



Queen's Economics Department Working Paper No. 954

Bootstrap Tests of Nonnested Linear Regression Models

Russell Davidson
GREQAM, Queen's University

James G. MacKinnon
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

09-1997

Bootstrap Tests of Nonnested Linear Regression Models

by

Russell Davidson

GREQAM
Centre de la Vieille Charité
2 rue de la Charité
13002 Marseille, France

Department of Economics
Queen's University
Kingston, Ontario, Canada
K7L 3N6

russell@ehess.cnrs-mrs.fr

and

James G. MacKinnon*

Department of Economics
Queen's University
Kingston, Ontario, Canada
K7L 3N6

jgm@qed.econ.queensu.ca

Abstract

The J test for nonnested regression models often works badly as an asymptotic test, but it generally works very well when bootstrapped. We provide a theoretical analysis of the J test which explains both of these phenomena. We also propose a modified version of the test which works even better than the ordinary J test when bootstrapped. Using our theoretical results to make simulation much faster, we obtain extremely accurate Monte Carlo results which demonstrate just how well the bootstrapped tests perform.

* Corresponding author.

This research was supported, in part, by grants from the Social Sciences and Humanities Research Council of Canada.

September, 1997

1. Introduction

Numerous procedures for testing nonnested regression models have been developed, directly or indirectly, from the pathbreaking work of Cox (1961, 1962). The most widely used, because of its simplicity, is the J test proposed in Davidson and MacKinnon (1981); see McAleer (1995) for evidence on this point. Like almost all nonnested hypothesis tests, the J test is not exact in finite samples. Indeed, as has been documented in numerous Monte Carlo experiments, its finite-sample distribution can be very far from its asymptotic one.

Numerous ways have been proposed to improve the finite-sample properties of the J test. Fisher and McAleer (1981) proposed a variant of the J test, called the J_A test, which is exact in finite samples under the usual conditions for t tests in linear regression models to be exact; see Godfrey (1983). Unfortunately, the J_A test is often very much less powerful than other nonnested tests; see, among others, Davidson and MacKinnon (1982) and Godfrey and Pesaran (1983). The latter paper suggested a different approach, applied not to the J test but to variants of the Cox test based on the work of Pesaran (1974). This approach first corrects the bias in the numerator of the test statistic, then estimates the variance of the corrected numerator, and finally calculates a t -like statistic. It does not yield exact tests, but it does yield tests that perform considerably better than the J test under the null and have good power.

More recently, Fan and Li (1995) and Godfrey (1997) have suggested bootstrapping the J test and other nonnested hypothesis tests. Because the J test is cheap and easy to compute, this is very easy to do. The Monte Carlo results in these papers suggest that bootstrapping the J test works very well, although they are not based on enough replications to be definitive. Moreover, neither paper provides any explanation of why it apparently works so well.

In this paper, we show precisely what determines the finite-sample distribution of the J test. We explain why the J test often works badly without bootstrapping and why it generally works well when bootstrapped. We also propose a modified J test which works extraordinarily well when bootstrapped. The performance of the J and modified J tests when bootstrapped is investigated by means of Monte Carlo experiments. Because the tests perform so well, these experiments need to be very accurate. By using the theoretical results of the paper in the design of the Monte Carlo experiments, we are able to use extremely large numbers of replications at low cost, and we are thus able to obtain very accurate results. Our results are much more accurate than any published Monte Carlo results on any form of bootstrap test of which we are aware.

2. The J Test

Although the J test can be applied to both linear and nonlinear regression models, we shall, like almost all the papers that have dealt with its finite-sample properties,

consider only the linear case. Suppose that there are two nonnested, linear regression models with i.i.d. normal errors:

$$\begin{aligned} H_1: \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad \mathbf{u} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \quad \text{and} \\ H_2: \mathbf{y} &= \mathbf{Z}\boldsymbol{\gamma} + \mathbf{v}, \quad \mathbf{v} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \end{aligned} \tag{1}$$

where \mathbf{y} , \mathbf{u} , and \mathbf{v} are $n \times 1$, \mathbf{X} and \mathbf{Z} are $n \times k$ and $n \times l$, respectively, $\boldsymbol{\beta}$ is $k \times 1$, and $\boldsymbol{\gamma}$ is $l \times 1$. The J test statistic is the ordinary t statistic for $\alpha = 0$ in the regression

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \alpha \mathbf{P}_Z \mathbf{y} + \text{residuals}, \tag{2}$$

where $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top$. Thus $\mathbf{P}_Z \mathbf{y}$ is the vector of fitted values from OLS estimation of the H_2 model. Asymptotically, this test statistic is distributed as $N(0, 1)$ under H_1 . In practice, the $t(n - k - 1)$ distribution is generally used for finite-sample inference although, except in one special case, there is no formal justification for doing so.

If \acute{s} denotes the estimated standard error from regression (2), the J statistic for testing H_1 can be written as

$$J = \frac{\mathbf{y}^\top \mathbf{P}_Z \mathbf{M}_X \mathbf{y}}{\acute{s}(\mathbf{y}^\top \mathbf{P}_Z \mathbf{M}_X \mathbf{P}_Z \mathbf{y})^{1/2}}, \tag{3}$$

where $\mathbf{P}_X \equiv \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ and $\mathbf{M}_X \equiv \mathbf{I} - \mathbf{P}_X$. Thus $\mathbf{M}_X \mathbf{y}$ is the vector of residuals from OLS estimation of the H_1 model. In the special case in which \mathbf{Z} has only one column that is not in $\mathcal{S}(\mathbf{X})$, the subspace spanned by the columns of \mathbf{X} , J is distributed as $t(n - k - 1)$. Except in this special case, it does not have a known finite-sample distribution, and the mean of J is nonzero. This nonzero mean appears to be the principal reason why the J test often performs poorly in finite samples. Under the null that H_1 generated the data, the numerator of (3) is

$$\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{P}_Z \mathbf{M}_X \mathbf{u} + \mathbf{u}^\top \mathbf{P}_Z \mathbf{M}_X \mathbf{u}, \tag{4}$$

and the second factor in the denominator is the square root of

$$\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{P}_Z \mathbf{M}_X \mathbf{P}_Z \mathbf{X} \boldsymbol{\beta} + 2\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{P}_Z \mathbf{M}_X \mathbf{P}_Z \mathbf{u} + \mathbf{u}^\top \mathbf{P}_Z \mathbf{M}_X \mathbf{P}_Z \mathbf{u}. \tag{5}$$

The first term in (4) has mean zero, but the second term has mean $\sigma^2 \text{Tr}(\mathbf{P}_Z \mathbf{M}_X)$, which will be greater than zero except in the special case just mentioned. This is the source of the bias in J .

Because the first term in (4) is $O(n^{1/2})$ and the second term is $O(1)$, the bias in the numerator vanishes asymptotically. Similarly, because the first term in (5) is $O(n)$, the second is $O(n^{1/2})$, and the third is $O(1)$, the denominator of (3) tends to the square root of the variance of the numerator as $n \rightarrow \infty$. For finite-sample performance to be good, we want the first term in (5) to be large relative to the remaining terms in (5) and relative to the second term in (4). This will be the case if the sum of the squared elements of $\mathbf{M}_X \mathbf{P}_Z \mathbf{X} \boldsymbol{\beta}$ is large relative to $\sigma^2 \text{Tr}(\mathbf{P}_Z \mathbf{M}_X)$.

3. The Distribution of the J Test

In this section, we examine precisely what determines the distribution of the J test under the assumption that the error vector \mathbf{u} is normally and independently distributed. The normality assumption may seem to be a strong one, but some distributional assumption is always essential if we are to say anything about finite-sample distributions. We are not able to obtain a closed-form expression for the distribution of the J test, but we are able to show that J depends in a fairly simple way on a few parameters and a few independent random variables with known distributions.

It is possible, and indeed likely, that the intersection of $\mathcal{S}(\mathbf{X})$ and $\mathcal{S}(\mathbf{Z})$ may have dimension greater than zero. Let this dimension be k_0 , and let the intersection be spanned by the columns of the $n \times k_0$ matrix \mathbf{X}_0 , which can be chosen without loss of generality to be such that $\mathbf{X}_0^\top \mathbf{X}_0 = N\mathbf{I}$, where $N = O(n)$ is a normalization factor that will be chosen later. As we shall see, the space $\mathcal{S}(\mathbf{X}_0)$ has no impact on the J statistic, other than reducing the number of degrees of freedom. The distribution will be just the same as that of a J statistic for which $\mathcal{S}(\mathbf{X})$ is replaced by the orthogonal complement of $\mathcal{S}(\mathbf{X}_0)$ in $\mathcal{S}(\mathbf{X})$, $\mathcal{S}(\mathbf{Z})$ is replaced by the orthogonal complement of $\mathcal{S}(\mathbf{X}_0)$ in $\mathcal{S}(\mathbf{Z})$, and n is replaced by $n - k_0$. In general, let the former of these orthogonal complements be spanned by the $k_1 \equiv k - k_0$ columns of the matrix \mathbf{X}_1 , and the latter by the $l - k_0$ columns of \mathbf{Z}_1 . Without loss of generality, suppose that

$$\mathbf{X}_1^\top \mathbf{X}_1 = N\mathbf{I}_{k_1}, \quad \mathbf{Z}_1^\top \mathbf{Z}_1 = N\mathbf{I}_{l-k_0},$$

By construction, $\mathbf{X}_0^\top \mathbf{X}_1 = \mathbf{0}$, and $\mathbf{X}_0^\top \mathbf{Z}_1 = \mathbf{0}$.

Now consider the k_1 columns of $\mathbf{P}_Z \mathbf{X}_1$. Since $\mathbf{P}_Z = \mathbf{P}_{X_0} + \mathbf{P}_{Z_1}$, and since $\mathbf{X}_0^\top \mathbf{X}_1 = \mathbf{0}$ by construction, we find that $\mathbf{P}_Z \mathbf{X}_1 = \mathbf{P}_{Z_1} \mathbf{X}_1$. The dimension of the span of these columns cannot exceed k_1 , but it may be less. In particular, it must be less if $l - k_0 < k_1$. To deal conveniently with this, consider the subspace of $\mathcal{S}(\mathbf{X}_1)$ annihilated by \mathbf{P}_Z , that is, the subspace of $\mathcal{S}(\mathbf{X})$ orthogonal to $\mathcal{S}(\mathbf{Z})$. Let the dimension of this space be k_3 , and let it be generated by the k_3 columns of \mathbf{X}_3 , with, as usual, $\mathbf{X}_3^\top \mathbf{X}_3 = N\mathbf{I}$. It is immediate that $\mathbf{X}_3^\top \mathbf{X}_0 = \mathbf{0}$. As with $\mathcal{S}(\mathbf{X}_0)$, the space $\mathcal{S}(\mathbf{X}_3)$ will turn out to have no impact on the distribution of the J statistic other than reducing the degrees of freedom.

We may now introduce a matrix \mathbf{X}_2 , orthogonal to both \mathbf{X}_0 and \mathbf{X}_3 and normalized as usual so that $\mathbf{X}_2^\top \mathbf{X}_2 = N\mathbf{I}$, such that its columns, along with those of \mathbf{X}_3 , span $\mathcal{S}(\mathbf{X}_1)$. This implies that the columns of \mathbf{X}_0 , \mathbf{X}_2 , and \mathbf{X}_3 span all of $\mathcal{S}(\mathbf{X})$, the three matrices being mutually orthogonal. Let us denote by l_2 the number of columns of \mathbf{X}_2 . Thus $k = k_0 + l_2 + k_3$. Note at this point that \mathbf{X}_2 is an arbitrary orthogonal basis, appropriately normalized, of the space spanned by its columns. Later, however, we will make a specific choice of \mathbf{X}_2 .

It now follows that the space spanned by the l_2 columns of $\mathbf{P}_Z \mathbf{X}_2 = \mathbf{P}_{Z_1} \mathbf{X}_2$ is of dimension l_2 , since by construction any vectors in $\mathcal{S}(\mathbf{X}_1)$ annihilated by \mathbf{P}_Z are in $\mathcal{S}(\mathbf{X}_3)$, and are thus orthogonal to \mathbf{X}_2 . We may therefore choose an orthogonal basis for this space, normalized as usual. Let the $n \times l_2$ matrix \mathbf{Z}_2 constitute this basis. For

each possible choice of such an orthogonal, normalized, basis, there exists an $l_2 \times l_2$ nonsingular matrix Δ such that $P_{Z_1} \mathbf{X}_2 = \mathbf{Z}_2 \Delta$. Note that \mathbf{Z}_2 is orthogonal to \mathbf{X}_0 , since

$$\mathbf{X}_0^\top \mathbf{Z}_2 = \mathbf{X}_0^\top P_{Z_1} \mathbf{X}_2 \Delta^{-1} = \mathbf{0};$$

recall that \mathbf{X}_0 and \mathbf{Z}_1 are orthogonal by construction. \mathbf{Z}_2 is also orthogonal to \mathbf{X}_3 , since by construction \mathbf{X}_3 is orthogonal to the whole of $\mathcal{S}(\mathbf{Z})$.

Finally, we can complete the space $\mathcal{S}(\mathbf{Z})$ by yet another suitably normalized orthogonal basis, given by the l_3 columns of the matrix \mathbf{Z}_3 , which is chosen to be orthogonal to both \mathbf{X}_0 and \mathbf{Z}_2 . We have then that $\mathcal{S}(\mathbf{Z}_1)$ is spanned by the columns of \mathbf{Z}_2 and \mathbf{Z}_3 , that $\mathcal{S}(\mathbf{Z})$ is spanned by the columns of the three mutually orthogonal matrices \mathbf{X}_0 , \mathbf{Z}_2 , and \mathbf{Z}_3 , and that $l = k_0 + l_2 + l_3$. In addition, \mathbf{Z}_3 is orthogonal to \mathbf{X}_3 , which is orthogonal to all of \mathbf{Z} , and also to \mathbf{X}_2 , since

$$\mathbf{Z}_3^\top \mathbf{X}_2 = \mathbf{Z}_3^\top P_Z \mathbf{X}_2 = \mathbf{Z}_3^\top \mathbf{Z}_2 \Delta = \mathbf{0};$$

recall that \mathbf{Z}_2 and \mathbf{Z}_3 are orthogonal by construction.

It will be our goal to express the J statistic as a function of a set of mutually independent random variables. Since \mathbf{Z}_2 and \mathbf{X}_2 are not orthogonal, it will be necessary to introduce the orthogonal complement of $\mathcal{S}(\mathbf{X}_2)$ in the space, of dimension $2l_2$, spanned by the columns of \mathbf{X}_2 and \mathbf{Z}_2 . This orthogonal complement is spanned by the l_2 columns of $M_X \mathbf{Z}_2$. Define a normalized orthogonal basis of this space by the columns of the $n \times l_2$ matrix \mathbf{W} . For any such choice of \mathbf{W} , there exists an $l_2 \times l_2$ nonsingular matrix \mathbf{A} such that $M_X \mathbf{Z}_2 = M_{X_2} \mathbf{Z}_2 = \mathbf{W} \mathbf{A}$. Since both \mathbf{X}_2 and \mathbf{Z}_2 are orthogonal to both \mathbf{X}_0 and \mathbf{X}_3 , it follows that \mathbf{W} is also orthogonal to these two matrices.

Suppose that \mathbf{y} is generated by a DGP satisfying the null hypothesis

$$\mathbf{y} = \mathbf{X}_0 \beta_0 + \mathbf{X}_2 \beta_2 + \mathbf{X}_3 \beta_3 + \mathbf{u}, \quad (6)$$

where we may assume without loss of generality that $\mathbf{u} \sim N(\mathbf{0}, \mathbf{I})$, since J is clearly scale-invariant. The J statistic (3) can be rewritten as

$$J = \frac{(n - k - 1)^{1/2} \mathbf{y}^\top P_Z M_X \mathbf{y}}{\left[(\mathbf{y}^\top M_X \mathbf{y}) (\mathbf{y}^\top P_Z M_X P_Z \mathbf{y}) - (\mathbf{y}^\top P_Z M_X \mathbf{y})^2 \right]^{1/2}}. \quad (7)$$

This statistic depends on three stochastic scalars:

$$\mathbf{y}^\top M_X \mathbf{y}, \quad \mathbf{y}^\top P_Z M_X \mathbf{y}, \quad \text{and} \quad \mathbf{y}^\top P_Z M_X P_Z \mathbf{y}. \quad (8)$$

It is easy to see that, under (6), $\mathbf{y}^\top M_X \mathbf{y} = \mathbf{u}^\top M_X \mathbf{u}$. It is not quite so easy to deal with the other two expressions in (8), both of which depend on the quantity

$$\begin{aligned} \mathbf{R} \equiv M_X P_Z \mathbf{y} &= M_X (\mathbf{X}_0 \beta_0 + P_Z \mathbf{X}_2 \beta_2) + M_X P_Z \mathbf{u} \\ &= M_X \mathbf{Z}_2 \Delta \beta_2 + M_X P_Z \mathbf{u} \\ &= \mathbf{W} \mathbf{A} \Delta \beta_2 + M_X P_Z \mathbf{u}. \end{aligned} \quad (9)$$

We can see already that the distribution of J depends neither on β_0 nor on β_3 . It is only through \mathbf{R} that J depends on β , and from (9) it is clear that β_2 is the only part of β on which \mathbf{R} depends.

The second term of (9) is

$$\begin{aligned} \mathbf{M}_X \mathbf{P}_Z \mathbf{u} &= (\mathbf{I} - \mathbf{P}_{X_0} - \mathbf{P}_{X_2} - \mathbf{P}_{X_3})(\mathbf{P}_{X_0} + \mathbf{P}_{Z_2} + \mathbf{P}_{Z_3})\mathbf{u} \\ &= (\mathbf{P}_{X_0} + \mathbf{P}_{Z_2} + \mathbf{P}_{Z_3} - \mathbf{P}_{X_0} - \mathbf{P}_{X_2} \mathbf{P}_{Z_2})\mathbf{u} \\ &= \mathbf{M}_{X_2} \mathbf{P}_{Z_2} \mathbf{u} + \mathbf{P}_{Z_3} \mathbf{u}. \end{aligned} \quad (10)$$

Expanding \mathbf{P}_{Z_2} and recalling the normalization of \mathbf{Z}_2 , we get

$$\mathbf{M}_{X_2} \mathbf{P}_{Z_2} = N^{-1} \mathbf{M}_{X_2} \mathbf{Z}_2 \mathbf{Z}_2^\top = N^{-1} \mathbf{W} \mathbf{A} \mathbf{Z}_2^\top. \quad (11)$$

Since we wish to express everything in terms of \mathbf{X}_2 and \mathbf{W} rather than of \mathbf{X}_2 and \mathbf{Z}_2 , we replace \mathbf{Z}_2 in the last expression above by

$$\mathbf{Z}_2 = \mathbf{P}_{X_2} \mathbf{Z}_2 + \mathbf{M}_{X_2} \mathbf{Z}_2 = \mathbf{P}_{X_2} \mathbf{Z}_2 + \mathbf{W} \mathbf{A}. \quad (12)$$

Now recall that

$$\mathbf{P}_{Z_1} \mathbf{X}_2 = \mathbf{P}_{Z_2} \mathbf{X}_2 = \mathbf{Z}_2 \mathbf{\Delta}. \quad (13)$$

The middle expression above is clearly just $N^{-1} \mathbf{Z}_2 \mathbf{Z}_2^\top \mathbf{X}_2$. Thus

$$\mathbf{\Delta} = N^{-1} \mathbf{Z}_2^\top \mathbf{X}_2. \quad (14)$$

Consequently,

$$\mathbf{P}_{X_2} \mathbf{Z}_2 = \mathbf{X}_2 \mathbf{\Delta}^\top, \quad (15)$$

and (12) becomes

$$\mathbf{Z}_2 = \mathbf{W} \mathbf{A} + \mathbf{X}_2 \mathbf{\Delta}^\top.$$

Substituting this into (11) yields

$$\mathbf{M}_{X_2} \mathbf{P}_{Z_2} = N^{-1} \mathbf{W} \mathbf{A} (\mathbf{A}^\top \mathbf{W}^\top + \mathbf{\Delta} \mathbf{X}_2^\top). \quad (16)$$

It now is convenient to define a new parameter vector $\boldsymbol{\theta}$ by the formula

$$\boldsymbol{\theta} = \mathbf{A} \mathbf{\Delta} \beta_2. \quad (17)$$

From (9), (10), (16), and (17), we have

$$\mathbf{R} = \mathbf{W} \boldsymbol{\theta} + N^{-1} \mathbf{W} \mathbf{A} (\mathbf{A}^\top \mathbf{W}^\top \mathbf{u} + \mathbf{\Delta} \mathbf{X}_2^\top \mathbf{u}) + N^{-1} \mathbf{Z}_3 \mathbf{Z}_3^\top \mathbf{u}. \quad (18)$$

This expression suggests the following definitions:

$$\begin{aligned} \mathbf{x} &\equiv N^{-1/2} \mathbf{X}_2^\top \mathbf{u} \sim N(\mathbf{0}, \mathbf{I}_{l_2}); \\ \mathbf{w} &\equiv N^{-1/2} \mathbf{W}^\top \mathbf{u} \sim N(\mathbf{0}, \mathbf{I}_{l_2}); \\ \mathbf{z} &\equiv N^{-1/2} \mathbf{Z}_3^\top \mathbf{u} \sim N(\mathbf{0}, \mathbf{I}_{l_3}). \end{aligned} \quad (19)$$

These three vectors of standard normal random variables are clearly mutually independent, on account of the mutual orthogonality of \mathbf{X}_2 , \mathbf{W} , and \mathbf{Z}_3 . In terms of these variables, then, (18) becomes

$$\mathbf{R} = \mathbf{W}\boldsymbol{\theta} + N^{-1/2}(\mathbf{W}\mathbf{A}\mathbf{A}^\top\mathbf{w} + \mathbf{W}\mathbf{A}\boldsymbol{\Delta}\mathbf{x} + \mathbf{Z}_3\mathbf{z}). \quad (20)$$

We may now proceed to evaluate the various pieces of (7). First, by use of (20),

$$\begin{aligned} \mathbf{y}^\top \mathbf{P}_Z \mathbf{M}_X \mathbf{y} &= \mathbf{R}^\top \mathbf{u} \\ &= N^{1/2} \boldsymbol{\theta}^\top \mathbf{w} + \mathbf{w}^\top \mathbf{A} \mathbf{A}^\top \mathbf{w} + \mathbf{w}^\top \mathbf{A} \boldsymbol{\Delta} \mathbf{x} + \mathbf{z}^\top \mathbf{z}. \end{aligned} \quad (21)$$

Similarly,

$$\begin{aligned} \mathbf{y}^\top \mathbf{P}_Z \mathbf{M}_X \mathbf{P}_Z \mathbf{y} &= \mathbf{R}^\top \mathbf{R} \\ &= N \|\boldsymbol{\theta}\|^2 + 2N^{1/2} (\boldsymbol{\theta}^\top \mathbf{A} \mathbf{A}^\top \mathbf{w} + \boldsymbol{\theta}^\top \mathbf{A} \boldsymbol{\Delta} \mathbf{x}) \\ &\quad + \mathbf{w}^\top \mathbf{A} \mathbf{A}^\top \mathbf{A} \mathbf{A}^\top \mathbf{w} + 2\mathbf{w}^\top \mathbf{A} \mathbf{A}^\top \mathbf{A} \boldsymbol{\Delta} \mathbf{x} + \mathbf{x}^\top \boldsymbol{\Delta}^\top \mathbf{A}^\top \mathbf{A} \boldsymbol{\Delta} \mathbf{x} + \mathbf{z}^\top \mathbf{z}. \end{aligned} \quad (22)$$

In order to simplify these rather forbidding expressions, we can make special choices of the matrices \mathbf{X}_2 , \mathbf{Z}_2 , and \mathbf{W} , which are at this point quite arbitrary normalized orthogonal bases of the spaces spanned by their columns. To this end, consider the $l_2 \times l_2$ matrix $\boldsymbol{\Delta}^\top \boldsymbol{\Delta}$, which by (14) is

$$N^{-1} \mathbf{X}_2^\top \mathbf{P}_{Z_2} \mathbf{X}_2.$$

This is a symmetric, positive definite matrix, which can be diagonalized by means of an orthogonal matrix, \mathbf{U} , say:

$$\boldsymbol{\Delta}^\top \boldsymbol{\Delta} = \mathbf{U} \mathbf{D}^2 \mathbf{U}^\top,$$

where $\mathbf{D} \equiv \text{diag}(d_1, \dots, d_{l_2})$ is a diagonal matrix, the elements of which are the canonical correlations defined between \mathbf{X}_2 and \mathbf{Z}_2 .

Note further that, from (12) and (15),

$$\mathbf{W}\mathbf{A} = \mathbf{Z}_2 - \mathbf{X}_2 \boldsymbol{\Delta}^\top. \quad (23)$$

Thus

$$\begin{aligned} \mathbf{A}^\top \mathbf{W}^\top \mathbf{W} \mathbf{A} &= \mathbf{N} \mathbf{A}^\top \mathbf{A} \\ &= \mathbf{Z}_2^\top \mathbf{Z}_2 + \boldsymbol{\Delta} \mathbf{X}_2^\top \mathbf{X}_2 \boldsymbol{\Delta}^\top - \mathbf{Z}_2^\top \mathbf{X}_2 \boldsymbol{\Delta}^\top - \boldsymbol{\Delta} \mathbf{X}_2^\top \mathbf{Z}_2 \\ &= \mathbf{N} \mathbf{I} + \mathbf{N} \boldsymbol{\Delta} \boldsymbol{\Delta}^\top - \mathbf{N} \boldsymbol{\Delta} \boldsymbol{\Delta}^\top - \mathbf{N} \boldsymbol{\Delta} \boldsymbol{\Delta}^\top \\ &= \mathbf{N} (\mathbf{I} - \boldsymbol{\Delta} \boldsymbol{\Delta}^\top). \end{aligned}$$

Consequently,

$$\boldsymbol{\Delta}^\top \mathbf{A}^\top \mathbf{A} \boldsymbol{\Delta} = \boldsymbol{\Delta}^\top \boldsymbol{\Delta} - (\boldsymbol{\Delta}^\top \boldsymbol{\Delta})^2 = \mathbf{U} (\mathbf{D}^2 - \mathbf{D}^4) \mathbf{U}^\top. \quad (24)$$

Consider the term $\mathbf{x}^\top \boldsymbol{\Delta}^\top \mathbf{A}^\top \mathbf{A} \boldsymbol{\Delta} \mathbf{x}$ in (22). Clearly it is

$$\mathbf{x}^\top \mathbf{U} (\mathbf{D}^2 - \mathbf{D}^4) \mathbf{U}^\top \mathbf{x}. \quad (25)$$

We now make our specific choice of the normalized orthogonal basis of the space spanned by \mathbf{X}_2 . The new \mathbf{X}_2 will be the old one postmultiplied by \mathbf{U} . Because \mathbf{U} is orthogonal, so that $\mathbf{U}^\top \mathbf{U} = \mathbf{U} \mathbf{U}^\top = \mathbf{I}$, the columns of the new \mathbf{X}_2 still constitute a normalized orthogonal basis of the same space as that spanned by the columns of \mathbf{X}_2 . When \mathbf{x} is defined in terms of the new \mathbf{X}_2 , expression (25) becomes just $\mathbf{x}^\top (\mathbf{D}^2 - \mathbf{D}^4) \mathbf{x}$, and $\boldsymbol{\Delta}^\top \boldsymbol{\Delta} = \mathbf{D}^2$.

This last relation now suggests a specific choice for \mathbf{Z}_2 . It implies that $\boldsymbol{\Delta} = \mathbf{V} \mathbf{D}$, for some orthogonal matrix \mathbf{V} . By (14), $\boldsymbol{\Delta} = N^{-1} \mathbf{Z}_2^\top \mathbf{X}_2$, and so choosing for our new \mathbf{Z}_2 the old one postmultiplied by \mathbf{V} causes the old $\boldsymbol{\Delta}$ to be premultiplied by \mathbf{V}^\top , so that the new $\boldsymbol{\Delta}$ is just \mathbf{D} .

Having fixed these choices of \mathbf{X}_2 and \mathbf{Z}_2 , we can proceed to make a specific choice for \mathbf{W} . Whatever the choice of \mathbf{W} may be, we have, from (24) that, with the new \mathbf{X}_2 ,

$$\boldsymbol{\Delta}^\top \mathbf{A}^\top \mathbf{A} \boldsymbol{\Delta} = \mathbf{D}^2 - \mathbf{D}^4.$$

But, with the new \mathbf{Z}_2 , $\boldsymbol{\Delta} = \mathbf{D}$, and so

$$\mathbf{A}^\top \mathbf{A} = \mathbf{I} - \mathbf{D}^2.$$

If we define $\mathbf{B} = \text{diag}((1 - d_i^2)^{1/2})$, it follows that $\mathbf{A} = \mathbf{Q} \mathbf{B}$ for yet another orthogonal matrix \mathbf{Q} . Now, for a given choice of \mathbf{W} , \mathbf{A} is defined by (23). What that equation actually defines is not \mathbf{A} , but $\mathbf{W} \mathbf{A}$. If then we define a new \mathbf{W} equal to the old one postmultiplied by \mathbf{Q} , then the new \mathbf{A} is just the diagonal matrix \mathbf{B} .

We have thus succeeded in finding appropriate forms for \mathbf{X}_2 , \mathbf{Z}_2 , and \mathbf{W} , such that $\boldsymbol{\Delta}$ and \mathbf{A} are simultaneously diagonal, with typical diagonal elements d_i and $(1 - d_i^2)^{1/2}$. This allows us to rewrite (21) and (22) in simpler forms. First, to ease the notational burden, define the l_2 -vectors \mathbf{v}_1 , \mathbf{v}_2 , and \mathbf{v} with typical elements

$$v_{1i} = (1 - d_i^2) w_i, \quad v_{2i} = d_i (1 - d_i^2)^{1/2} x_i, \quad v_i = v_{1i} + v_{2i}. \quad (26)$$

Then we obtain:

$$\mathbf{y}^\top \mathbf{P}_Z \mathbf{M}_X \mathbf{y} = N^{1/2} \boldsymbol{\theta}^\top \mathbf{w} + \mathbf{w}^\top \mathbf{v} + \mathbf{z}^\top \mathbf{z},$$

and

$$\mathbf{y}^\top \mathbf{P}_Z \mathbf{M}_X \mathbf{P}_Z \mathbf{y} = N \|\boldsymbol{\theta}\|^2 + 2N^{1/2} \boldsymbol{\theta}^\top \mathbf{v} + \mathbf{v}^\top \mathbf{v} + \mathbf{z}^\top \mathbf{z}.$$

Just before we use these expressions in (7), note that

$$\mathbf{y}^\top \mathbf{M}_X \mathbf{y} = \mathbf{u}^\top \mathbf{M}_X \mathbf{u} = \mathbf{u}^\top (\mathbf{M}_X - \mathbf{P}_{Z_3} - \mathbf{P}_W) \mathbf{u} + \mathbf{w}^\top \mathbf{w} + \mathbf{z}^\top \mathbf{z}, \quad (27)$$

where $S^2 \equiv \mathbf{u}^\top (\mathbf{M}_X - \mathbf{P}_{Z_3} - \mathbf{P}_W) \mathbf{u}$ is independent of \mathbf{x} , \mathbf{w} , and \mathbf{z} , and is distributed as $\chi^2(n - k_0 - 2l_2 - k_3 - l_3)$.

At this point, we can finally make a sensible choice for the normalization factor N : We define $N = n - k_0 - k_3$. Then S^2 has $N - 2l_2 - l_3$ degrees of freedom. Provided that $n > k_0 + 2l_2 + l_3 + k_3$, so that S^2 has a positive number of degrees of freedom, (7) becomes:

$$J = \frac{(N - l_2 - 1)^{1/2} (N^{1/2} \boldsymbol{\theta}^\top \mathbf{w} + \mathbf{w}^\top \mathbf{v} + \mathbf{z}^\top \mathbf{z})}{\left[(S^2 + \mathbf{z}^\top \mathbf{z} + \mathbf{w}^\top \mathbf{w}) (N \|\boldsymbol{\theta}\|^2 + 2N^{1/2} \boldsymbol{\theta}^\top \mathbf{v} + \mathbf{v}^\top \mathbf{v} + \mathbf{z}^\top \mathbf{z}) - (N^{1/2} \boldsymbol{\theta}^\top \mathbf{w} + \mathbf{w}^\top \mathbf{v} + \mathbf{z}^\top \mathbf{z})^2 \right]^{1/2}} \quad (28)$$

This remarkable formula has a great many interesting properties. It expresses J in terms of deterministic quantities and the independent random variables \mathbf{x} , \mathbf{w} , \mathbf{u} , and S^2 , all of which have *known* distributions. It is therefore possible to simulate J directly from (28), by drawing these random variables from their respective distributions. This will turn out to be extremely useful when we undertake the Monte Carlo experiments of Sections 5 and 7. Note that \mathbf{v} is *not* independent of \mathbf{w} ; see (26). It is used in place of \mathbf{x} , which is independent of \mathbf{w} , only because doing so makes (28) much easier to write.

Notice that (28) depends on n only through N . This, along with the obvious fact that (28) depends neither on \mathbf{X}_0 nor on \mathbf{X}_3 , justifies the earlier remarks that the regressors in \mathbf{X}_0 and \mathbf{X}_3 have no effect other than to reduce the number of degrees of freedom. The actual design of the two regressor matrices \mathbf{X} and \mathbf{Z} influence (28) only through the canonical correlation coefficients d_i , $i = 1, \dots, l_2$. Degrees of freedom do matter, of course, and there are precisely three of them that influence (28) independently: N , l_2 , and l_3 .

The other thing that influences (28) is the parameter vector $\boldsymbol{\theta}$. It can be seen that J depends on $\boldsymbol{\theta}$ through $\|\boldsymbol{\theta}\|^2$, $\boldsymbol{\theta}^\top \mathbf{w}$, and $\boldsymbol{\theta}^\top \mathbf{v}$. It appears from (17) that $\boldsymbol{\theta}$ depends only on $\boldsymbol{\beta}_2$, in keeping with the observation that J is unaffected by \mathbf{X}_0 and \mathbf{X}_3 . Although (17) takes on a very simple form in our special bases in which \mathbf{A} and $\boldsymbol{\Delta}$ are diagonal, it is useful to see what some of the expressions in (28) are in terms of the raw regressor matrices \mathbf{X} and \mathbf{Z} . From (17), we have

$$\mathbf{W}\boldsymbol{\theta} = \mathbf{W}\mathbf{A}\boldsymbol{\Delta}\boldsymbol{\beta}_2 = \mathbf{M}_{X_2}\mathbf{Z}_2\boldsymbol{\Delta}\boldsymbol{\beta}_2 = \mathbf{M}_{X_2}\mathbf{P}_{Z_2}\mathbf{X}_2\boldsymbol{\beta}_2.$$

The last expression here is valid however the null hypothesis is parametrized. Since \mathbf{W} is normalized, we can directly compute $\|\boldsymbol{\theta}\|^2$:

$$\|\boldsymbol{\theta}\|^2 = N^{-1} \boldsymbol{\theta}^\top \mathbf{W}^\top \mathbf{W} \boldsymbol{\theta} = N^{-1} \boldsymbol{\beta}_2^\top \mathbf{X}_2^\top \mathbf{P}_{Z_2} \mathbf{M}_{X_2} \mathbf{P}_{Z_2} \mathbf{X}_2 \boldsymbol{\beta}_2.$$

It is easy to see that this last expression can be written as

$$V \equiv N^{-1} \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{P}_Z \mathbf{M}_X \mathbf{P}_Z \mathbf{X} \boldsymbol{\beta}, \quad (29)$$

which is $1/N$ times the leading-order term in (5). Similarly, we have

$$\boldsymbol{\theta}^\top \mathbf{w} = N^{-1/2} \boldsymbol{\theta}^\top \mathbf{W}^\top \mathbf{u} = N^{-1/2} \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{P}_Z \mathbf{M}_X \mathbf{u},$$

and the last expression here is $N^{-1/2}$ times the leading-order term in (4). If we divide numerator and denominator in (28) by N , and then let $N \rightarrow \infty$, noting that $\text{plim}_{N \rightarrow \infty} S^2/N = 1$, we find that the limit of (28) is

$$\text{plim}_{N \rightarrow \infty} J = \frac{\boldsymbol{\theta}^\top \mathbf{w}}{\|\boldsymbol{\theta}\|}. \quad (30)$$

It is clear that this has the standard normal distribution.

It is illuminating to see what (28) reduces to in several special cases. If all the canonical correlations d_i tend to zero, we have the case of orthogonal models, in which the J statistic does not have its usual asymptotic distribution. In that case, $\mathbf{A} = \mathbf{I}$ and $\boldsymbol{\Delta} = \mathbf{0}$, which further implies that $\boldsymbol{\theta} = \mathbf{0}$. In addition, it follows from (26) that $\mathbf{v} = \mathbf{w}$. With this, (28) gives

$$\begin{aligned} J &= \frac{(N - l_2 - 1)^{1/2}(\mathbf{w}^\top \mathbf{w} + \mathbf{z}^\top \mathbf{z})}{\left[(S^2 + \mathbf{w}^\top \mathbf{w} + \mathbf{z}^\top \mathbf{z})(\mathbf{w}^\top \mathbf{w} + \mathbf{z}^\top \mathbf{z}) - (\mathbf{w}^\top \mathbf{w} + \mathbf{z}^\top \mathbf{z})^2 \right]^{1/2}} \\ &= \left(\frac{\mathbf{w}^\top \mathbf{w} + \mathbf{z}^\top \mathbf{z}}{S^2/(N - l_2 - 1)} \right)^{1/2}, \end{aligned}$$

the square of which is proportional to the F statistic. This special case, and others related to it, have been studied by Michelis (1995, 1996).

If we go to the other extreme, and let all the d_i tend to unity, then in the limit the null is nested in the alternative, since we have $\mathbf{X}_2 = \mathbf{Z}_2$. In this case, $\mathbf{A} = \mathbf{0}$, and $\boldsymbol{\Delta} = \mathbf{I}$. Once more $\boldsymbol{\theta} = \mathbf{0}$, and this time $\mathbf{v} = \mathbf{0}$. (28) becomes

$$\begin{aligned} J &= \frac{(N - l_2 - 1)^{1/2} \mathbf{z}^\top \mathbf{z}}{\left[(S^2 + \mathbf{w}^\top \mathbf{w} + \mathbf{z}^\top \mathbf{z}) \mathbf{z}^\top \mathbf{z} - (\mathbf{z}^\top \mathbf{z})^2 \right]^{1/2}} \\ &= \left(\frac{\mathbf{z}^\top \mathbf{z}}{(S^2 + \mathbf{w}^\top \mathbf{w})/(N - l_2 - 1)} \right)^{1/2}, \end{aligned}$$

the square of which is proportional to the F statistic against a nested alternative with the \mathbf{Z}_3 as extra regressors.

Next consider the special case with $l_2 = 1$, $l_3 = 0$, that is, the case in which each of the models contains just one regressor that is not in the other model. In this case, \mathbf{w} , \mathbf{v} and $\boldsymbol{\theta}$ all become scalars, \mathbf{z} vanishes, and (28) becomes

$$J = \frac{(N - 2)^{1/2} (N^{1/2} \theta w + wv)}{\left[(S^2 + w^2)(N\theta^2 + 2N^{1/2}\theta v + v^2) - (N^{1/2}\theta w + wv)^2 \right]^{1/2}}.$$

The denominator here is the square root of

$$(S + w^2)(N^{1/2}\theta + v)^2 - w^2(N^{1/2}\theta + v)^2 = S^2(N^{1/2}\theta + v)^2,$$

so that

$$J = \frac{(N-2)^{1/2}w(N^{1/2}\theta + V)}{S(N^{1/2}\theta + v)} = \frac{(N-2)^{1/2}w}{S},$$

which is just the usual t statistic. This is the one case in which the J statistic actually has the Student's t distribution in finite samples.

4. A Modified J Test

A great deal of Monte Carlo work, including Godfrey and Pesaran (1983) and Godfrey (1997), has found that the main problem with the J test is that it has a nonzero mean. It therefore seems natural to correct the test by subtracting an estimate of this mean. In order to obtain a corrected test statistic that has a mean of 0 through order $N^{-1/2}$, we need to perform an asymptotic expansion. This will be done under the assumption that the d_i are bounded away from both 0 and 1.

First, define an order unity variable η , independent of the normal random vectors, by the relation

$$S^2/N = 1 + N^{-1/2}\eta.$$

Clearly, $E(\eta) = O(N^{-1})$. Dividing the numerator and denominator of (28) by N , we can rewrite J as

$$\begin{aligned} & (\boldsymbol{\theta}^\top \mathbf{w} + N^{-1/2}(\mathbf{w}^\top \mathbf{v} + \mathbf{z}^\top \mathbf{z})) \left((1 + N^{-1/2}\eta) (\|\boldsymbol{\theta}\|^2 + 2N^{-1/2}\boldsymbol{\theta}^\top \mathbf{v}) \right)^{-1/2} + O(N^{-1}) \\ &= \frac{1}{\|\boldsymbol{\theta}\|} (\boldsymbol{\theta}^\top \mathbf{w} + N^{-1/2}(\mathbf{w}^\top \mathbf{v} + \mathbf{z}^\top \mathbf{z})) (1 - N^{-1/2}\eta/2) \left(1 - N^{-1/2} \frac{\boldsymbol{\theta}^\top \mathbf{v}}{\|\boldsymbol{\theta}\|^2} \right) + O(N^{-1}) \\ &= \frac{\boldsymbol{\theta}^\top \mathbf{w}}{\|\boldsymbol{\theta}\|} + \frac{N^{-1/2}}{\|\boldsymbol{\theta}\|} \left(\mathbf{w}^\top \mathbf{v} + \mathbf{z}^\top \mathbf{z} - \frac{1}{2}\eta \boldsymbol{\theta}^\top \mathbf{w} - \frac{1}{\|\boldsymbol{\theta}\|^2} \boldsymbol{\theta}^\top \mathbf{w} \mathbf{v}^\top \boldsymbol{\theta} \right) + O(N^{-1}). \end{aligned} \quad (31)$$

The leading-order term here is (30), and we have already seen that it has the standard normal distribution. Thus, what we are interested in is the second term in (31).

There are four terms inside the parentheses in the second term in (31). The second has expectation l_3 and the third has expectation 0. It can be shown that

$$E(\mathbf{w}^\top \mathbf{v}) = \text{Tr}(\mathbf{M}_{X_2} \mathbf{P}_{Z_2}) = \text{Tr}(\mathbf{M}_X \mathbf{P}_Z) - l_3 \quad (32)$$

and that

$$E(\boldsymbol{\theta}^\top \mathbf{w} \mathbf{v}^\top \boldsymbol{\theta}) = W \equiv N^{-1} \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{P}_Z \mathbf{M}_X \mathbf{P}_Z \mathbf{M}_X \mathbf{P}_Z \mathbf{X} \boldsymbol{\beta}. \quad (33)$$

Now recall (29), which defined $V = \|\boldsymbol{\theta}\|^2$. This, along with (32) and (33), shows that, to order $N^{-1/2}$, the expectation of J is

$$N^{-1/2} \frac{1}{V^{1/2}} \left(\text{Tr}(\mathbf{M}_X \mathbf{P}_Z) - \frac{W}{V} \right). \quad (34)$$

The obvious way to modify the J statistic is to subtract an estimate of (34) from it. However, this approach is probably not a good one to use. The reason is that, when $\|\boldsymbol{\theta}\| = 0$, $V = W = 0$, and so (34) tends to infinity for finite N . Even if $\|\hat{\boldsymbol{\theta}}\| = 0$ with probability zero, we may still expect that (34) will tend to dominate the leading-order term in (31) if $\boldsymbol{\theta}$ is close to zero. The obvious way to avoid this problem is to subtract something from the numerator of (28) only, in such a way as to make the expectation of the result zero through order $N^{-1/2}$. If such a term is incorporated into the stochastic expansion (31), it is easy to see that the quantity to be subtracted is the expectation of

$$\mathbf{w}^\top \mathbf{v} + \mathbf{z}^\top \mathbf{z} - \frac{1}{2} \eta \boldsymbol{\theta}^\top \mathbf{w} - \frac{1}{\|\boldsymbol{\theta}\|^2} \boldsymbol{\theta}^\top \mathbf{w} \mathbf{v}^\top \boldsymbol{\theta}, \quad (35)$$

which, as we have seen, equals

$$\text{Tr}(\mathbf{M}_X \mathbf{P}_Z) - \frac{W}{V}. \quad (36)$$

The estimate of σ^2 used to compute the test statistic could be either the estimate from the J test regression or the estimate from OLS estimation of H_1 . Since the former is biased downwards because of the correlation between the error terms and the test regressor, it seems more attractive to use the latter.

Because of the various assumptions that were made to make the analysis easier, the preceding discussion does not provide a clear prescription for computing the modified J test. The test statistic is actually defined as follows:

$$J_M = \frac{\mathbf{y}^\top \mathbf{M}_X \mathbf{P}_Z \mathbf{y} - s^2 (\text{Tr}(\mathbf{M}_X \mathbf{P}_Z) + \hat{W}/\hat{V})}{s (\mathbf{y}^\top \mathbf{P}_Z \mathbf{M}_X \mathbf{P}_Z \mathbf{y})^{1/2}}, \quad (37)$$

where s is the estimate of the regression standard error from OLS estimation of H_1 ,

$$\hat{V} \equiv \mathbf{y}^\top \mathbf{P}_X \mathbf{P}_Z \mathbf{M}_X \mathbf{P}_Z \mathbf{P}_X \mathbf{y},$$

and

$$\hat{W} \equiv \mathbf{y}^\top \mathbf{P}_X \mathbf{P}_Z \mathbf{M}_X \mathbf{P}_Z \mathbf{M}_X \mathbf{P}_Z \mathbf{P}_X \mathbf{y}.$$

These estimates of V and W are simply (29) and (33) with $\mathbf{X}\boldsymbol{\beta}$ replaced by the fitted values $\mathbf{P}_X \mathbf{y}$ and the factors of N^{-1} , which cancel in (37), omitted. It is necessary to multiply the correction term in the numerator by s^2 because we are no longer making any normalizing assumption about the value of σ^2 . What is new about (37) is the second part of the correction term. Without this second part, we would simply be subtracting an estimate of the mean of the J test numerator from the latter, and then dividing by an estimate of its variance. Such a procedure would be even simpler, but it would not yield a test statistic with mean zero through $O(N^{-1/2})$.

Although it seems likely that J_M will perform much better than J as an asymptotic test, we do not claim that it will perform acceptably. However, as we shall demonstrate in Section 7, J_M does perform extremely well, and significantly better than J , when both are bootstrapped.

5. Finite-Sample Performance of the Tests

In this section, we use simulation methods to study the finite-sample performance of the J and J_M tests. Neither test performs particularly well when it is not bootstrapped, but the performance of J_M is much less sensitive to parameter values. As we shall see in Section 7, this has important consequences for the performance of the two tests when they are bootstrapped.

The theoretical results of Section 4 make it possible to perform Monte Carlo experiments that employ extremely large numbers of replications. This is essential if we are to obtain results accurate enough to distinguish among competing bootstrap tests. The key result is (28). Instead of generating regressors and error terms, we can simply generate realizations of \mathbf{x} , \mathbf{w} , \mathbf{z} , and S^2 and plug them into expression (28) to calculate a realization of the J statistic. Moreover, since S^2 is the only random variable in (28) that depends on N , we can inexpensively generate realizations of the J statistic for a large number of different sample sizes at the same time. For the J_M statistic, an expression similar to (28) is available and can be used in the same way.

This procedure does have one limitation. As can be seen from (19), the results that the vectors \mathbf{x} , \mathbf{w} , and \mathbf{z} are normally distributed depend crucially on the assumption that the error vector \mathbf{u} is normally distributed. If that assumption were dropped, (28) would still be true, but we could no longer generate \mathbf{x} , \mathbf{w} , and \mathbf{z} as i.i.d. normal vectors. For this reason, we maintain the normality assumption in all of our Monte Carlo experiments.

As we have seen, the distributions of J and J_M depend on N , l_2 , l_3 , the canonical correlations d_i , and the parameter vector $\boldsymbol{\theta}$. If we allow σ^2 to differ from unity, $\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\Delta}\boldsymbol{\beta}_2/\sigma$. Thus all elements of $\boldsymbol{\theta}$ will tend to be small when the error variance is large, and most elements will tend to be large when the error variance is small.

The most interesting feature of the finite-sample performance of the J test is the way in which it depends on $\boldsymbol{\theta}$. This is illustrated in Figures 1, 2, and 3. All of these figures are based on 250,000 replications for each of a large number of values of $\boldsymbol{\theta}$. They show rejection frequencies for a one-tailed test at a nominal level of 5% based on the Student's t distribution with $N - l_2 - 1$ degrees of freedom. Figures 1 and 2 pertain to the case $l_2 = 2$, $l_3 = 4$, $d_1 = 0.9$, and $d_2 = 0.5$. Changing any of these parameters would change the details of the figures, but not their basic shapes, provided of course that the case $l_2 = 1, l_3 = 0$ is avoided. In Figure 1, both θ_1 and θ_2 vary together, which is equivalent to varying σ^2 while holding $\boldsymbol{\beta}$ fixed. In Figure 2, θ_1 is fixed at 1 and θ_2 varies. Because there are 6 regressors in H_2 that are not in H_1 , we might expect the J test to perform rather poorly in this case, and that is indeed what happens, at least for smaller values of the θ_i .

Figure 1 is striking. For large values of θ_1 and θ_2 , the J test performs reasonably well. However, for values near 0, it overrejects very severely. This is true even for moderately large sample sizes. As N increases, the range of values for which the overrejection is severe diminishes, but the extent of the overrejection near 0 does not diminish very much. This is precisely what we would expect from the asymptotic expansion (31), since it is not defined when the denominator $\|\boldsymbol{\theta}\|$ is equal to zero.

It seems safe to conclude that there is a singularity at $\boldsymbol{\theta} = \mathbf{0}$. On the other hand, in Figure 2, where only θ_2 varies, no singularity is evident, a result that is again consistent with (31). Note the different vertical scale of the two figures. While the finite-sample behavior of the J test in Figure 2 is certainly not very good, and clearly deteriorates as θ_2 gets closer to 0, it does improve markedly as the sample size increases.

Figure 3 demonstrates that the performance of the J test does indeed deteriorate markedly as the number of regressors in H_2 that are not in H_1 increases. As l_3 varies from 0 to 4, this number, which is equal to $l_2 + l_3$, varies from 2 to 6. The height and the steepness of the hump near $\boldsymbol{\theta} = \mathbf{0}$ both increase markedly as l_3 increases.

Figures 4, 5, and 6 are based on the same experiments as Figures 1, 2, and 3, respectively, but they contain results for the J_M test instead of the J test. Note the very different scale of the vertical axis in these figures. We see that the J_M test tends to underreject in small samples, but that this tendency goes away quite rapidly as the sample size increases. The rejection frequencies in all the figures are based on the $t(N - l_2 - 1)$ distribution, and the J_M test would actually work better if the $N(0, 1)$ distribution were used instead. From Figure 4, we see that the J_M test apparently does have a singularity at $\boldsymbol{\theta} = \mathbf{0}$, but a much less severe one than the J test. As expected, no singularity is to be seen in Figure 5, and there is only a very weak dependence on the value of θ_2 . Finally, Figure 6 demonstrates that the performance of the J_M test, unlike that of the J test, is not very sensitive to the number of regressors in H_2 that are not in H_1 .

The fact that the J_M test generally performs very much better than the J test is interesting. However, the principal objective of this paper is not to propose yet another nonnested test, but to explain why the J test performs so well when it is bootstrapped. In the remainder of the paper, we therefore turn our attention to bootstrap tests.

6. Bootstrapping the J and J_M Tests

In this section, we discuss how to bootstrap the J and J_M tests. For a bootstrap test, the actual test statistic, which will be denoted $\hat{\tau}$, is computed in the usual way. Instead of comparing this with critical values from some theoretical distribution, we decide whether or not to reject the null hypothesis by comparing $\hat{\tau}$ with a distribution estimated from a large number of bootstrap test statistics computed using simulated data sets. Because we are assuming normality, we will limit our attention to the parametric bootstrap.

There are many ways to bootstrap a hypothesis test, and the choice among them seems to be largely a matter of taste. In our view, the easiest and most informative way to use the bootstrap in most hypothesis testing situations is to compute bootstrap P values; for another method, see Horowitz (1994). The procedure we advocate works as follows:

1. Estimate H_1 by OLS. Then obtain $\hat{\tau}$, the test statistic of interest, which may be either J or J_M , and the parameter estimates $\hat{\beta}$ and s^2 .
2. Draw B sets of bootstrap error terms \mathbf{u}_j^* from the $N(0, s^2)$ distribution and use them, along with the parameter estimates $\hat{\beta}$, to generate B bootstrap samples \mathbf{y}_j^* , $j = 1, \dots, B$, according to the equation

$$\mathbf{y}_j^* = \mathbf{X}\hat{\beta} + \mathbf{u}_j^*.$$

3. Using each of these bootstrap samples, calculate a test statistic τ_j^* in exactly the same way as $\hat{\tau}$ was computed from the actual sample. Then compute the estimated bootstrap P value

$$\hat{p}^*(\hat{\tau}) = \frac{1}{B} \sum_{j=1}^B I(\tau_j^* \geq \hat{\tau}), \quad (38)$$

where $I(\cdot)$ is an indicator function, equal to 1 if its argument is true and equal to 0 otherwise. This assumes that the J test is being treated as a one-tailed test.

In the procedure just described, B may be any positive integer which is chosen so that $\alpha(B + 1)$ is an integer, where α is the desired test size. Larger values of B will make the test results less dependent on the realizations of the bootstrap samples and will increase test power somewhat; see Davidson and MacKinnon (1997b). Unless computing costs are a problem, we suggest using at least $B = 999$, and preferably a larger value, in applications. For purposes of Monte Carlo experiments, however, smaller values of B work perfectly well, because errors tend to cancel out across the replications. In the next section, we will use $B = 399$ in our experiments.

Unless a test statistic is pivotal, which means that its distribution does not depend on unknown parameters, a bootstrap test based upon it will not be exact. Beran (1988) is the classic reference; see also Hall (1992). The problem is that the bootstrap P value (38) is based on the distribution of the bootstrap statistics τ_j^* , and this will differ from the true distribution of $\hat{\tau}$ whenever the statistic is not pivotal and the parameter estimates used to generate the bootstrap samples differ from the true values of the parameters.

As we discuss in Davidson and MacKinnon (1997a), the finite-sample performance of a bootstrap test at level α under the null hypothesis depends, to leading order, on the bias of the bootstrap estimate of the size- α critical value. In the case of the parametric bootstrap, such a bias can arise for two reasons. Firstly, the parameter estimates may themselves be biased. Thus, the bootstrap samples will be drawn from distributions that, on average, have the wrong parameter values. This is a problem for the J test, but only to a very limited extent. Both β and σ^2 are estimated unbiasedly by $\hat{\beta}$ and s^2 , respectively, but $\hat{\beta}/s$ does not provide an unbiased estimate of β/σ . However, as we shall see, the extent of the bias is very small. Secondly, the bootstrap estimate of the size- α critical value will generally be biased whenever the critical value depends on the parameters in a nonlinear way. The extent

of this bias depends on the variance of the parameter estimates, becoming larger as the variance increases. In both cases, what matters is the shape of the “critical value function,” which relates the critical value for a test at a specified level to the values of the parameters.

As Figures 1 through 3 make clear, the J test is emphatically not pivotal. Figures 4 through 6 suggest that the distribution of the J_M test depends much less on the value of θ than does the distribution of J , but it too is not pivotal. Thus, one might conjecture that the bootstrap J test would not perform particularly well in small samples, and that the bootstrap J_M test would perform much better, but still by no means perfectly. As we shall see in the next section, these conjectures are only partially correct. It is true that the bootstrap J_M test performs distinctly better than the bootstrap J test. But the latter nevertheless performs remarkably well. The reason for this is simply that the only parameters which matter for the distribution, namely, the vector θ , are always estimated with small bias and small variance, relative to the rate at which the critical values vary with θ . This is illustrated in Figures 7 and 8, which show, respectively, the bias and the standard error of $\hat{\theta}_1$ as a function of N for various values of (θ_1, θ_2) for the same case as Figure 1. We see that both the bias and the standard error are particularly small when θ_1 and θ_2 are close to zero.

Although Figures 1 through 6 do not actually show critical value functions, the latter will have essentially the same shapes as do the rejection frequencies shown in these figures. Thus, from Figures 1 through 3, we see that the critical value function for the J test has almost no curvature for values of the θ_i greater than 2 or 3, is strongly concave for values very near 0, and is somewhat convex for values between about 0.4 and 1 or 2, depending on the sample size. The theory developed in Davidson and MacKinnon (1997a) suggests that the J test should overreject for values of θ very near zero, where the critical value function is concave, and underreject for somewhat larger values, where it is convex. However, because θ is estimated so precisely, especially in the region where its value matters most, the theory also suggests that the extent of overrejection or underrejection should be modest.

7. Monte Carlo Results for Bootstrap Tests

In this section, we report the results of several Monte Carlo experiments which largely confirm the predictions made in the previous section. If anything, the bootstrap tests work even better than we might expect them to. Each of the experiments used 500,000 replications, with 399 bootstraps per replication. Thus a total of 20 million J and 20 million J_M statistics were computed for each sample size. The extremely large number of replications ensures that experimental error is very small, which it is essential to ensure in this case because the bootstrap tests perform so well. We obtained results for all values of N between 10 and 50 by using the technique discussed in Section 5. Each experiment required about 17 hours of CPU time on a Pentium Pro 200 computer. It would have been totally infeasible to use anything close to 500,000 replications, or to deal with so many sample sizes, without making use of

the result (28). One disadvantage of this procedure, however, is that the results for the various values of N within a single set of experiments are positively correlated, probably quite highly correlated for values of N that are close to each other.

The results of four experiments are shown in Figure 9. These experiments correspond to the case dealt with in Figures 1 through 6, in which $d_1 = 0.9$ and $d_2 = 0.5$. We see that, as predicted, the bootstrap J test overrejects for very small values of θ_1 and θ_2 and underrejects a little bit for somewhat larger values. The overrejection remains noticeable even for moderate sample sizes, while the underrejection vanishes rapidly as N increases. In contrast, the J_M test performs almost perfectly in all cases. Notice the scale of the vertical axes in Figure 9. By most standards, even the J test performs extraordinarily well.

Figure 10 presents the results of four more experiments. These experiments are similar to those reported in Figure 9, except that $d_1 = 0.1$ and $d_2 = 0.2$. The results for $\theta_1 = \theta_2 = 0.1$ and for $\theta_1 = \theta_2 = 1.0$ are quite similar to the previous ones, but the results for the other two cases are somewhat different. For $\theta_1 = \theta_2 = 0.2$, the bootstrap J test actually underrejects for very small values of N , and the J_M test underrejects very slightly for all values of N . For $\theta_1 = \theta_2 = 0.5$, both tests tend to underreject slightly for small values of N , and there is a range over which the bootstrap J test actually performs a little bit better than the bootstrap J_M test.

Like all results of Monte Carlo experiments, the results in Figures 9 and 10 cover only a small fraction of the space of possible models and parameter values. However, we believe that they should be taken much more seriously than most results of Monte Carlo experiments, because the experimental design is based on a detailed understanding of what determines the finite-sample distributions of the tests. We have deliberately chosen cases for which the original J test works very badly and for which the bootstrap tests could be expected to work badly. The H_2 model contains 6 regressors that are not in the H_1 model, and in all the experiments θ is relatively small. It would only be possible to make the bootstrap tests perform substantially worse than they do in these experiments if we could find a case for which the critical value function was much more nonlinear, or for which the estimates of θ were much less precise, than they are in these experiments. If such a case exists, we believe that it must be quite pathological.

8. Summary and Conclusion

The original J test for nonnested, linear regression models often performs very poorly in finite samples. Previous work, based on rather imprecise Monte Carlo results, has suggested that bootstrapping the test greatly improves its finite-sample performance. In this paper, we have explained why it does so. In Section 3, we provided a detailed analysis of what determines the finite-sample distribution of the J test. In Section 4, we used this analysis to propose a new test, called the J_M test, that should generally perform even better when bootstrapped. In Sections 5 and 6, we explained why the J and J_M tests can be expected to perform well when

bootstrapped. We then confirmed our predictions via Monte Carlo experiments in Section 7. Our theoretical analysis made it possible for these Monte Carlo experiments to employ a very large number of replications, and thus to yield very accurate results. Because the performance of the bootstrap tests is so good, this is essential if the experimental results are not to be dominated by experimental noise.

From the point of view of practitioners, the message of this paper is quite simple. Unless the sample size is quite large, the H_1 model fits well, and the number of regressors in the H_2 model that are not also in H_1 is small, one should be extremely wary of the ordinary J test. It is very easy to bootstrap the J test, and, in almost all cases, the bootstrap J test should be perfectly reliable. However, if the sample size is extremely small and the model under test fits very poorly, it may be worthwhile to use the bootstrap J_M test instead.

References

- Beran, R. (1988). "Prepivoting test statistics: a bootstrap view of asymptotic refinements," *Journal of the American Statistical Association*, 83, 687–697.
- 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. and J. G. MacKinnon (1981). "Several tests for model specification in the presence of alternative hypotheses," *Econometrica*, 49, 781–793.
- Davidson, R. and J. G. MacKinnon (1982). "Some non-nested hypothesis tests and the relations among them," *Review of Economic Studies*, 49, 551–565.
- Davidson, R. and J. G. MacKinnon (1997a). "The size distortion of bootstrap tests," Queen's Institute for Economic Research Discussion Paper No. 936 (revised).
- Davidson, R. and J. G. MacKinnon (1997b). "Bootstrap tests: How many bootstraps?," Queen's Institute for Economic Research Discussion Paper No. 951.
- Fan, Y., and Q. Li (1995). "Bootstrapping J -type tests for non-nested regression models," *Economics Letters*, 48, 107–112.
- Fisher, G. R., and M. McAleer (1981). "Alternative procedures and associated tests of significance for non-nested hypotheses," *Journal of Econometrics*, 16, 103–119.

- Godfrey, L. G. (1983). “Testing non-nested models after estimation by instrumental variables or least squares,” *Econometrica*, 51, 355–365.
- Godfrey, L. G. (1997). “Tests of non-nested regression models: Some results on small sample behaviour and the bootstrap,” *Journal of Econometrics*, forthcoming.
- Godfrey, L. G., and M. H. Pesaran (1983). “Tests of non-nested regression models: small sample adjustments and Monte Carlo evidence,” *Journal of Econometrics*, 21, 133–154.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, New York, Springer-Verlag.
- Horowitz, J. L. (1994). “Bootstrap-based critical values for the information matrix test,” *Journal of Econometrics*, 61, 395–411.
- McAleer, M. (1995). “The significance of testing empirical non-nested models,” *Journal of Econometrics*, 67, 149–171.
- Michelis, L. (1995). “The null distribution of nonnested tests with nearly orthogonal regression models,” *Econometric Theory*, 11, 1177–1178.
- Michelis, L. (1996). “The null distribution of nonnested tests with nearly orthogonal regression models — solution,” *Econometric Theory*, 12, 870–871.
- Pesaran, M. H. (1974). “On the general problem of model selection,” *Review of Economic Studies*, 41, 153–171.

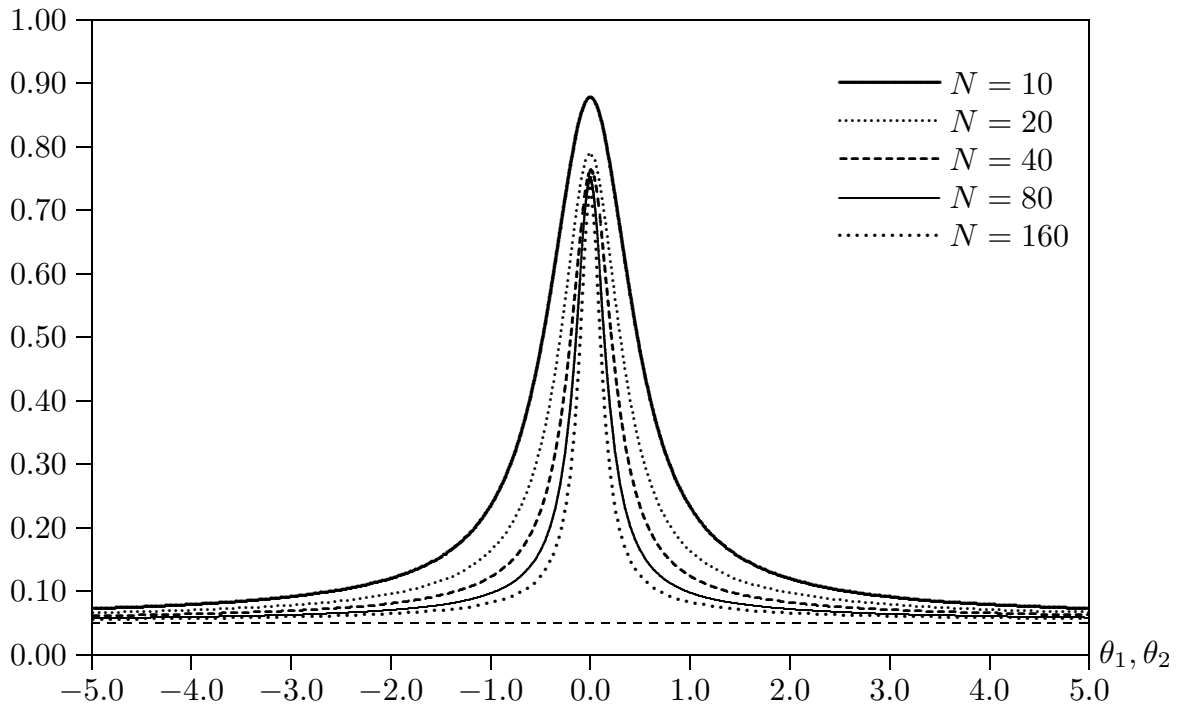


Figure 1. Rejection Frequencies for J Test, Both Parameters Varying

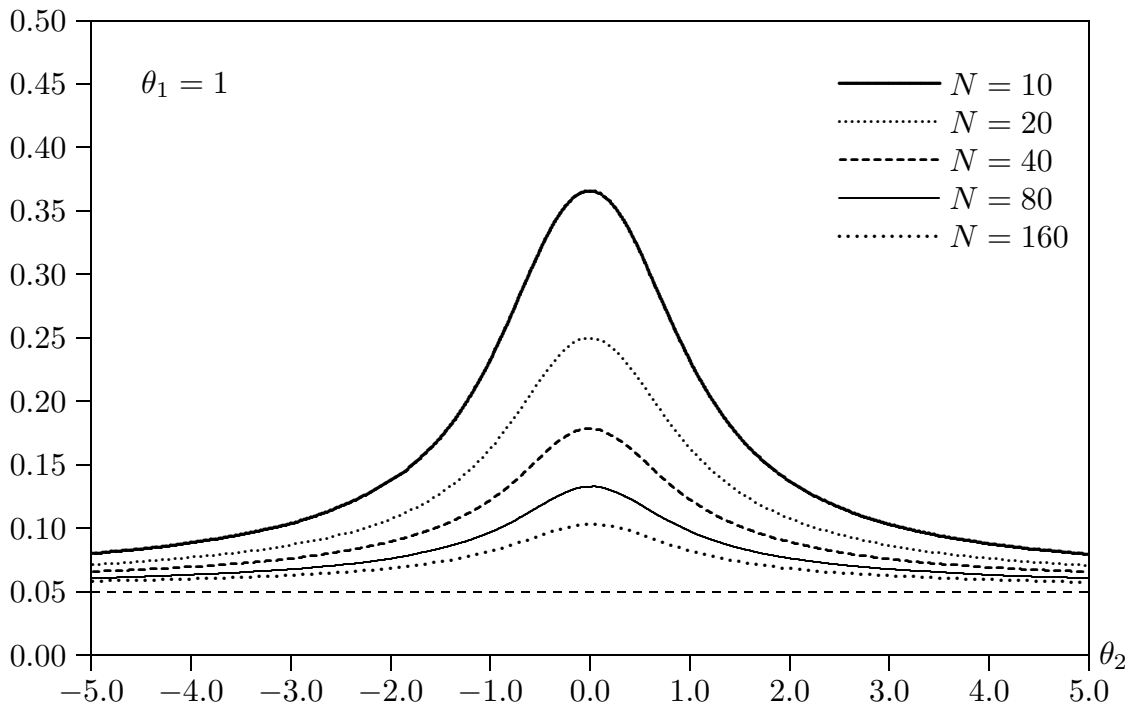


Figure 2. Rejection Frequencies for J Test, One Parameter Varying

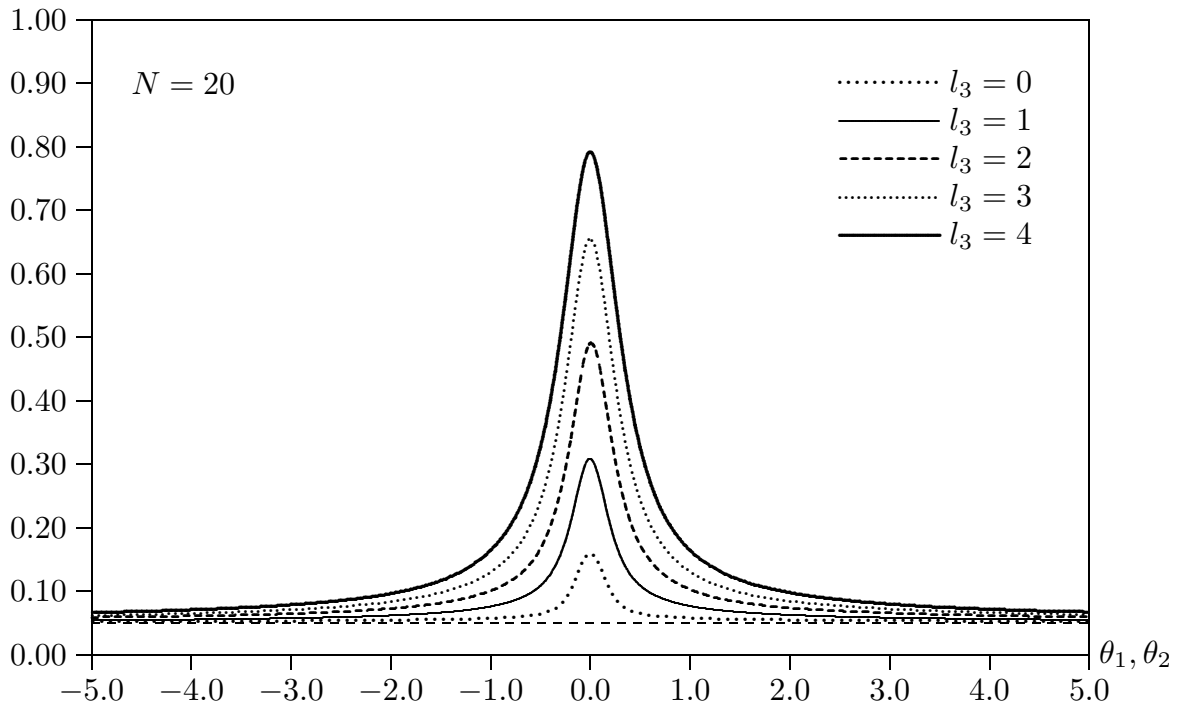


Figure 3. Rejection Frequencies for J Test, Both Parameters Varying

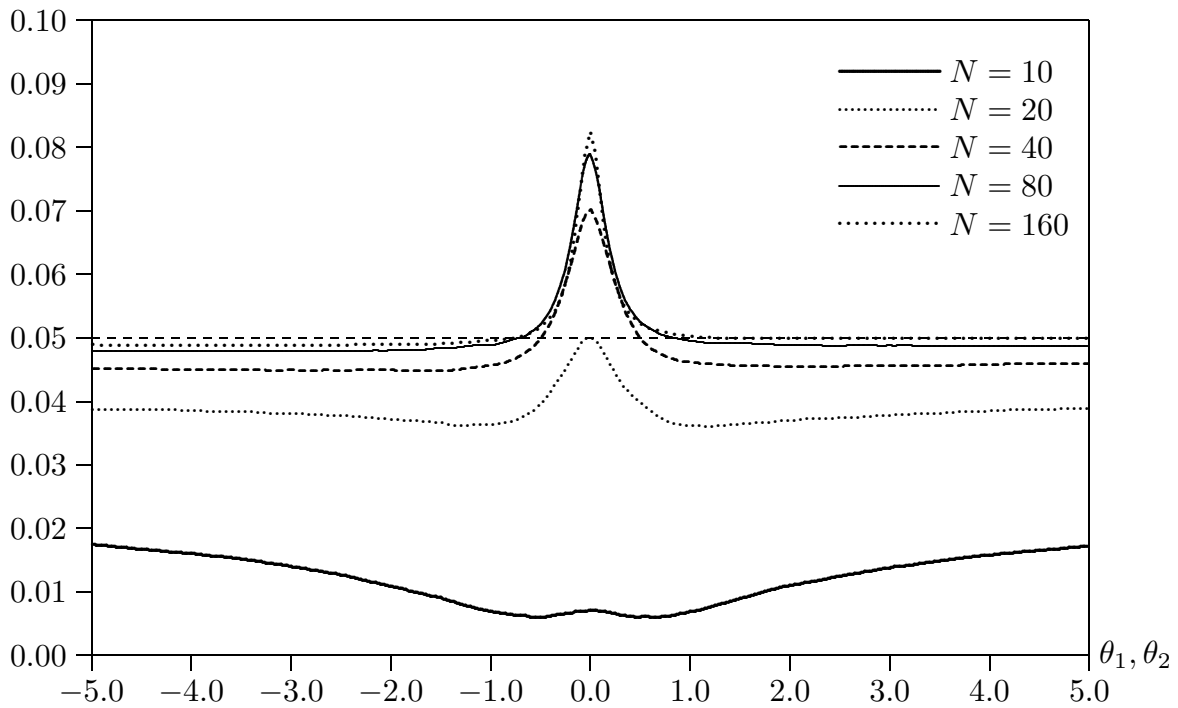


Figure 4. Rejection Frequencies for J_M Test, Both Parameters Varying

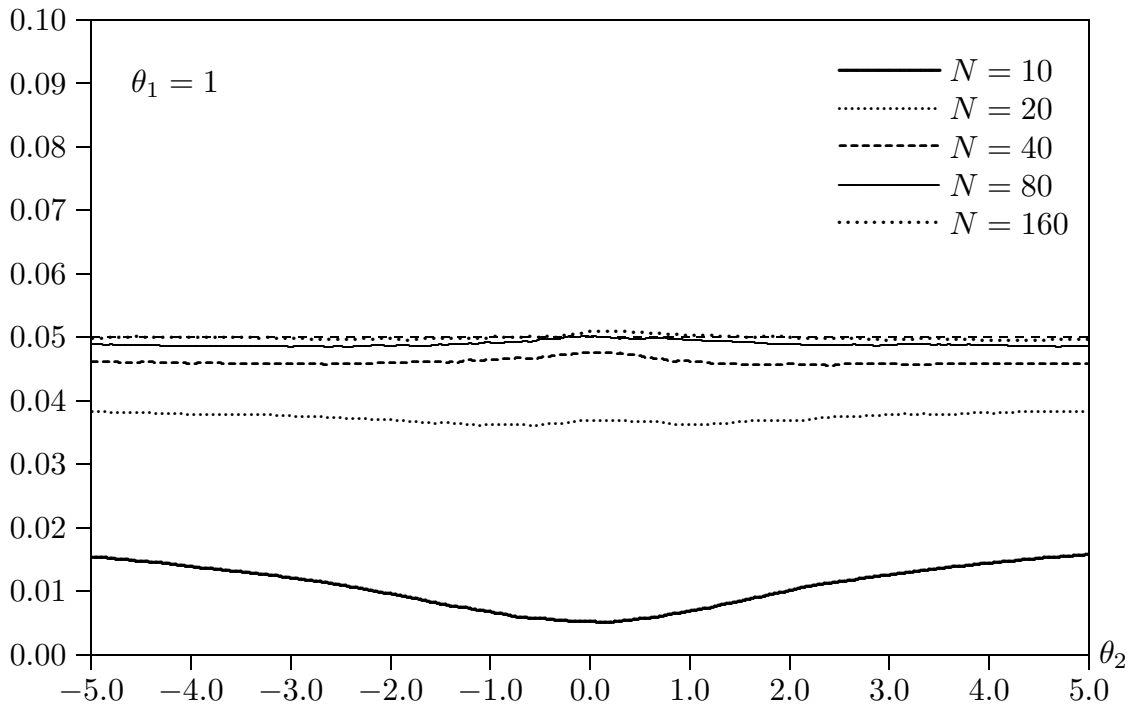


Figure 5. Rejection Frequencies for J_M Test, One Parameter Varying

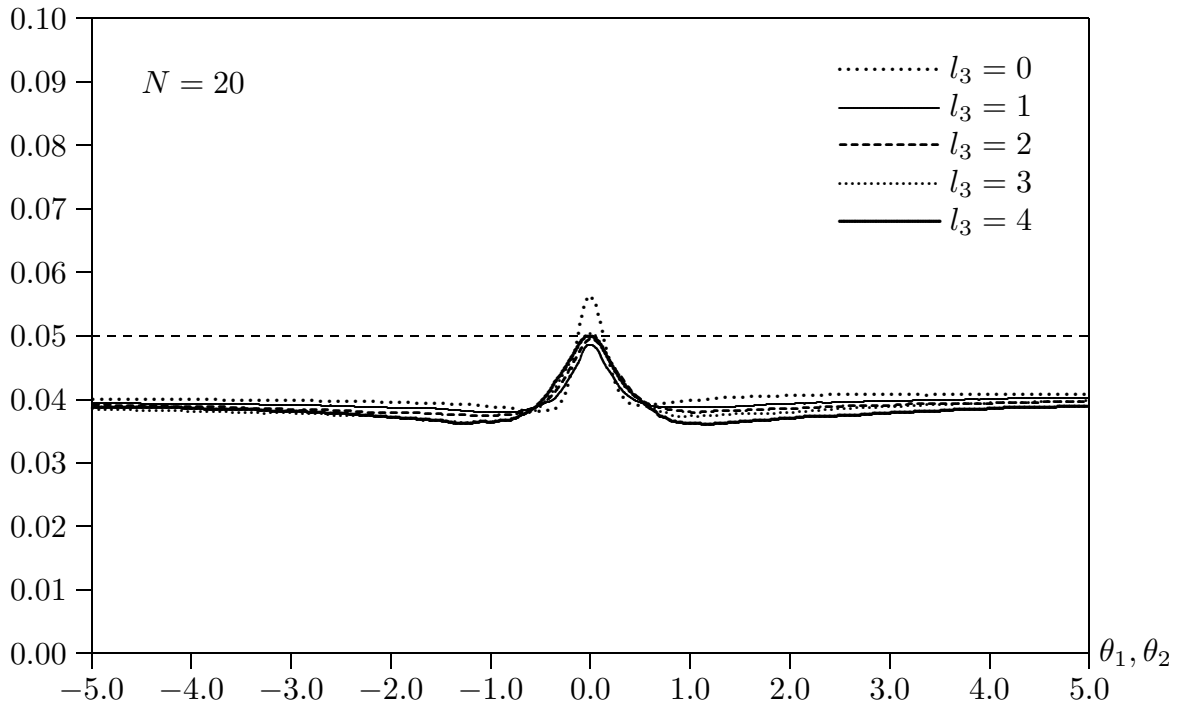


Figure 6. Rejection Frequencies for J_M Test, Both Parameters Varying

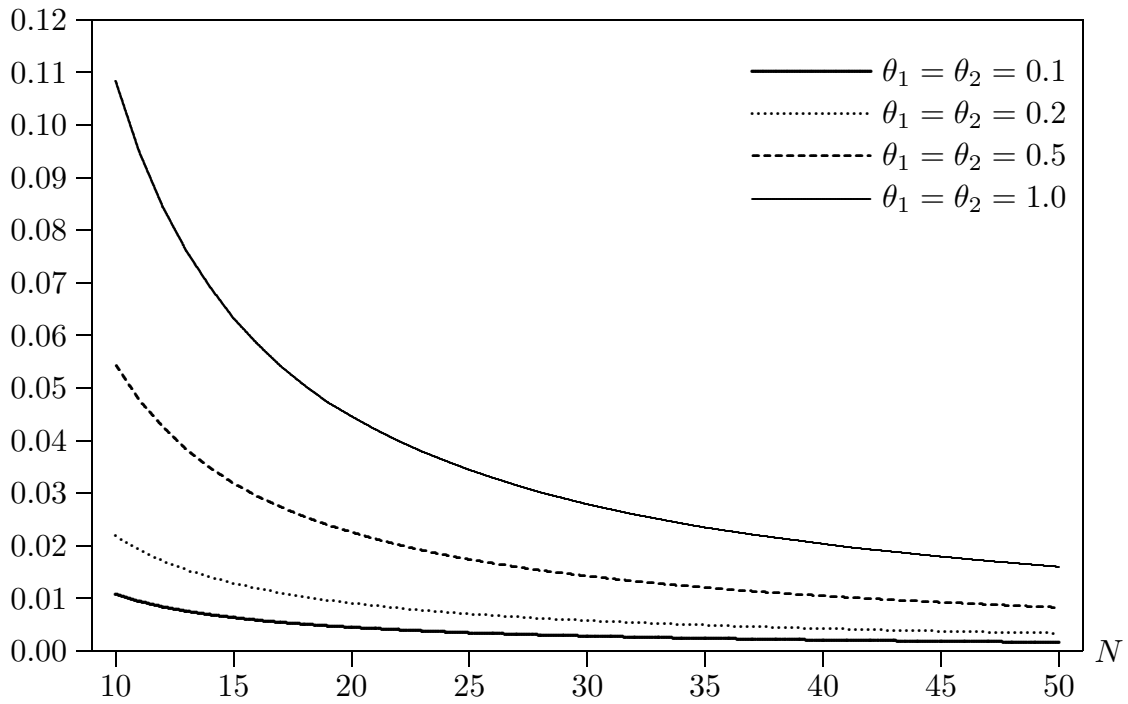


Figure 7. Biases of Estimates of θ_1

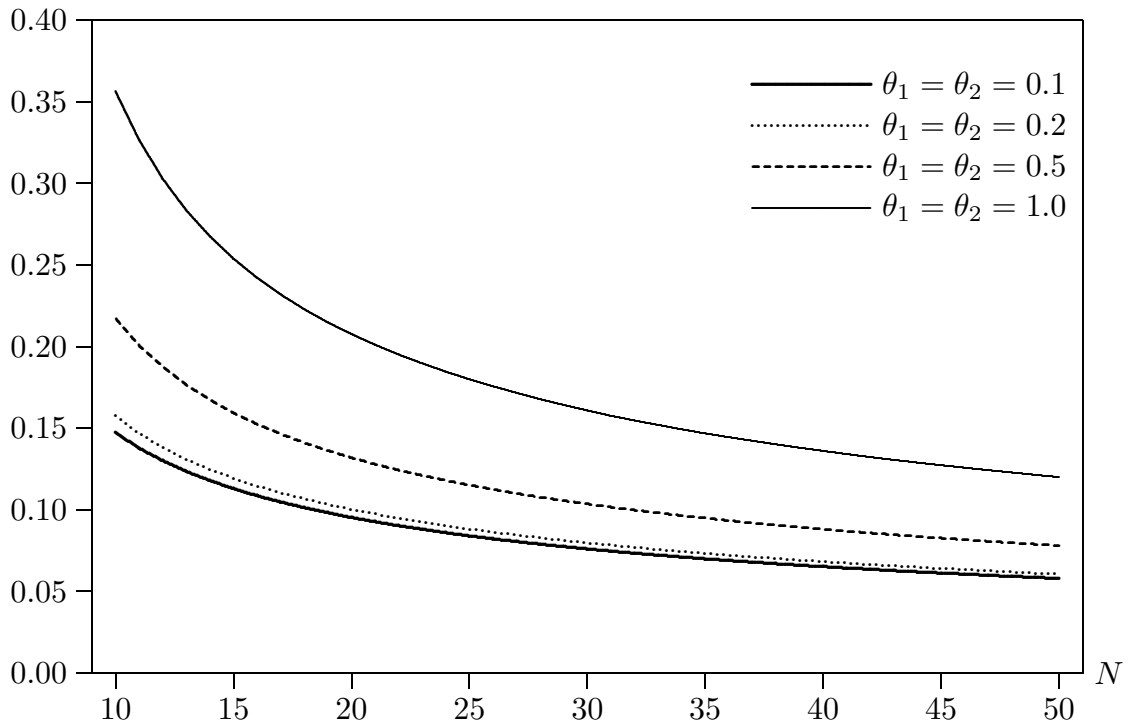


Figure 8. Standard Errors of Estimates of θ_1

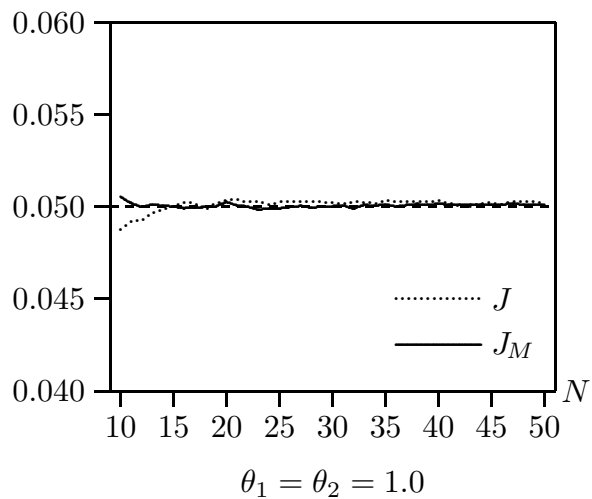
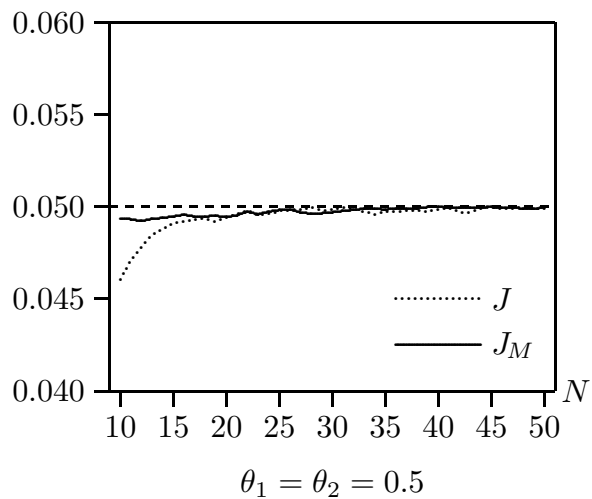
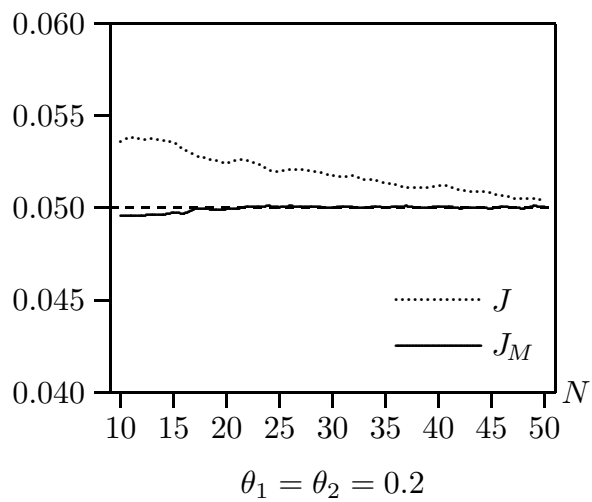
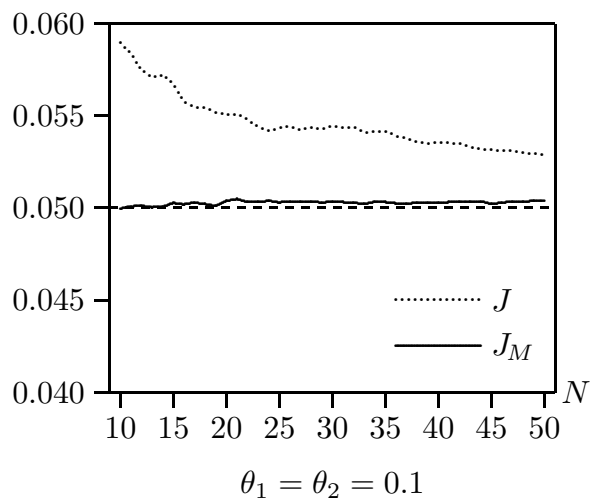


Figure 9. Rejection Frequencies for Bootstrap J and J_M Tests at .05 Level, Case 1

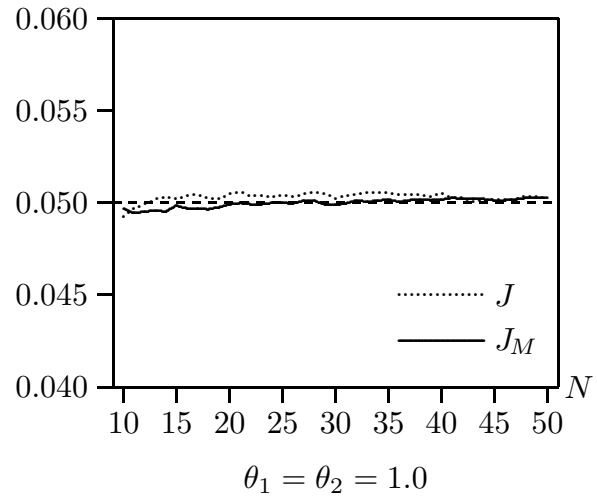
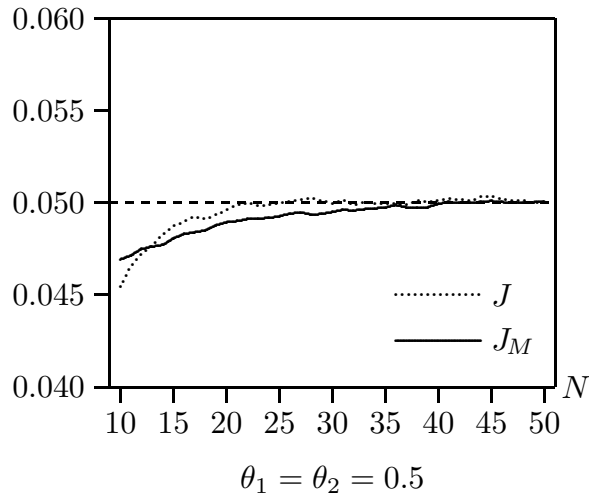
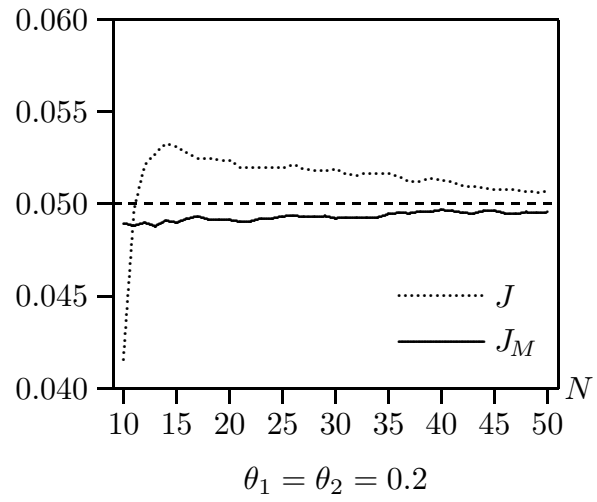
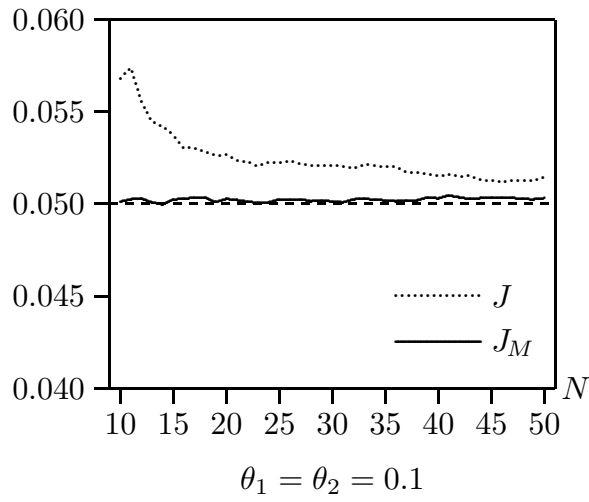


Figure 10. Rejection Frequencies for Bootstrap J and J_M Tests at .05 Level, Case 2