1983) for non-stochastic regressors. On the other hand, a poor fit of the true DGP, as measured by a low value for R^2, will make it more difficult for the tests to identify the 'true' model.

4. Computational Details

All necessary computations of the new modified tests and power performance indicators of the new test were conducted using FORTRAN programming. The generation of the null and alternative hypotheses requires several independently and identically distributed normal deviates; the matrices X_T, u_T, e_T, v_{oi}, and V_{1j} were generated by invoking the DRNNOA subroutine of the IMSL/STAT Library. For each of the 320 experiments *with 500* replications were performed. By setting a 95% probability that the estimated and nominal Type-I errors will differ by no more than 0.02: on this point.

For the chosen instrument $W_0 = W_1 = [Z_f \mid X_f]$, the non-nested test statistics G_0, $\tilde{\eta}^{m}_{w0}$ and $\tilde{\eta}_{LM0}$ do not exist: see theorem 4.14. Also in this instance, $\tilde{J}_0 = J\tilde{A}_0$ and $\tilde{J}_0^m = J\tilde{A}_0^m$. Computationally, it is easier to work with the squares of these test statistics, so that the statistics computed for these experiments will be $\tilde{J}_0^2$ and $\tilde{J}_0^{m^2}$ each following a X_1^2 distribution under H_0. By interchanging the roles of H_0 and H_1, the values $\tilde{J}_1^2$ and $\tilde{J}_1^{m^2}$ serve as the test of H_1.

The performance indicators calculated for each experiment are the mean, variance, skewness and kurtosis of the empirical distribution, the x^2 goodness-of-fit test, and Type-I error, Type-II error, and power measures for $\tilde{J}_0^2$ and $\tilde{J}_0^{m^2}$.

The mean, variance and coefficients of skewness and kurtosis for the sampling distributions of $\tilde{J}_0^2$ and $\tilde{J}_0^{m^2}$ are calculated by invoking the DUVSTA subroutine of the IMSL/ STAT Library. Similarly, by calling DCHIGF, the Chi-square goodness-of-fit measures are obtained. Since the sample observations for each statistic are divided into 20 equal groups, the goodness-of-fit test will have 19 degrees of freedom.

The Type-I errors, Type-II errors, and powers of these tests are tabulated at the three conventional significance levels: α = 0.01, 0.05, 0.10. In order to determine appropriate critical values for each significance level, the IMSL/STAT subroutine DCHIIN was invoked. The empirical Type-I and Type-II errors for a test, say $\tilde{J}^2$, are respectively defined as the proportion of times in 500 replications that $\tilde{J}_0^2$ is greater than, and $\tilde{J}_1^2$ is less than, the critical value of X_1^2 for a certain value of α. On the other hand, the power of a test represents the proportion of times that neither a Type-I nor Type-II error occurs. The standard errors for the Type-I error, Type-II error and power of each test are also estimated. The estimated standard error, say for the size of a test, is given by $\sqrt{\hat{p}(1-\hat{p})/500}$ where $\hat{p}$ denotes the estimated Type-I error.

5. Results of the Experiments

The full sets of Monte Carlo results are very long and hence here for the sake of simplicity in explosion we have summarized these results in tables 5.1 to 5.3. For our computational purposes, we have computed the sizes and powers for the $\tilde{J}_0^2$ and $\tilde{J}_0^{m^2}$ for all values of n = 20, 40, 60, and 100, permitting a direct evaluation of the approach to the asymptotic distribution. In our computation we have considered all combinations of the values of R^2 = (0.99, 0.95, 0.80, 0.50, 0.30) and ρ^2 = (0.30, 0.50, 0.80, 0.95) for a given set of values for k, m, and σ_m^2. In particular, for table 5.1, set k = 2, m = 1, σ_m^2 = 1.00, table 5.2, set k = 4, m = 2, σ_m^2 = 1.00, table 5.3, set k = 2, m = 1, σ_m^2 = 0.25, and table 5.4.1 - 5.4.20 set k = 4, m = 2, σ_m^2 = 0.25.

For different sample sizes (n) and a given set of other parametric values, each table reports the estimated mean (M), variance (V), coefficients of skewness (S) and kurtosis (K), and the goodness-of-fit measure (x^2) for $\tilde{J}_0^2$ and $\tilde{J}_0^{m^2}$. The estimated Type-I and Type-II errors and power of each statistic at 1%, 5% and 10% significance levels, with standard errors in parentheses, are reported in the last nine columns of each table.

The observed size estimates in all tables that depart significantly (i.e., differ by more than two standard errors) from their nominal values are marked with an asterisk. Similarly, the goodness-of-fit test (x^2) exceeding X^2_{19} at the a = 5% level is so marked ($X^2_{19,0.05} \cong$ 30.144), indicating a lack of fit between the empirical distribution and the theoretical x^2_1 distribution. The theoretical mean of these distributions is one (the degrees of freedom of the x^2_1 distribution), so that the amount by which the empirical mean (M) differs from 1 is the estimated bias. Analogously, the theoretical variance is two (twice the degrees of freedom of the x^2_1distribution). Thus, a cursory examination of each table reveals the behavior of a test for different sample sizes as well as a comparison between $\tilde{J}^2$ and $\tilde{J}^{m^2}$ for a given set of parameters. However, evaluating the tests for different parametric values involves inter-table comparisons. Some of the interesting results of this experiment are now discussed.

As expected, the observed means, variances, measures of goodness-of-fit, and sizes of both tests generally approach their theoretical counterparts as the sample size gets larger. For example, the size of $\tilde{J}^2$ in Table 5.1 moves from 0.046 at n = 20 to 0.082 at n = 100 when α = 0.10; similarly, the size of $\tilde{J}^{m^2}$ ranges between 0.068 and 0.074. Increasing sample sizes also yield higher powers for those tests with correct sizes; e.g., in table 5.1 at α = 0.10, $\tilde{J}^{m^2}$ has the correct size for all sample values while its power increases from 27% at n = 20 to 67% at n = 100.

The relative performance of $\tilde{J}^2$ and $\tilde{J}^{m^2}$ can also be seen from these results. A careful examination of each table, in all of the four sets, reveals that the observed significance levels of $\tilde{J}^2$ are generally less than the corresponding values of $\tilde{J}^{m^2}$. However, the means of both tests are not usually farther apart (with the exception of the results reported in tables below for R^2 = 0.80, 0.95 and ρ^2 = 0.95), suggesting that the lower size of $\tilde{J}^2$ is most likely caused by the lower values of its variance and kurtosis. Moreover, the power of $\tilde{J}^{m^2}$ generally exceeds that of $\tilde{J}^2$ (when each test has the correct size). As an example, in those cases where $\tilde{J}^2$ and $\tilde{J}^{m^2}$ each have correct Type-I errors, selected powers of these tests have been reproduced below when n = 100.

Table 5. 1: Selected Estimated Powers

		α = 0.01	**α = 0.01**	**α = 0.05**	**α = 0.01**
(1)	$\tilde{J}^2$	0.374	0.320	0.890	0.824
	$\tilde{J}^{m^2}$	0.670	0.676	0.942	0.970
(2)	$\tilde{J}^2$	0.020	0.018	0.058	0.082
	$\tilde{J}^{m^2}$	0.024	0.024	0.066	0.086

Notes: (1) R^2 = 0.99 and ρ^2 =0.30, (2) R^2 = 0.30 and ρ^2 =0.9

The higher power of $\tilde{J}^{m^2}$ over $\tilde{J}^2$ is quite encouraging, since the modified test, $\tilde{J}^{m^2}$, is also the easiest of the two tests to compute. (Recall that $\tilde{J}^m$ can be viewed as the direct t-ratio in a compound model, whereas $\tilde{J}$ cannot be calculated on existing software). An important caveat to this claim of high power also needs to be made - in some instances, the $\tilde{J}^{m^2}$ test fails to reach the correct size even in samples as large as 100.

The Monte Carlo results in Godfrey and Pesaran (1983) measure an expected decline in the power of non-nested tests *ceteris paribus* with decreasing ρ^2. The same trend applies to the EVM non-nested tests. For example, the power of $\tilde{J}^{m^2}$ as R^2 decreases when n = 100 and a = 0.10 are 0.810, 0.792, 0.746, 0.622 and 0.464. On the other hand, the effect of ρ^2 in this experiment is not the same as those reported by Godfrey and Pesaran for the case of non-stochastic regressor models, since ρ^2 plays two competing roles here. First, the increased value of ρ^2 increases the efficiency of the IV estimators of H_0 which may increase the power of the modified tests. Second, with increasing values of ρ^2, differentiating between the null and fixed alternative hypotheses becomes increasingly difficult and the power of a non-nested test may decline. The power of $\tilde{J}^2$ and $\tilde{J}^{m^2}$ (when these tests have correct sizes) increases in all four sets when ρ^2 increases from 0.30 to 0.50 while other parameters are fixed. For example, the powers of $\tilde{J}^2$ are reproduced below for those instances in which R^2 = 0.99, n = 100 and α = 0.01. From these results, it is noted that the power of $\tilde{J}^2$ (and also that of $\tilde{J}^{m^2}$ not tabulated here) increases up to a point (ρ^2 = 0.50) and then begins to decline. In one instance, the drop in power is quite dramatic — note for group 4 that power drops from 94% when ρ^2 = 0.50 to 45% when ρ^2 = 0.95. Thus, it is interesting to conclude that for this particular experimental design, at lower levels of ρ^2, an increase from 0.30 to 0.50 causes the efficiency effect of the IV estimators of H_0 to dominate, causing an increase in the power of $\tilde{J}^2$ and $\tilde{J}^{m^2}$. Further increase in the values of ρ^2 (i.e., greater than 0.50) will cause in declining the powers. This is due to the increased closeness between the null and fixed alternative hypotheses.

Table 5.2 Selected estimated power

	k= 2, m = 1, $\sigma_m^2 = 1.00$	k= 4, m = 2, $\sigma_m^2 = 1.00$	k= 2, m = 1, $\sigma_m^2 = 0.25$	k= 4, m = 2, $\sigma_m^2 = 0.25$
$\rho^2 = 0.30$	0.374	0.320	0.832	0.824
$\rho^2 = 0.50$	0.484	0.458	0.950	0.940
$\rho^2 = 0.80$	0.310	0.300	0.938	0.942
$\rho^2 = 0.95$	0.052	0.052	0.470	0.454

The effects of (k, m) on the performance of $\tilde{J}^2$ and $\tilde{J}^{m^2}$ can also be isolated. In general, both tests exhibit the correct Type-I error more often when (k, m) = (2, 1) than when (k, m) = (4, 2). Moreover, for these cases in which valid comparisons can be made, power declines when (k, m) increases from (2, 1) to (4, 2). However, the differences in powers are usually less pronounced for higher values of n and R^2 and low values of ρ^2, i.e., n = (60, 100) , R^2 = (0.99, 0.95, 0.80) and ρ^2 = (0.30, 0.50), whereas differences in power are otherwise quite substantial. Finally, a divergence between the estimated powers of the $\tilde{J}^2$ and $\tilde{J}^{m^2}$ tests is more likely to be found as k and m increase.

The performance indicators generally reveal marked improvements when σ_m^2, the mismeasurement error variance, declines from 1.00 to 0.25. In these tables where σ_m^2 = 0.25, the observed values of various measures are mostly in close proximity to their true values; the notable discrepancies in the mean, variance, and size of $\tilde{J}^{m^2}$ occur when n = 20, R^2 = 0.95, 0.80, and ρ^2 = 0.95, refer table

5.3.The effect of σ_m^2 on power is typified in the following, where the powers of $\tilde{J}^2$ for n = 20, 100, *a* = 1%, R^2 = 0.99, and ρ^2 = 0.30. Interestingly, power actually declines when σ_m^2 declines in the small sample case (n = 20). It is only when n = 100 that improvements in power are associated with declining mismeasurement errors for the $\tilde{J}^{m^2}$ test. Finally, the size and power of $\tilde{J}^{m^2}$ for σ_m^2 = 0.25 are compared with those of Type-I Error and Power Comparisons of $\tilde{J}^m$ and J^2. The size or Type-I errors and powers of J^2 (in the columns marked σ_m^2 = 0) are reproduced from McAleer (1987) as originally computed by Godfrey and Pesaran (1983). The asterisk here denotes the Type-I errors that are significantly different from *a* = 0.05.

Table 5.3: Type-I Error and Power Comparisons of $\tilde{J}^{m^2}$ and J^2

n = 20, R^2 = 0.50, *a*=0.05

		Size		**Power**	
k	ρ^2	σ_m^2 = 0.25	σ_m^2 = 0.00	σ_m^2 = 0.25	σ_m^2 = 0.00
2	0.30	0.050	0.062	0.376	0.834
	0.50	0.070	0.044	0.324	0.766
	0.80	0.072	0.054	0.200	0.418
	0.95	0.054	0.052	0.098	0.126
4	0.30	0.052	0.142*	0.236	0.778
	0.50	0.060	0.108*	0.264	0.732
	0.80	0.070	0.074*	0.160	0.416
	0.95	0.044	0.072	0.094	0.108

Godfrey and Pesaran's (1983) values for J^2 where σ_m^2 = 0 (i.e., the case of non- stochastic regressors). Fixing R^2 = 0.50 and *a* = 0.05, provides an expected decrease in power with increasing values of ρ^2, σ_m^2 and k. There exists a marked difference between the powers of J^2 and $\tilde{J}^{m^2}$ for lower values of ρ^2 while it is insignificant when ρ^2 = 0.95. It may also be expected that such differences will diminish in large samples (n = 100) or with higher values of R , since the power of $\tilde{J}^m$ in these cases reaches the maximum. However, the published results of Godfrey and Pesaran (1983) are not available for such parametric values and no direct comparisons are possible to prove the point. Finally, the small sample over-rejection of H_0 seems to be a serious problem for J^2 when k = 4 and ρ^2 = 0.30, 0.50, and 0.80 whereas $\tilde{J}^{m^2}$ has correct sizes.

In general, the EVM non-nested tests have performed as expected. For this particular Monte Carlo design, these tests appear to be most useful when R^2 and n are relatively large (i.e., $R^2 > 0.80$ and n > 40) and σ_m^2 is moderately low (i.e., σ_m^2 = 0.25). The effect of ρ^2 on power is ambiguous.

6. Concluding Remarks

This study, for the first time in the econometric literature, provides Monte Carlo evidence on the behavior of non-nested tests that involve IV estimation of mismeasured regression models. The design of the Monte Carlo experiment involves a true DGP and a fixed alternative hypothesis suggested by Godfrey and Pesaran (1983), while some of the explanatory variables in both models are then designed to be mismeasured. Instrumental variables are chosen to be the correctly measured regressors in the competing model. Different combinations of the values of n, R^2, k, m and σ_m^2 are specified; the mean, variance, coefficients of skewness and kurtosis, measure of goodness-of-fit, Type-I errors, Type-II errors and powers of two modified tests are then derived.

The results from the 320 outcomes are obtained for the squared values of $\tilde{J} = J\tilde{A}$ and $\tilde{J}^m = J\tilde{A}^m$. The mean, variance, size, goodness-of-fit measure, and power of $\tilde{J}$ and $\tilde{J}^m$ significantly improve as the sample size increases. For all combinations of R^2 and ρ^2, both tests have the correct size in large samples. The powers of $\tilde{J}^m$ are generally higher than those of $\tilde{J}$, *ceteris paribus*. Powers of the $\tilde{J}$ and $\tilde{J}^m$ tests are reasonably high for those combinations involving n = (100, 60), R^2 = 0.99, 0.95, 0.80) and ρ^2 = (0.30, 0.50). A substantial increase in the powers of $\tilde{J}$ and $\tilde{J}^m$, for these parametric values, occurs when σ_m^2 decreases from 1.0 to 0.25. In general, power is directly related to n and R^2, and inversely related to σ_m^2 and k. One interesting outcome of these experiments involves the two competing roles of ρ^2 in the determination of the power of $\tilde{J}$ and $\tilde{J}^m$. An increase in ρ^2 from 0.30 to 0.50 causes a significant increase in the power of $\tilde{J}$ and $\tilde{J}^m$, while any further increase in ρ^2 (i.e., 0.80 or 0.95) causes power to decline. The former increase in power may be attributed to increased efficiency of the IV estimators of H_0, and the latter decrease in power may be related to the increased closeness between the null and fixed alternative hypotheses. These Monte Carlo results are necessarily limited to the chosen experimental design. Nonetheless, they establish a plausible case for testing non-nested mismeasured regression models using readily available instrumental variables. The computational convenience of $\tilde{J}^m$, supplemented with its generally high power, enhances its practical utility for applied researchers.

References

Berger J O, Mortera J (1999) Default Bayes factors for nonnested hypothesis testing. Journal of American Statistical Association 94:542-554

Bernanke B S, Reiss P C, Bohn H (1986) Alternative nonnested specification tests of time series investment models. Research paper No. 890, Graduate School of Business, Stanford University, USA

Bodla, M.A. and Bhatti, M. I. (2007), Use of Instrumental Variables in Non-nested EV-Models" jointly with M A Bodla, Journal of Statistical Theory and Application, Vol. 6(3), pp. 254-263.

Camilli, G. (2006), Examination of Simple Errors-in-Variable Model: A demonstration of Marginal likelihood", Journal of Educational and Behavioral Statistics, 31(3): 311-325.

Cox D R (1961) Tests of separate families of hypotheses. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 1:105-123

Cox D R (1962) Further results on tests of separate families of hypotheses. Journal of the Royal Statistical Society, Series B, 24:406-424

Davidson R, MacKinnon J G (1981) Several tests for model specification in the presence of alternative hypotheses. Econometrica 49:781-793
Davidson R, MacKinnon J G (2001) Improving the reliability of bootstrap tests. Queen's University Institute for Economic Research Discussion Paper No. 995, revised
Davidson R, MacKinnon J G (2002) Bootstrap J tests of nonnested linear regression models. Journal of Econometrics 109: 167-193
Dubbelman C, Louter A S, Abrahamse A P J (1978) On typical Characteristics of Economic Time Series and the Relative Qualities of Five Autocorrelation Tests. Journal of Economrtrics 8:295-306
Ericsson N R (1986) Post-simulation Analysis of Monte Carlo Experiments: Interpreting Pesaran's (1974) Study of Non-nested Hypothesis Test Statistics. Review of Economic Studies LIII:691-707
Fan Y, Li Q (1995) Bootstrapping J-type tests for non-nested regression models. Economic Letters 48:107-112
Godfrey L G (1998) Tests of non-nested regression models: some results on small sample behaviour and the bootstrap. Journal of Econometrics 84:59-74
Godfrey L G, Pesaran M H (1983) Tests of Non-nested Regression Models: Small Sample Adjustments and Monte Carlo Evidence. Journal of Econometrics 21:133-154
Granger C W J (1966) The Typical Spectral Shape of an Economic Variable. Econometrica 34:150-161.
Haslag G H, Hein S E (1990) Economic activity and two monetary base measures. The Review of Economics and Statistics 72:672-677
Huang, X., Stefanski, L. A., and Davidian, M. (2006), Latent Model Robustness in Structural Measurement Error Models", Biometrika, 93(1), 53-64.
Huth P, Gelpi C, Bennett D S (1993) The escalation of great power militarized disputes: Testing rational deterrence theory and structural realism. American Political Science Review 87:609-623
King G (1989) Unifying political methodology. Cambridge, Cambridge University Press
King, M. L. (1998) The likelihood ratio test of non-nested linear regression models
McAleer M (1995) The significance of testing empirical non-nested models. Journal of Econometrics 67: 149-171
McAleern M (1987) Specification Tests for Separate Models: A Survey, in King, M.L. and D.E.A.Giles (Eds) Specification Analysis in the Linear Model, 146-195
Pesaran M H, Pesaran B (1993) A simulation approach to the problem of computing Cox's statistic for testing nonnested models. Journal of Econometrics 57:377-392
Shumway C R, Gottret P E (1991) Numeraire choice in normalized quadratic functional forms. Applied Econometrics 23: 1597-1602
Silvapulle P, King M L (1993) Nonnested testing for autocorrelation in the linear regression model. Journal of Econometrics 58:295-314