Center for Research on Economic and Social Theory
CREST Working Paper

Testing Hypotheses about Regression Coefficients in Misspecified Models

Dilip K. Bhattacharyya
Jan Kmenta

December 15, 1989
Number 90–5

DEPARTMENT OF ECONOMICS
University of Michigan
Ann Arbor, Michigan 48109

Testing Hypotheses about Regression Coefficients in Misspecified Models

Dilip K. Bhattacharyya and Jan Kmenta*

If in a classical regression model $y=X_1\beta_1 + X_2\beta_2 + U$ the values of X_2 are omitted, the least squares estimator of β_1 is biased (unless $X_1'X_2 = 0$) and the conventional tests of significance concerning β_1 are not valid. The purpose of this paper is to propose a method for estimating the noncentrality parameters of the F distribution involved in testing hypotheses about β_1 when there is no information about $X_2\beta_2$. The estimated noncentrality parameters can be used to approximate the true size of the tests. Further uses of the estimates include choosing the most suitable proxies for X_2 and ranking competing models according to their proximity to the true model. Our theoretical results are supported by a suitably designed Monte Carlo experiment.

KEY WORDS: Misspecified models; Noncentral F distribution; Omitted variables; Regression models; RESET.

***Dilip K. Bhattacharyya is Lecturer, Department of Economics, University of Leicester, Leicester LE1 7RH, England. Jan Kmenta is Professor, Department of Economics, University of Michigan, Ann Arbor, MI 48109-1220. This work was completed while Bhattacharyya was visiting Wayne State University.** **The authors are grateful to Rabi Bhattacharyya and Jayanta Ghosh for comments on an earlier version of the paper.**

December 15, 1989

1. INTRODUCTION

In empirical research most researchers start with a theoretical model showing that an outcome (say Y) depends on many variables. It is also common that the researcher's primary interest lies in studying the effect of subset of explanatory variables on Y. At the same time in many situations some of the explanatory variables are not directly observable, or their measurements are not available, or they are not properly identified by the researcher. Given that the researcher is interested mainly in the interrelationship between Y and the specified set of variables for which measurements are available, a question arises as to what one should do about the variables omitted from the model. There are two obvious choices in this situation; either use proxy variables for the unobserved variables or ignore the existence of the omitted variables and estimate the truncated model without them. However, in both cases the estimators of the coefficients of the included variables will be biased as long as the omitted explanatory variables are not orthogonal to the included explanatory variables and the proxies are not perfect; see, e.g., Kmenta (1986). As orthogonality of data or existence of perfect proxies are not likely in empirical research, McCallum (1972) and Wickens (1972) compared the two biases and concluded, on the basis of the size of the asymptotic bias, that models with proxies for the omitted variables produce better estimates of the coefficients associated with the included variables than the truncated models. However, using a mean square error criterion Aigner (1974) found that under certain conditions the estimated parameters of the truncated model are better; see also Kinal and Lahiri (1983) and Terasvirta (1987). Thus it is not possible to generalize the results in favor of the 'truncated' or the 'proxy' model. It is also important to note that all of the results in the existing literature are derived with the assumption that the proxies represent a linear combination of the omitted variables and a white noise error term. Since this assumption is frequently questionable, we derive our results with proxies without assigning any prior structure to their relationship with the omitted variables.

It is well known that biased estimates also distort the test statistics normally used in testing hypotheses about the respective coefficients. For example, a bias will turn a conventional central χ^2 into a noncentral χ^2. The effect of this change in distribution changes the size of the test. Therefore, one can argue that the larger the noncentrality parameters the more does the estimated model differ from the true model. Hence, it seems reasonable to discriminate among models on the basis of the magnitude of the noncentrality parameters.

In this paper we propose a method of estimating the noncentrality parameters of the F distribution involved in testing hypotheses about the coefficients of the included variables when there is no information about the omitted variables and their coefficients. The estimated noncentrality parameters can be used

a) to approximate the true significance level of the tests;

b) to choose the most suitable proxies for the omitted variables among all available candidates; and

c) to rank competing models according to their estimated proximity to the true model.

All of these uses are important in practical situations. The unreliability of testing hypotheses in misspecified regression models is notorious and any improvement in the standard procedure should be welcome. Further, having a criterion for choosing among all available proxies for the omitted variables when the introduction of proxies is desirable also has a lot of potential use in practice. Finally, having a criterion for ranking competing models on the basis of sample of evidence would certainly add a new dimension to the existing practices of discriminating among models.

The plan of the paper is as follows. In Section 2 we derive the F statistics and their noncentrality parameters for regression models with omitted relevant explanatory variables. In Section 3 we propose a method for estimating the noncentrality parameters derived in

Section 2. Section 4 contains the results of a stylized simulation study designed to illustrate the estimation method of Section 3. The paper ends with some concluding remarks presented in Section 5.

2. DERIVATION OF THE PROPERTIES OF CONVENTIONAL TESTS UNDER MISSPECIFICATION

Suppose the true model is

$$Y = X_1\beta_1 + X_2\beta_2 + U \tag{1}$$

where

$Y \rightarrow (n \times 1)$, $X_1 \rightarrow (n \times K_1)$, $X_2 \rightarrow (n \times K_2)$, $\beta_1 \rightarrow (K_1 \times 1)$, $\beta_2 \rightarrow (K_2 \times 1)$, and $U \rightarrow (n \times 1)$.

Further, we assume that $E(U) = 0$ and, without any loss of generality, that $E(UU') = I_n$. We also assume that U is normally and independently distributed and that X_1 and X_2 are fixed. (If X_1 and X_2 are stochastic, our results can be viewed as conditional on the sample values of these variables.)

According to our presumption the researcher fails to implement the model because X_2 is unobservable. The available alternatives then are:

(i) a truncated model

$$Y = X_1\beta_1 + U_1 \tag{2}$$

or (ii) a proxied model

$$Y = X_1\beta_1 + Z\gamma + U_2 \tag{3}$$

where Z is a (nxK_3) matrix of the chosen proxy variable and γ is a (K_3x1) vector of constants. Thus $Z\gamma$ is used in place of $X_2\beta_2$. Given (1), both models (2) and (3) are, of course, misspecified.

From equation (2) the least squares estimator of β_1 is

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'Y \tag{2a}$$

and its bias is

$$E(\hat{\beta}_1 - \beta_1) = (X_1'X_1)^{-1}X_1'X_2\beta_2. \tag{2b}$$

The matrix mean square error of $\hat{\beta}_1$ is

$$\text{Mtx. MSE}(\hat{\beta}_1) = V(\hat{\beta}_1) + (X_1'X_1)^{-1}X_1'X_2\beta_2\beta_2'X_1(X_1'X_1)^{-1} \tag{2c}$$

where

$$V(\hat{\beta}_1) = (X_1'X_1)^{-1}.$$

On the other hand using equation (3) the estimator of β_1 is

$$\tilde{\beta} = (X_1'M_zX_1)^{-1}X_1'M_zY \tag{3a}$$

and its bias is

$$E(\tilde{\beta}_1-\beta_1) = (X_1'M_zX_1)^{-1}X_1M_z\beta_2,$$

where

$$M_z = I - Z\,(Z'Z)^{-1}Z'$$

The matrix mean square error of $\tilde{\beta}_1$ is

$$\text{Mtx. MSE}(\tilde{\beta}_1) = V(\tilde{\beta}_1) + (X_1'M_zX_1)^{-1}X_1'M_zX_2\beta_2\beta_2'X_2'M_zX_1(X_1'M_zX_1)^{-1} \tag{3c}$$

where

$$V(\tilde{\beta}_1) = (X_1'M_zX_1)^{-1}$$

It is clear from (2a)-(2c) and (3a)-(3c) that the difference in biases and mean square errors depends very largely on Z. The biases in estimating β_1 also bring problems in testing hypotheses concerning β_1. As our parameter vector of interest is β_1, we propose to choose between the alternative misspecified models on the basis of 'test distortion'. We shall demonstrate below that the tests of hypotheses concerning β_1 for both models in (2) and (3) should be based on noncentral distributions. For the true model, of course, the tests of hypotheses concerning β_1 are based on central distributions. Hence one can argue that the further the noncentrality parameters are from zero, the worse is the model as an approximation of the true model.

To perform a suitable test of the null hypothesis H_0: $\beta_1=\beta_1^*$ against H_1: $\beta_1\neq\beta_1^*$ in terms of the truncated model in equation (2), we have

$$F_1 = \frac{(\hat{\beta}_1-\beta_1^*)'(X_1'X_1)(\hat{\beta}_1-\beta_1^*)/K_1}{\hat{U}_1'\hat{U}_1/(n-K_1)} \tag{4}$$

where

$$\hat{U}_1 = Y - X_1\hat{\beta}_1 = M_1Y \text{ and } M_1 = I - X_1(X_1'X_1)^{-1}X_1'.$$

The researcher would generally regard F_1 as having a central F distribution with $(K_1, n\text{-}K_1)$ degrees of freedom.

In terms of the true model (1) we have, under H_0,

$$(\hat{\beta}_1 - \beta_1^*)'(X_1'X_1)(\hat{\beta}_1 - \beta_1^*) = (U + X_2\beta_2)'(I\text{-}M_1)(U + X_2\beta_2) \tag{4a}$$

and

$$\hat{U}_1'\hat{U}_1 = (U + X_2\beta_2)'M_1(U+X_2\beta_2). \tag{4b}$$

When H_0 is true, we know from standard theorems (for references see, e.g., Bhattacharyya (1985)) the expression in (4a) is distributed as noncentral χ^2 with degrees of freedom given by $Tr(I\text{-}M_1)$ and with the noncentrality parameter equal to $\beta_2'X_2'(I\text{-}M_1)X_2\beta_2$. Similarly the expression in (4b) is distributed as noncentral χ^2 with degrees of freedom given by $Tr(M_1)$ and with the noncentrality parameter equal to $\beta_2'X_2'M_1X_2\beta_2$. Hence the expression for F_1 in (4) is distributed as doubly noncentral F with noncentrality parameters $[\beta_2'X_2'(I\text{-}M_1)X_2\beta_2, \beta_2'X_2'M_1X_2\beta_2]$ and degrees of freedom $[K_1, n\text{-}K_1]$. It is clear that the researcher can use the correct degrees of freedoms but cannot calculate the noncentrality parameters without the knowledge of $X_2\beta_2$. If the values of noncentrality parameters are known then following

Johnson and Kotz (1970, pp. 178–179), we can approximate F_1 as follows:

$$F_1 \simeq \frac{(n-K_1)[K_1+\beta_2'X_2'(I-M_1)X_2\beta_2]}{K_1\ [(n-K_1)+\beta_2'X_2'M_1X_2\beta_2]}\ F_{n_1,n_2}, \tag{4c}$$

where

$$n_1 = \frac{[K_1+\beta_2'X_2'(I-M_1)X_2\beta_2]^2}{K_1+2\beta_2'X_2'(I-M_1)X_2\beta_2} \tag{4d}$$

and

$$n_2 = \frac{[(n-K_1)+\beta_2'X_2'M_1X_2\beta_2]^2}{(n-K_1)+2\beta_2'X_2'M_1X_2\beta_2}, \tag{4e}$$

and F_{n_1,n_2} stands for central F distribution with n_1 and n_2 degrees of freedom.

Similarly, to test H_0: $\beta_1=\beta_1^*$ against H_1: $\beta_1\neq\beta_1^*$ using the proxied model of equation (3), the research would generally use,

$$F_2 = \frac{(\tilde{\beta}_1-\beta_1^*)'(X_1'M_zX_1)(\tilde{\beta}_1-\beta_1^*)/K_1}{\tilde{U}_2'\tilde{U}_2/(n-K_1-K_3)} \tag{5}$$

where $\tilde{U}_2 = Y - X_1\tilde{\beta}_1 - Z\tilde{\gamma}$ and $\tilde{\gamma} = (Z'M_1Z)^{-1}(Z'M_1Y)$.

Thus, under H_0

$$(\hat{\beta}_1 - \beta_1^*)'(X_1' M_z X_1)(\hat{\beta}_1 - \beta_1^*) = (U_2 + X_2\beta_2)'M(U + X_2\beta_2), \tag{5a}$$

where

$$M = M_z X_1 (X_1' M_z X_1)^{-1} X_1' M_z$$

is distributed as noncentral χ^2 with $Tr(M) = K_1$ degrees of freedom and with the noncentrality parameter equal to $\beta_2' X_2 M X_2 \beta_2$.

Further,

$$\hat{U}_2' \hat{U}_2 = (U_2 + X_2\beta_2)'(M_z - M)(U + X_2\beta_2)$$

is distributed as noncentral χ^2 with degrees of freedom equal to $tr(M_z - M) = n - K_1 - K_3$ and noncentrality parameters equal to $\beta_2' X_2' (M_z - M) X_2 \beta_2$.

Again, when the noncentrality parameters are known, the doubly noncentral distribution of F_2 can be approximated as follows:

$$F_2 \simeq \frac{[(n - K_1 - K_2)[K_1 + \beta_2' X_2' M X_2 \beta_2]}{K_1[(n - K_1 - K_2) + 2\beta_2' X_2' (M_z - M) X_2 \beta_2]} \; F_{m_1, m_2} \tag{5c}$$

where

$$m_1 = \frac{(K_1 + \beta_2' X_2' M X_2 \beta_2)^2}{K_1 + 2\beta_2' X_2' M X_2 \beta_2} \tag{5d}$$

and

$$m_2 = \frac{[(n-K_1-K_2) + \beta_2' X_2' (M_z - M) X_2 \beta_2]^2}{(n-K_1-K_2) + 2\beta_2' X_2' (M_z - M) X_2 \beta_2}, \tag{5e}$$

and F_{m_1, m_2} is distributed as central F with m_1 and m_2 degrees of freedom.

It may be useful to note that a test of hypothesis H_0: $\gamma=\gamma^*$ against H_1; $\gamma \neq \gamma^*$ will also lead to a doubly noncentral F distribution.

From the above results, it is clear that in order to obtain the correct significance level for a test it is necessary to know the noncentrality parameters. In the next section we suggest a procedure for obtaining consistent estimate of these parameters.

3. AN EMPIRICAL METHOD OF ESTIMATING NONCENTRALITY PARAMETERS

Consider the model with proxy variables in equation (3), i.e.,

$$Y = X_1\beta_1 + Z\gamma + U_2 .$$

We define 'pseudo' least squares residuals as

$$\hat{V} = Y - X_1\tilde{\beta}_1 \tag{6}$$

where

$$\tilde{\beta}_1 = (X_1'M_zX_1)^{-1}X_1'M_zY$$

Substituting for Y from equation (1) into (6) we find that

$$\hat{V} = N(X_2\beta_2) + NU \tag{7}$$

where

$$N = I-X_1(X_1'M_zX_1)^{-1}X_1'M_z \tag{8}$$

Hence, we have

$$E(\hat{V}) = NX_2\beta_2$$

Given the assumptions of the classical regression model concerning X and Z, it can be shown that for large samples the matrix N converges towards I.

A consistent estimator of $X_2\beta_2$ can then be found with the help of the following theorems.

<u>Theorem 1</u>:

Let $Ax = b$ represent a mathematically inconsistent system that may have no solution. This system may then be expressed as $r(x)=Ax - b$. An approximate solution of the system by the least squares method is given by $x^* = Bb$
where B satisfies the following conditions: (i) $ABA = A$, and (ii) AB is symmetric.

Proof: See Dhrymes (1978), Appendix 3.5.

Theorem 2: **(Singular Value Decomposition)**

Let A be a (n×n) matrix of rank (n-k). Then there exist (n×n) orthogonal matrices H and V and a (n×n) diagonal matrix S such that

$H'AV = S$ and $A = HSV'$

The successive diagonal entries of S are positive and non-increasing. The matrix S takes the form

$$S = \begin{bmatrix} S_{11} & 0 \\ 0 & 0 \end{bmatrix}$$

where S_{11} is (n-k)×(n-k) matrix of rank (n-k).

Proof: See Lawson and Hanson (1974), pp. 20–21.

Theorem 3:

The solutions to the problem of minimizing $||Ax-b||$ (the least squares solutions) are of the form

$$\hat{X} = V \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{bmatrix}$$

where $\tilde{y}_2$ is arbitrary. Further, $\tilde{y}_1$ is the unique solution of

$$S_{11}\, y_1 = g_1$$

where g_1 is obtained from

$$H'b = g = \begin{bmatrix} g_1 \\ \\ g_2 \end{bmatrix} \begin{matrix} n-k \\ \\ k \end{matrix}$$

<u>Proof</u>: See Lawson and Hanson (1974), pp. 5–7.

Thus by minimizing $||\hat{V}\text{-}NX_2\beta_2||$ we can obtain the least squares solution for equation (7). This particular procedure, as justified by the three theorems above, is used in our empirical work. Let the estimate of $X_2\beta_2$, given by as the least squares solution of $||\hat{V}\text{-}NX_2\beta_2||$, be called Q. Then the following estimates of biases and of the noncentrality parameters are available using the observed variables:

(I) $\widehat{\text{Bias}}$ 1: $(X_1'X_1)^{-1}X_1'Q$

(II) $\widehat{\text{Bias}}$ 2: $(X_1'M_zX_1)^{-1}X_1'M_zQ$

(III) $\hat{\delta}_{11}$: $Q'(I\text{-}M_1)Q$

(IV) $\hat{\delta}_{12}$: $Q'M_1Q$

(V) $\hat{\delta}_{21}$: $Q'MQ$

(VI) $\hat{\delta}_{22}$: $Q'(M_z\text{-}M)Q$

Here (I) and (II) are obtained from equations (2b) and (3b), δ_{11} and δ_{12} are the noncentrality parameters of the respective numerator and denominator of F_1 in (4), and δ_{21} and δ_{22} are the noncentrality parameters of the respective numerator and denominator of F_2 in (5).

Using (III) through (VI), one can calculate the approximate values of the doubly noncentral F's in terms of the standard F distribution, as shown in equations (4c) and (5c). With respect to choosing between the truncated and the proxied model, suppose we find

$$\hat{\delta}_{11} < \hat{\delta}_{21} \text{ and } \hat{\delta}_{12} < \hat{\delta}_{22}.$$

Then, for the given set of data, the truncated model is closer to the true model. If the inequality is reversed, then the proxied model is preferable to the truncated model. However, these easy choice situations may not arise and we may find that the two inequalities under consideration appear in different directions. In that case our decision can be made in terms of the distance from the origin, i.e.,

$$\Delta_1 = \hat{\delta}_{11}^2 + \hat{\delta}_{12}^2 \text{ and } \Delta_2 = \hat{\delta}_{21}^2 + \hat{\delta}_{22}^2$$

where the model with lower Δ values is to be preferred.

4. MONTE CARLO RESULTS

To illustrate the usefulness of our approach in empirical situations, we carry out a small Monte Carlo experiment based on a design that has been used in previous studies dealing with testing for specification errors. Our procedure applies to any choice of proxy variables for which N converges to identity, but for designing our experiment we have to make a particular choice. Since we wish to use a reasonably generic model, we use uninformative proxies as proposed by Ramsey (1969) and modified by Thursby and Schmidt (1977). The design of the experiment is also taken from Thursby and Schmidt (1977). The true model is assumed to be

$$Y_t = 10 + 5\,X_{1t} - 2\,X_{2t} + U_t;\ U_t \sim \text{NID}(0,1). \tag{A}$$

In accordance with our earlier discussion, the truncated model is

$$Y_t = a_1 + b_1X_{1t} + U_{1t}, \tag{B}$$

and the proxy model is

$$Y_t = a_2 + b_2X_{1t} + c_1Z_{1t} + c_2Z_{2t} + c_3Z_{3t} + U_{2t}, \tag{C}$$

where

$$Z_{1t} = X_{1t}^2,\ Z_{2t} = X_{1t}^3 \text{ and } Z_{3t} = X_{1t}^4.$$

This particular choice of proxies is recommended by Thursby and Schmidt (1977) who found them to be, in general, the most satisfactory in comparison with other alternatives within the RESET principle as discussed by Ramsey (1969).

In conducting the experiment we generated fixed samples of sizes 30 and 50. The values of X_1 and X_2 were chosen to be the same as those used by Thursby and Schmidt (1977) and Ramsey and Gilbert (1972). They are given as follows:

X_1 = 1.0, 3.7, 0.8, 9.9, 1.2, 6.6, 3.1, 8.5, 6.3, and 7.3

X_2 = 9.0, 9.0, 8.0, 1.0, 8.0, 6.0, 6.0, 8.0, 12.0, and 16.0.

We replicated three or five times the basic 10 observations to obtain sample sizes 30 and 50, respectively. The actual values of biases and the values of noncentrality parameters are calculated using these fixed values of X_1 and X_2 and the parameters of the true model in (A). These results are presented in Table 1. Finally, the random values of Y_t were generated using the NAG (Numerical Algorithm Group) library routine for the given values of X_1 and X_2 and model (A). The estimation and calculations are done as described in section 3. The experiments were repeated 1000 times for each sample size.

The figures in Table 1 are quite informative and can be taken as a guideline for empirical study. First, we observe that the values of the noncentrality parameters are higher for sample size 50 than for sample size 30. Further, the noncentrality parameter associated with the denominator of the truncated (OVM) model is higher than that for the proxied (PVM) model, whereas the noncentrality parameter associated with the numerator of the test statistic is lower for the OVM than for the PVM model. The calculated biases have higher values for the PVM than for the OVM specification.

With respect to our Monte Carlo experiment, we first obtained the least squares solution of $||\hat{V}-NX_2\beta_2||$ where N is the observed matrix defined in (8). The solution procedure is explained in connection with Theorems 2 and 3. To estimate $X_2\beta_2$, we used the program F04JAF in the NAG library. The averages of 1000 replications are presented in Table 1. The estimates of the biases and of the noncentrality parameters (see expressions (I) through (VI) in Section 3) are close to the actual values. It is interesting to note that both models used, OVM and PVM, perform well in estimating the noncentrality parameter of the denominator of the F statistic, but that the proxied model (PVM) does better in estimating the noncentrality parameter of the numerator than the truncated model (OVM). Also, in terms of the distance measure Δ discussed in Section 3 the proxied model performs better than the truncated model.

If the noncentrality parameters are ignored and the conventional critical values of the F statistic are used, the critical points will differ from the conventional ones. However, this divergence is substantially more pronounced for large values of the noncentrality parameter of the <u>numerator</u> (δ_1) of the F statistic than for large values of the noncentrality parameter of the denominator (δ_2). The magnitude of the difference between the true and the conventional value of the F statistic is much more sensitive to increases in δ_1 than to increases in δ_2. This can be clearly seen from Table 2 in which some stylized results are presented. Hence it is considerably more important to have a good estimate of δ_1 than of δ_2. On this count the proxied model performs better than the truncated model. Thus according to the results of our experiment the proxied model is to be recommended over the truncated model when it comes to testing hypothesis about β_1, the coefficients of the included variables.

5. OTHER APPLICATIONS OF THE ESTIMATED NONCENTRALITY PARAMETERS

In the preceding sections we concentrated on the problem of estimating the noncentrality parameters associated with the F statistic for the truncated and the proxy model. The implied purpose was to carry out hypotheses concerning β_1. However, the estimated noncentrality parameters have a wider potential use when it comes to model selection. For instance, suppose the researcher is faced with a choice between two sets of proxy variables Z_1 and Z_2. On the presumption supported by our earlier results that the proxied model is to be preferred to the truncated model, one can calculate the suggested pseudo-residuals using Z_1 and Z_2 in turn, i.e.,

$$\hat{V} = Y - X_1\tilde{\beta}_{11}$$

where

$$\tilde{\beta}_{11} = (X_1' M_{Z_1} X_1)^{-1} X_1' M_{Z_1} Y.$$

$$M_{Z_1} = I - Z_1 (Z_1' Z_1)^{-1} Z_1',$$

and

$$\tilde{V}_2 = Y - X_1\tilde{\beta}_{12}$$

$$\tilde{\beta} = (X_1' M_{Z_2} X_1)^{-1} X_1' M_{Z_2} Y$$

$$M_{Z_2} = I - Z_2(Z_2' Z_2)^{-1} Z_2'.$$

Then

$$\hat{V}_i = N_i(X_2\beta_2 + U) \tag{9}$$

and

$$N_i = I - X_1(X_1' M_{Z_i} X_1)^{-1} X_1' M_{Z_i} \qquad \text{for } i = 1, 2\,.$$

Using $\hat{V}_i$ (i = 1, 2), two alternative least squares estimates of $X_2\beta_2$ can be derived, yielding two sets of estimates of the respective noncentrality parameters. Then the set of proxy variables that produces a lower distance measure Δ can be regarded as providing a better approximation to the true model for the given sample size. To allow for sampling fluctuations, an appropriate test of significance could be developed.

Let us consider the situation when we are concerned about deciding between two nonnested models:

Model I: $\qquad Y = X_1\beta_1 + U_1,$

Model II: $\qquad Y = X_2\beta_2 + U_2.$

Neither of the two models can be asserted to be true. In this case we can choose a set of proxy variables Z that could be used for both models. Given Z we can obtain two sets of pseudo-residuals from the two competing models. Using the procedure outlined in Section 3 above, we can then calculate two sets of estimates of the respective noncentrality parameters, yielding two distance measures Δ. The model with the lower distance measure can then be regarded as being closer to the true model.

Table 1.
ACTUAL AND ESTIMATED VALUES OF THE NONCENTRALITY PARAMETERS AND BIASES

	Model	Sample Size	δ_1	δ_2	Bias	Δ
ACTUAL	OVM	30	23.464	1599.13	0.2812	1599.30
	PVM	30	954.695	667.90	1.7476	1165.14
	OVM	50	39.107	2665.21	0.2812	2665.50
	PVM	50	1591.159	1113.16	1.7476	1941.88
ESTIMATED	OVM	30	0.00	1629.78	0.000	1629.78
	PVM	30	935.69	694.09	1.2828	1165.02
	OVM	50	0.00	2713.07	0.000	2713.07
	PVM	50	1555.36	1157.71	1.5569	1938.93

δ_1: Noncentrality parameter associated with the numerator of the test statistic;

δ_2: Noncentrality parameter associated with the denominator of the test statistic;

Δ: Square root of the distance measure as reported in (8);

OVM: Truncated (omitted variable) model;

PVM: Proxied (proxy variable) model.

Table 2
EFFECT OF THE NONCENTRALITY PARAMETERS ON THE 5% CRITICAL POINTS

DF1	DF2	δ_1	δ_2	m_1	m_2	CF	$F_{5\%}(m_1,m_2)$	True $F_{5\%}$
1	29	0	0	1	29	1	4.183	4.183
1	29	50	1500	26	772	0.9672	1.5	1.451
1	29	1500	50	751	48	.551	1.504	822.70
1	60	0	0	1	60	1	4.0012	4.0012
1	60	50	1500	26	795	1.9301	1.6427	3.1704
1	60	1500	50	751	75	812.47	1.3554	1101.22
5	60	0	0	5	60	1	2.3683	2.3683
5	60	50	1500	29	792	0.3891	1.4037	0.5462
5	60	1500	50	754	71	157.67	1.3642	215.09

DF1 = Degrees of freedom of the numerator

DF2 = Degrees of freedom of the denominator

δ_1 = Noncentrality parameter of the numerator

δ_2 = Noncentrality parameter of the denominator

m_1 = Degrees of freedom of the numerator for the approximate noncentral F distribution

m_2 = Degrees of freedom of the denominator for the approximate noncentral F distribution

CF = Multiplicative correction factor for the approximate F distribution

$F_{5\%}(m_1,m_2)$ = Table value from Biometrika table.

True $F_{5\%}$ = CF multiplied by the table value of F (see Section 3).

REFERENCES

Aigner, D. J. (1974). "MSE Dominance of Least Squares with Errors of Observation", Journal of Econometrics, 2, 365–372.

Bhattacharyya, D. K. (1985). "A Clarifying Note on the Effects of Linear Approximation on Hypothesis Testing", Economics Letters, 17. 231–235.

Dhrymes, P. J. (1978). Introductory Econometrics, New York: Springer-Verlag Inc..

Johnson, N. L. and Kotz, S. (1970). Continuous Univariate Distributions – 2, Boston: Houghton-Mifflin Co.

Kinal, T. and Lahiri, K. (1983). "Specification Error Analysis with Stochastic Regressions", Econometrica 51, 1209–1218.

Kmenta, J. (1986). Elements of Econometrics, 2nd ed., New York: Macmillan.

Lawson, C. L. and Hanson, R. J. (1974). Solving Least Squares Problems, Englewood Cliffs: Prentice-Hall, Inc..

McCallum, B. T. (1972). "Relative Asymptotic Bias from Errors of Omission and Measurement", Econometrica, 40, 757–758.

Ramsey, J. B. (1969). "Test for Specification Errors in Classical Linear Least-Squares Regression Analysis", Journal of Royal Statistical Society, Series B, 31, 350–371.

Ramsey, J. B. and Gilbert, R. F. (1972). "A Monte Carlo Study of Some Small Sample Properties of Tests for Specification Error", Journal of the American Statistical Association, 67, 180–186.

Thursby, J. G. and Schmidt, P. (1977). "Some Properties of Tests for Specification Error in a Linear Regression Model", Journal of the American Statistical Association, 72, 636–641.

Terasvirta, T. (1987). "Usefulness of Proxy Variables in Linear Models with Stochastic Regressions," Journal of Econometrics, 36, 377–382.

Wickens, M. R. (1972). "A Note on the Use of Proxy Variables", Econometrica, 40, 759–761.

Recent CREST Working Papers

89–01: Mark Bagnoli, Severin Borenstein, "Carrot and Yardstick Regulation: Enhancing Market Performance with Output Prizes," October, 1988.

89–02: Ted Bergstrom, Jeffrey K. MacKie–Mason, "Some Simple Analytics of Peak-Load Pricing," October, 1988.

89–03: Ken Binmore, "Social Contract I: Harsanyi and Rawls," June, 1988.

89–04: Ken Binmore, "Social Contract II: Gauthier and Nash," June, 1988.

89–05: Ken Binmore, "Social Contract III: Evolution and Utilitarianism," June, 1988.

89–06: Ken Binmore, Adam Brandenburger, "Common Knowledge and Game Theory," July, 1989.

89–07: Jeffrey A. Miron, "A Cross Country Comparison of Seasonal Cycles and Business Cycles," November, 1988.

89–08: Jeffrey A. Miron, "The Founding of the Fed and the Destabilization of the Post-1914 Economy," August, 1988.

89–09: Gérard Gaudet, Stephen W. Salant, "The Profitability of Exogenous Output Contractions: A Comparative-Static Analysis with Application to Strikes, Mergers and Export Subsidies," July, 1988.

89–10: Gérard Gaudet, Stephen W. Salant, "Uniqueness of Cournot Equilibrium: New Results from Old Methods," August, 1988.

89–11: Hal R. Varian, "Goodness-of-fit in Demand Analysis," September, 1988.

89–12: Michelle J. White, "Legal Complexity," October, 1988.

89–13: Michelle J. White, "An Empirical Test of the Efficiency of Liability Rules in Accident Law," November, 1988.

89–14: Carl P. Simon, "Some Fine-Tuning for Dominant Diagonal Matrices," July, 1988.

89–15: Ken Binmore, Peter Morgan, "Do People Exploit Their Bargaining Power? An Experimental Study," January, 1989.

89–16: James A. Levinsohn, Jeffrey K. MacKie–Mason, "A Simple, Consistent Estimator for Disturbance Components in Financial Models," April 25, 1989.

89–17: Hal R. Varian, "Sequential Provision of Public Goods," July, 1989.

89–18: Hal R. Varian, "Monitoring Agents with Other Agents," June, 1989.

89–19: Robert C. Feenstra, James A. Levinsohn, "Distance, Demand, and Oligopoly Pricing," July 17, 1989.

89–20: Mark Bagnoli, Shaul Ben-David, Michael McKee, "Voluntary Provision of Public Goods," August, 1989.

89–21: N. Gregory Mankiw, David Romer, Matthew D. Shapiro, "Stock Market Forecastability and Volatility: A Statistical Appraisal," August, 1989.

89–22: Arthur J. Robson, "Efficiency in Evolutionary Games: Darwin, Nash and the Secret Handshake," 1989.

89–23: Mark Bagnoli, Ted Bergstrom, "Log-Concave Probability and Its Applications," September 7, 1989.

89–24: Gérard Gaudet, Stephen W. Salant, "Towards a Theory of Horizontal Mergers," July, 1989.

89–25 (evolved from 87-35): Stephen W. Salant, Eban Goodstein, "Predicting Committee Behavior in Majority-Rule Voting Experiments," July, 1989.

89–26: Ken Binmore, Martin J. Osborne, Ariel Rubinstein, "Noncooperative Models of Bargaining," 1989.

89–27: Avery Katz, "Your Terms or Mine? The Duty to Read the Fine Print in Contracts," February 19, 1989.

Recent CREST Working Papers

90–01: Hal R. Varian, "A Solution to the Problem of Externalities and Public Goods when Agents are Well–Informed," January 17, 1990.

90–02: Hal R. Varian, "Sequential Provision of Public Goods," January 17, 1990.

90–03: Hal R. Varian, "Goodness-of-Fit Optimizing Models," December 11, 1989.

90–04: Paul N. Courant, Alan Deardorff, "International Trade with Lumpy Countries," August 22, 1989.

90–05: Dilip K. Bhattacharyya, Jan Kmenta, "Testing Hypotheses about Regression Coefficients in Misspecified Models," December 15, 1989.

90–06: Greg Shaffer, "Slotting Allowances and Resale Price Maintenance: A Comparison of Facilitating Practices," November 1989.

90–07: Michelle J. White, "Why are Taxes So Complex and Who Benefits?" January 19, 1990.

90–08: Daniel P. O'Brien, "Endogenous Price Leadership: A Bargaining Model of International Telecommunications Settlements," May 1989.

90–09: Mark Bagnoli, Barton L. Lipman, "Stock Price Manipulation Through Takeover Bids," November 1989.

90–10: Ken Binmore, "Bargaining Theory Without Tears," December 1, 1989.

90–11: Hyun Song Shin, "Non-Partitional Information On Dynamic State Spaces and the Possibility of Speculation," December 1989.