Bootstrap J Tests of Nonnested Linear Regression Models

by

Russell Davidson*

and

James G. MacKinnon**

*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

**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 overrejects very severely as an asymptotic test. We provide a theoretical analysis which explains why and when it performs badly. This analysis implies that, except in certain extreme cases, the J test will perform very well when bootstrapped. Using several methods to speed up the simulations, we obtain extremely accurate Monte Carlo results on the finite-sample performance of the bootstrapped J test. These results fully support the predictions of our theoretical analysis, even in contexts where the analysis is not strictly applicable.

This research was supported, in part, by grants from the Social Sciences and Humanities Research Council of Canada. We are grateful to several referees, Joel Horowitz, and workshop participants at GREQAM, Dalhousie University, the University of Rochester, CORE, the University of Cyprus, the Swedish School of Economics, Cornell University, Yale University, and the University of Iowa for comments on earlier versions.

First version, December, 2000; final revision, July, 2001

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 many Monte Carlo experiments have shown, its finite-sample distribution can be very far from the $N(0,1)$ distribution that it follows asymptotically.

Several ways have been proposed to improve the finite-sample properties of the J test. Fisher and McAleer (1981) proposed a variant, 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 (1998) 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 often works very well. However, neither paper provides any theoretical explanation of why it does so.

In this paper, we develop a theoretical approach that enables us to show precisely what determines the finite-sample distribution of the J test. We explain why it often works very badly without bootstrapping and why it almost always works very well indeed when bootstrapped. The theory allows us to identify situations in which the tests can be expected to achieve their worst behavior, and our Monte Carlo experiments focus on these. Since the tests perform very well even in such situations, the experiments need to be very accurate. Fortunately, our theory provides a low-cost way to perform experiments that use extremely large numbers of replications.

The assumptions needed for our theoretical analysis are fairly restrictive: The errors are assumed to be normally distributed, and the regressors are assumed to be exogenous. However, additional Monte Carlo experiments strongly suggest that these assumptions are not crucial. Even when both of them are violated, the bootstrap J test performs in almost exactly the same way as it does when they are satisfied.

In the next section, we briefly describe the J test. In Section 3, we derive a theoretical expression for the test statistic and use it to obtain a number of interesting results. In Section 4, we use a combination of theory and simulation to study the finite-sample properties of the asymptotic J test. In Section 5, we study the finite-sample properties of the bootstrap J test. In Section 6, we relax the restrictive assumptions made up to this point and show that the bootstrap J test works extraordinarily well in almost every case in which a nonnested test is worth doing. Finally, in Section 7, we briefly discuss the effect of bootstrapping on the power of the J test.

2. The J Test

Although the J test can be applied to both linear and nonlinear regression models, we restrict our attention to the linear case, since it would be extremely difficult to obtain general results about the finite-sample properties of the J test in the nonlinear case. Consider two nonnested, linear regression models with IID normal errors:

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

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

$$\boldsymbol{y} = \boldsymbol{Xb} + \alpha \boldsymbol{P_Z y} + \text{residuals}, \tag{2}$$

where $\boldsymbol{P_Z} \equiv \boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top$, so that $\boldsymbol{P_Z y}$ is the vector of fitted values from OLS estimation of H_2. Asymptotically, the J statistic is distributed as $N(0,1)$ under H_1. In practice, the $t(n-k-1)$ distribution is often used for finite-sample inference although, except in one special case, there is no formal justification for doing so.

The J statistic for testing H_1 can be written as

$$J = \frac{(n-k-1)^{1/2}\boldsymbol{y}^\top\boldsymbol{P_Z M_X y}}{\left[(\boldsymbol{y}^\top\boldsymbol{M_X y})(\boldsymbol{y}^\top\boldsymbol{P_Z M_X P_Z y}) - (\boldsymbol{y}^\top\boldsymbol{P_Z M_X y})^2\right]^{1/2}}. \tag{3}$$

where $\boldsymbol{P_X} \equiv \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ and $\boldsymbol{M_X} \equiv \mathbf{I} - \boldsymbol{P_X}$. Like any t statistic, J is $(n-k-1)^{1/2}$ times the cotangent of the angle between two vectors in n–dimensional Euclidean space. The vectors are $\boldsymbol{M_X y}$, the vector of OLS residuals from estimating H_1, and $\boldsymbol{M_X P_Z y}$. As can be seen from (3), J can be expressed in terms of the three scalar products defined by these two vectors: $\boldsymbol{y}^\top\boldsymbol{M_X y}$, which is the sum of squared residuals from H_1, $\boldsymbol{y}^\top\boldsymbol{P_Z M_X y}$, and $\boldsymbol{y}^\top\boldsymbol{P_Z M_X P_Z y}$. Since J depends on the regressor matrices only through the projections $\boldsymbol{P_X}$ and $\boldsymbol{P_Z}$, it is invariant to changes in $\boldsymbol{X}$ and $\boldsymbol{Z}$ that do not change the linear spaces spanned by the columns of these matrices, which we will denote by $\mathcal{S}(\boldsymbol{X})$ and $\mathcal{S}(\boldsymbol{Z})$, respectively.

3. The Distribution of the J Statistic

Although no closed-form expression can be found for the distribution of the J statistic (3) under the hypothesis H_1, the statistic can be expressed as a function of a small number of standard normal variables and a chi-squared variable, all mutually independent. The function also depends on the true values of the parameters, $\boldsymbol{\beta}$ and σ^2, and on certain features of the regressors of the two nonnested models, which are assumed to be exogenous and are thus treated as nonrandom in our analysis. Some parts of the analysis, which necessarily involves some notational complexity, are relegated to the Appendix.

We begin by studying some geometrical features of the space $\mathcal{S}(\boldsymbol{X}, \boldsymbol{Z})$ spanned by the regressors $\boldsymbol{X}$ and $\boldsymbol{Z}$. We denote this space by S_0. The spaces $\mathcal{S}(\boldsymbol{X})$ and $\mathcal{S}(\boldsymbol{Z})$ spanned by the columns of $\boldsymbol{X}$ and $\boldsymbol{Z}$, respectively, will, in general, have an intersection of dimension greater than 0. Denote this intersection by S_1. Denote by S_2 the intersection of $\mathcal{S}(\boldsymbol{X})$ with the orthogonal complement $\mathcal{S}^{\perp}(\boldsymbol{Z})$ of $\mathcal{S}(\boldsymbol{Z})$, and by S_3 the orthogonal complement of the direct sum $S_1 \oplus S_2$ in $\mathcal{S}(\boldsymbol{X})$. Thus the spaces S_1, S_2, and S_3 are mutually orthogonal, and

$$\mathcal{S}(\boldsymbol{X}) = S_1 \oplus S_2 \oplus S_3.$$

Similarly, let S_4 be the intersection of $\mathcal{S}(\boldsymbol{Z})$ with $\mathcal{S}^{\perp}(\boldsymbol{X})$, and let S_5 be the orthogonal complement of $S_1 \oplus S_4$ in $\mathcal{S}(\boldsymbol{Z})$. Thus S_1, S_4, and S_5 are mutually orthogonal, and

$$\mathcal{S}(\boldsymbol{Z}) = S_1 \oplus S_4 \oplus S_5.$$

Further, since $S_4 \subseteq \mathcal{S}^{\perp}(\boldsymbol{X})$, S_4 is orthogonal to S_2 and S_3 as well as to S_1. Similarly, S_5 is orthogonal to S_2, but not to S_3, since all vectors in $\mathcal{S}(\boldsymbol{Z})$ that are orthogonal to all of $\mathcal{S}(\boldsymbol{X})$ are in S_4, and hence not in S_5.

If we denote by $\boldsymbol{P}_i$ the orthogonal projections on to the S_i, $i = 0, 1, \ldots, 5$, and by $\boldsymbol{M}_i \equiv \mathbf{I} - \boldsymbol{P}_i$ the corresponding complementary projections on to the orthogonal complements of the S_i, we have

$$\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{P}_1 + \boldsymbol{P}_2 + \boldsymbol{P}_3 \quad \text{and} \quad \boldsymbol{P}_{\boldsymbol{Z}} = \boldsymbol{P}_1 + \boldsymbol{P}_4 + \boldsymbol{P}_5. \tag{4}$$

In order to express $\boldsymbol{P}_0$ as a sum of projections on to mutually orthogonal spaces, we define $S_6 = \boldsymbol{M}_{\boldsymbol{X}} S_5$, that is, the image of S_5 under $\boldsymbol{M}_{\boldsymbol{X}}$. In fact,

$$S_6 = (\mathbf{I} - \boldsymbol{P}_1 - \boldsymbol{P}_2 - \boldsymbol{P}_3) S_5 = S_5 - \boldsymbol{P}_3 S_5 = \boldsymbol{M}_3 S_5,$$

since S_5 is already orthogonal to S_1 and S_2. It follows that S_6 is orthogonal to everything that is orthogonal to S_3 and S_5, and by construction also to S_3. Thus S_1, S_2, S_3, S_4, and S_6 are all mutually orthogonal, and

$$\boldsymbol{P}_0 = \boldsymbol{P}_1 + \boldsymbol{P}_2 + \boldsymbol{P}_3 + \boldsymbol{P}_4 + \boldsymbol{P}_6. \tag{5}$$

If we denote the dimensions of the spaces S_i by s_i, (5) implies that $s_0 = s_1 + s_2 + s_3 + s_4 + s_6$. From (4), it follows that $k = s_1 + s_2 + s_3$ and $l = s_1 + s_4 + s_5$. Since $\dim(S_0) = k + l - s_1$, it can be seen that $s_5 = s_6$. In fact, it is also true that $s_3 = s_5 = s_6$. This can be seen by considering $\boldsymbol{P}_{\boldsymbol{Z}}$ as a linear mapping from $\mathcal{S}(\boldsymbol{X})$ to $\mathcal{S}(\boldsymbol{Z})$. The null space of this mapping consists of all vectors in $\mathcal{S}(\boldsymbol{X})$ that are orthogonal to $\mathcal{S}(\boldsymbol{Z})$, which by construction is the space S_2. Thus the dimension of the image of $\mathcal{S}(\boldsymbol{X})$ under $\boldsymbol{P}_{\boldsymbol{Z}}$ is $k - s_2 = s_1 + s_3$. The orthogonal complement of this image in $\mathcal{S}(\boldsymbol{Z})$ is the set of vectors in $\boldsymbol{z} \in \mathcal{S}(\boldsymbol{Z})$ which are orthogonal to all vectors of the form $\boldsymbol{P_Z x}$, for any $\boldsymbol{x} \in \mathcal{S}(\boldsymbol{X})$. For such a $\boldsymbol{z}$, we have

$$\boldsymbol{z}^\top \boldsymbol{P_Z x} = \boldsymbol{z}^\top \boldsymbol{x} = 0 \quad \text{for all } \boldsymbol{x} \in \mathcal{S}(\boldsymbol{X}).$$

This is just the defining condition for a vector $\boldsymbol{z}$ to be in S_4. Thus the dimension of the orthogonal complement in $\mathcal{S}(\boldsymbol{Z})$ of the image of $\mathcal{S}(\boldsymbol{X})$ under $\boldsymbol{P_Z}$ is s_4, and so the dimension of the image itself is $l - s_4 = s_1 + s_5$. We saw above that this dimension is also equal to $s_1 + s_3$, and so $s_3 = s_5$, as we wished to show. Thus we see that S_5 is the image of $\mathcal{S}(\boldsymbol{X})$ under $\boldsymbol{P_Z}$.

We can now state our principal theoretical result.

Theorem 1

Suppose that the n–vector $\boldsymbol{y}$ is generated by the data-generating process H_1, defined in (1), for specific values of $\boldsymbol{\beta}$ and σ^2. Then the J statistic (3) can be written as

$$\frac{(n-k-1)^{1/2}\big(\boldsymbol{\theta}^\top \boldsymbol{v}_6 + \boldsymbol{v}_6^\top \boldsymbol{v}_5 + \|\boldsymbol{v}_4\|^2\big)}{\begin{bmatrix} \big(V^2 + \|\boldsymbol{v}_4\|^2 + \|\boldsymbol{v}_6\|^2\big)\big(\|\boldsymbol{\theta}\|^2 + 2\boldsymbol{\theta}^\top \boldsymbol{v}_5 + \|\boldsymbol{v}_4\|^2 + \|\boldsymbol{v}_5\|^2\big) \\ -\big(\boldsymbol{\theta}^\top \boldsymbol{v}_6 + \boldsymbol{v}_6^\top \boldsymbol{v}_5 + \|\boldsymbol{v}_4\|^2\big)^2 \end{bmatrix}^{1/2}}. \tag{6}$$

In (6), the random s_4–vector $\boldsymbol{v}_4$ is distributed as $N(\mathbf{0}, \mathbf{I})$, the s_5–vectors $\boldsymbol{v}_6$ and $\boldsymbol{v}_3$ are distributed as $N(\mathbf{0}, \mathbf{I})$, and the random scalar V^2 is distributed as $\chi^2(n - s_0)$. All these variables are mutually independent. The s_5–vector $\boldsymbol{v}_5$ is defined in terms of $\boldsymbol{v}_3$ and $\boldsymbol{v}_6$ by the formula

$$\boldsymbol{v}_5 = \boldsymbol{\Delta}(\boldsymbol{D}\boldsymbol{v}_3 + \boldsymbol{\Delta}\boldsymbol{v}_6), \tag{7}$$

where $\boldsymbol{D}$ is an $s_5 \times s_5$ diagonal matrix, the elements of which are the canonical correlations between the spaces S_3 and S_5, and $\boldsymbol{\Delta}$ is a positive definite diagonal matrix defined by $\boldsymbol{\Delta}^2 = \mathbf{I} - \boldsymbol{D}^2$. Further, the s_5–vector $\boldsymbol{\theta}$ satisfies the relation

$$\|\boldsymbol{\theta}\|^2 = \boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{P_Z M_X P_Z X \beta}/\sigma^2 = \|\boldsymbol{M_X P_Z X\beta}\|^2/\sigma^2. \tag{8}$$

Proof: See Appendix.

The formula (6) has a great many interesting properties. It expresses J in terms of deterministic quantities and independent random variables all of which have known distributions. It is therefore possible to simulate J directly from (6), by drawing these random variables from their respective distributions. This will turn out to be extremely useful when we undertake Monte Carlo experiments. Note that $\boldsymbol{v}_5$ is *not* independent of $\boldsymbol{v}_3$ and $\boldsymbol{v}_6$; see (7). It is used in (6) in place of $\boldsymbol{v}_3$, which is independent of $\boldsymbol{v}_6$, only because doing so makes (6) much easier to write.

Since $n - k = n - s_1 - s_2 - s_5$, the only dimensions, or degrees of freedom, that influence (6) are s_4, s_5, and $N \equiv n - s_1 - s_2$. To see this, observe that J depends on n or N only through $n - k - 1 = N - s_5 - 1$ in the numerator of (6), and through V^2, which has $n - s_0 = N - s_4 - 2s_5$ degrees of freedom, in the denominator.

In particular, the dimensions of the spaces S_1 and S_2 have no effect other than to reduce the number of degrees of freedom. For the remainder of the paper, therefore, we use N rather than n to measure the sample size. The actual design of the matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ influences (6) only through the canonical correlation coefficients d_i, which are determined by the s_5–dimensional spaces S_3 and S_5 alone.

The other thing that influences (6) is the parameter vector $\boldsymbol{\theta}$. In fact, J depends on $\boldsymbol{\theta}$ only through $\|\boldsymbol{\theta}\|^2$, $\boldsymbol{\theta}^\top \boldsymbol{v}_6$, and $\boldsymbol{\theta}^\top \boldsymbol{v}_5$. Since all components of $\boldsymbol{v}_5$ and $\boldsymbol{v}_6$ have distributions that are symmetric about the origin, it follows that the distribution of J is invariant to changes in the sign of any component of $\boldsymbol{\theta}$. From (8), it is clear that, unless for some reason $\boldsymbol{\theta} = \mathbf{0}$, $\|\boldsymbol{\theta}\|^2 = O(n) = O(N)$ as $N \to \infty$. Equivalently, $\boldsymbol{\theta} = O(N^{1/2})$. The length of the vector $\boldsymbol{\theta}$ will turn out to be crucial to the finite-sample properties of the J test. Evidently, $\|\boldsymbol{\theta}\|$ will tend to be smaller the smaller is N, the larger is σ^2 relative to $\boldsymbol{\beta}$, and the closer is $\boldsymbol{P_Z X}$ to lying in $\mathcal{S}(\boldsymbol{X})$.

In order to determine the asymptotic distribution of J, for $\boldsymbol{\theta} \neq \mathbf{0}$, we divide numerator and denominator in (6) by N and then let $N \to \infty$. Since $\operatorname{plim}_{N\to\infty} V^2/N = 1$ and $N^{-1/2}\boldsymbol{\theta} = O(1)$, it follows that the limit of (6) is

$$\operatorname*{plim}_{N\to\infty} J = \frac{\boldsymbol{\theta}^\top \boldsymbol{v}_6}{\|\boldsymbol{\theta}\|}. \tag{9}$$

It is clear that this has the standard normal distribution independently of $\boldsymbol{\theta}$.

By use of (6), J can be expressed as a cotangent between two vectors in a Euclidean space of dimension lower than n. It is easy to see that (6) is $(N - s_5 - 1)^{1/2}$ times the cotangent between two vectors in a space of dimension $s_5 + s_4 + 1$. These two vectors can be expressed as

$$\begin{bmatrix} \boldsymbol{\theta} + \boldsymbol{v}_5 \\ \boldsymbol{v}_4 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \boldsymbol{v}_6 \\ \boldsymbol{v}_4 \\ V \end{bmatrix}. \tag{10}$$

This fact will be seen later to provide an intuitive explanation of some properties of the J test.

When we discuss bootstrapping the J test in Section 5, it will be necessary to have an expression for the estimate $\hat{\boldsymbol{\theta}}$ of the parameter vector $\boldsymbol{\theta}$ given by OLS estimation of model H_1. It is shown in the Appendix that the appropriate expression is

$$\hat{\boldsymbol{\theta}} = \left(\frac{V^2 + \|\boldsymbol{v}_4\|^2 + \|\boldsymbol{v}_6\|^2}{N - s_5} \right)^{-1/2} \big(\boldsymbol{\theta} + \boldsymbol{\Delta} \boldsymbol{D} \boldsymbol{v}_3\big). \tag{11}$$

The two factors in this expression are independent. The first is a $\chi^2(N - s_5)$ variable divided by $N - s_5$ and raised to the power $-1/2$, and the second is normal with mean $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Delta}^2 \boldsymbol{D}^2$; recall that $\boldsymbol{\Delta}$ and $\boldsymbol{D}$ are diagonal, and they therefore commute.

It is possible to rederive many known results about the J test by considering special cases of (6). For example, if all the canonical correlations d_i tend to zero, we have the case of orthogonal models, because S_3 becomes orthogonal to S_5. The orthogonality implies that $\boldsymbol{X}^\top \boldsymbol{Z} = \mathbf{0}$, and so, from (8), $\boldsymbol{\theta} = \mathbf{0}$. Thus (9) is not defined and the J statistic does not have its usual asymptotic distribution. In addition, $\boldsymbol{D} = \mathbf{0}$ and $\boldsymbol{\Delta} = \mathbf{I}$, and it follows from (7) that $\boldsymbol{v}_5 = \boldsymbol{v}_6$. With this, (6) gives

$$
\begin{aligned}
J &= \frac{(n-k-1)^{1/2}(\|\boldsymbol{v}_6\|^2 + \|\boldsymbol{v}_4\|^2)}{\Big[\big(V^2 + \|\boldsymbol{v}_6\|^2 + \|\boldsymbol{v}_4\|^2\big)\big(\|\boldsymbol{v}_6\|^2 + \|\boldsymbol{v}_4\|^2\big) - \big(\|\boldsymbol{v}_6\|^2 + \|\boldsymbol{v}_4\|^2\big)^2\Big]^{1/2}} \\
&= \left(\frac{\|\boldsymbol{v}_6\|^2 + \|\boldsymbol{v}_4\|^2}{V^2/(n-k-1)}\right)^{1/2},
\end{aligned}
$$

the square of which is proportional to the F statistic that tests the null of H_1 against the alternative of a model in which the regressors span the whole of S_0. This special case, along with others related to it, has been studied by Michelis (1999).

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 $S_3 = S_5$. In this case, $\boldsymbol{D} = \mathbf{I}$, and $\boldsymbol{\Delta} = \mathbf{0}$. Once more $\boldsymbol{\theta} = \mathbf{0}$, and this time $\boldsymbol{v}_5 = \mathbf{0}$. The statistic (6) becomes

$$
\begin{aligned}
J &= \frac{(n-k-1)^{1/2}\|\boldsymbol{v}_4\|^2}{\big[\big(V^2 + \|\boldsymbol{v}_6\|^2 + \|\boldsymbol{v}_4\|^2\big)\|\boldsymbol{v}_4\|^2 - \|\boldsymbol{v}_4\|^4\big]^{1/2}} \\
&= \frac{\|\boldsymbol{v}_4\|}{\big[\big(V^2 + \|\boldsymbol{v}_6\|^2\big)/(n-k-1)\big]^{1/2}},
\end{aligned}
$$

the square of which is proportional to the F statistic against a nested alternative with extra regressors contained in S_4.

Next consider the special case in which $s_4 = 0$, $s_5 = 1$, that is, the case in which model H_2 contains just one regressor that is not in model H_1. In this case, $\boldsymbol{v}_6$, $\boldsymbol{v}_5$, and $\boldsymbol{\theta}$ all become scalars that we denote as v_6, v_5, and θ, $\boldsymbol{v}_4$ vanishes, and (6) becomes

$$
J = \frac{(n-k-1)^{1/2}(\theta v_6 + v_6 v_5)}{\big[(V^2 + v_6^2)(\theta^2 + 2\theta v_5 + v_5^2) - (\theta v_6 + v_6 v_5)^2\big]^{1/2}}.
$$

The denominator here is the square root of

$$
(V^2 + v_6^2)(\theta + v_5)^2 - v_6^2(\theta + v_5)^2 = V^2(\theta + v_5)^2,
$$

so that

$$
J = \frac{(n-k-1)^{1/2} v_6(\theta + v_5)}{V(\theta + v_5)} = (n-k-1)^{1/2}\frac{v_6}{V}. \tag{12}
$$

In this case, J does not depend on θ at all. It is in fact just the usual t statistic for a test of H_1 against an alternative with one extra regressor contained in S_5. This is the one case in which the J statistic actually has the Student's t distribution in finite samples under the null.

The J statistic could still be written as expression (6) if it were no longer assumed that the vector $\boldsymbol{u}$ is normally distributed. However, the quantities that appear in (6) have their specified distributions only under that assumption. A weaker and more plausible assumption is that $\boldsymbol{u} \sim \text{IID}(\mathbf{0}, \sigma^2\mathbf{I})$. All of the quantities in (6) that vanish asymptotically depend on the vectors $\boldsymbol{v}_4$, $\boldsymbol{v}_5$, and $\boldsymbol{v}_6$. Since these three vectors are suitably weighted sums of the components of $\boldsymbol{u}$ (see the Appendix), we can apply a central limit theorem to each of these vectors and conclude that it is approximately multivariate normal. Therefore, even though our results apply exactly only under the assumption of normal errors, there is good reason to expect that they will provide reasonable approximations under much weaker distributional assumptions.

4. Finite-Sample Performance of the *J* Test

Although analytic expressions for the distribution of the J statistic in finite samples are not available, efficient Monte Carlo methods based on the results of the last section can be used to study the properties of this distribution as functions of the specific aspects of the models H_1 and H_2 on which they depend. As Theorem 1 shows, these are just $N \equiv n - s_1 - s_2$, s_4, s_5, the canonical correlations d_i, and the parameter vector $\boldsymbol{\theta}$. It will be seen that, especially in the neighborhood of $\boldsymbol{\theta} = \mathbf{0}$, the statistic has a distribution very different from the $N(0, 1)$ or Student's $t(n - k - 1)$ distributions, except for the case $s_4 = 0$, $s_5 = 1$, in which J reduces to (12).

In our Monte Carlo experiments, instead of generating regressors and error terms, we generate realizations of $\boldsymbol{v}_3$, $\boldsymbol{v}_4$, $\boldsymbol{v}_6$, and V^2 and plug them into (6) to calculate realizations of the J statistic for specific choices of $\boldsymbol{\theta}$ and the d_i. By using N instead of n, we can ignore s_1 and s_2. Since V^2 is the only one of the random variables that depends on N, and V^2 for N is simply equal to V^2 for $N-1$ plus one more independent, squared, standard normal variable, we can inexpensively generate realizations of J for a large number of different sample sizes at the same time. In this way, it is feasible to conduct experiments with very large numbers of replications that yield extremely accurate results. This procedure does require us to assume that the error terms are normally distributed. However, that assumption will be relaxed in the Monte Carlo experiments of Section 6.

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 Figure 1, which is based on 250,000 replications. Figure 1 shows rejection frequencies for a one-tailed test at the nominal .05 level based on the Student's t distribution with $n - k - 1 = N - s_5 - 1$ degrees of freedom for a large number of values of $\boldsymbol{\theta}$ and some representative values of N. Figure 1 pertains to what we will use as the baseline case, with $s_4 = 4$, $s_5 = 2$, and $d_1 = d_2 = 0.55$. The J test can be expected to perform poorly in this case, because there are 6 regressors in H_2 that are not in H_1. The reason for the choice of $d_1 = d_2 = 0.55$ will be explained when we discuss bootstrap versions of the J test in the next section.

In Figure 1, both θ_1 and θ_2 vary together, which is equivalent to varying σ^2 while holding $\boldsymbol{\beta}$ fixed. Only positive values of θ_1 and θ_2 are considered, because the distribution of J is invariant under changes of sign of these components. Since $d_1 = d_2$ in this figure, it can be seen from (6) that the distribution of J is spherically symmetric with respect to $\boldsymbol{\theta}$, and so there is no loss of generality in setting $\theta_1 = \theta_2$, since the distribution depends only on $\|\boldsymbol{\theta}\|$. For the curve labelled $N = \infty$, formula (6) was replaced, not by the asymptotic formula (9), but by the limit of (6) as $N \to \infty$ with fixed $\boldsymbol{\theta}$, namely,

$$\frac{\boldsymbol{\theta}^\top \boldsymbol{v}_6 + \boldsymbol{v}_6{}^\top \boldsymbol{v}_5 + \|\boldsymbol{v}_4\|^2}{\left(\|\boldsymbol{\theta}\|^2 + 2\boldsymbol{\theta}^\top \boldsymbol{v}_5 + \|\boldsymbol{v}_4\|^2 + \|\boldsymbol{v}_5\|^2\right)^{1/2}}. \tag{13}$$

A striking feature of Figure 1 is that the curves for different values of N almost coincide for values of θ_i greater than about 3. This is not too surprising, since expression (6) depends on N only through the factor of $(n-k-1)^{1/2}$ in the numerator and through V^2 in the denominator. The implicit degrees-of-freedom correction resulting from the use of the $t(n-k-1)$ distribution rather than the standard normal appears to be very effective except in the immediate neighborhood of $\boldsymbol{\theta} = \mathbf{0}$.

It may appear from Figure 1 that the J test is not valid asymptotically, since the rejection frequencies appear to be converging, as $N \to \infty$, to a curve that is not a horizontal line at .05. This is an illusion caused by the fact that the elements of $\boldsymbol{\theta}$ are defined to be $O(N^{1/2})$; see (8). If we hold $\boldsymbol{\theta}$ constant as N increases, then we are implicitly making $\boldsymbol{\beta}$ smaller or making σ^2 larger. For $\|\boldsymbol{\theta}\| \neq 0$, increasing N without changing $\boldsymbol{\beta}$ or σ^2 will cause $\boldsymbol{\theta}$ to increase, and the rejection frequency will eventually converge to .05. However, if $\|\boldsymbol{\theta}\| = 0$, this does not happen, and in the baseline case of the figure, the test will reject nearly 80% of the time.

This is illustrated in Figure 2, which plots exactly the same experimental results as Figure 1, but with $N^{-1/2}$ times θ_1 and θ_2 on the horizontal axis. Thus, in this figure, $\boldsymbol{\beta}$ and σ^2 are held constant as N increases. We observe that the rejection frequency converges to .05 for all $N^{-1}\|\boldsymbol{\theta}\| \neq 0$, and to a very much larger value, 0.787, for $N^{-1}\|\boldsymbol{\theta}\| = 0$. Rejection frequencies for $N = \infty$ are not plotted; they simply form an L shape with the top of the L at the point $(0, 0.787)$ and the base extending rightwards from the point $(0, 0.05)$.

The extent to which the J test overrejects depends on the nominal level of the test. Figure 3 shows P value discrepancy plots, in the sense of Davidson and MacKinnon (1998), for the same baseline case as Figures 1 and 2, for $N = 10$ and $N = \infty$, for three different values of θ_1 and θ_2. These are based on 500,000 replications. The nominal level α is plotted on the horizontal axis, and the P value discrepancy, that is, the difference between the true rejection probability and the nominal level, is plotted on the vertical axis. As in Figure 1, the choice of N has only a modest effect on the results, especially when $\|\boldsymbol{\theta}\|$ is large. Since all the plots lie above the horizontal axis, the test universally overrejects. The nominal level for which the overrejection is most severe seems to increase with $\|\boldsymbol{\theta}\|$.

Figure 4 demonstrates that the dependence of the rejection frequencies on the canonical correlations d_i, $i = 1, \ldots, s_5$, is fairly moderate. The figure is based on 250,000 replications with $N = 10$. For the solid curves, $d_1 = d_2$ and both the d_i vary together. For the dotted curves, $d_1 = 0.55$ and only d_2 varies. The dependence on the d_i is strongest for small values of $\|\boldsymbol{\theta}\|$ and becomes very small as $\|\boldsymbol{\theta}\|$ becomes large.

The next two figures demonstrate how the finite-sample performance of the J test depends on s_4 and s_5. Figure 5, which is comparable to Figure 1, shows rejection frequencies as a function of $\theta_1 = \theta_2$ for six different values of s_4, with $s_5 = 2$. These results are based on 250,000 replications with $N = 20$, but other values of N would have given very similar results. We see that the finite-sample performance of the J test deteriorates as s_4 increases, dramatically so for small values of $\|\boldsymbol{\theta}\|$. Figure 6 is similar to Figure 5, except that $s_4 = 4$ and s_5 takes on three different values, 1, 3, and 5. It is clear that the rejection frequencies do depend on s_5, but that this dependence is much weaker than the dependence on s_4. The shape of the curve is somewhat different for $s_5 = 1$ compared with all other values of s_5. Results for $s_5 = 2$ and $s_5 = 4$, which are not shown to avoid cluttering the figure, were similar to those for $s_5 = 3$ and $s_5 = 5$.

The reason for which s_4 has such a strong effect compared with that of s_5, and the nature of that effect, can be seen by considering the two vectors in (10). Recall that J is, except for a degrees-of-freedom factor, the cotangent of the angle between these two vectors. As s_4, which is the dimension of $\boldsymbol{v}_4$, increases, these vectors acquire new components which are the same for each. In most circumstances, this will cause the angle between the two vectors to become smaller, thus making J, and the rejection probability for a given nominal level, larger. This is what we see in Figure 5. In contrast, as s_5, the dimension of $\boldsymbol{\theta}$, $\boldsymbol{v}_5$, and $\boldsymbol{v}_6$, increases, the two vectors gain new but different components. For small values of $\boldsymbol{\theta}$, the new components will be positively correlated, since $\boldsymbol{v}_5$ is positively correlated with $\boldsymbol{v}_6$. Thus we expect that increasing s_5 will increase the rejection probability, as it does in Figure 6. For large $\boldsymbol{\theta}$, the positive correlation will be less, and increasing s_5 may either increase or decrease the rejection probability. Both effects are observed in Figure 6.

If $\boldsymbol{\beta}$ and σ^2 are held fixed as the sample size is increased, the performance of the J test must improve, except when $\|\boldsymbol{\theta}\| = 0$, because $\|\boldsymbol{\theta}\|$ is $O(N^{1/2})$ except in that special case. Figure 7 shows rejection frequencies for a test at the .05 nominal level in the base case as a function of N for several fixed values of $N^{-1/2}\theta_1 = N^{-1/2}\theta_2$.

It is quite clear from Figure 7 that the J test does not have the $N(0,1)$ distribution asymptotically when $\|\boldsymbol{\theta}\| = 0$. From (13), it is easy to see that, as $N \to \infty$, the limiting distribution when $\|\boldsymbol{\theta}\| = 0$ is the distribution of

$$\frac{\boldsymbol{v}_6{}^\top \boldsymbol{v}_5 + \|\boldsymbol{v}_4\|^2}{\left(\|\boldsymbol{v}_4\|^2 + \|\boldsymbol{v}_5\|^2\right)^{1/2}}. \tag{14}$$

Figure 8 shows how greatly this distribution differs from the $N(0,1)$ distribution in several specific cases with all the d_i equal to 0.55. Since the asymptotic distribution is $N(0,1)$ for all $\|\boldsymbol{\theta}\| \neq 0$, there is a singularity in the asymptotic rejection probability function. As we will see in the next section, this causes the bootstrap to fail asymptotically in this special case.

5. Bootstrapping the *J* Test

One way to perform a bootstrap test is to calculate a bootstrap P value and reject the null hypothesis whenever it is less than the level of the test. This is equivalent to rejecting the null whenever the test statistic exceeds a bootstrap critical value. The latter approach is discussed by Horowitz (1994).

Because the error terms are assumed to be normally distributed, we will in this section use a parametric bootstrap. A semiparametric bootstrap procedure will be discussed in the next section. The first step in computing either type of bootstrap P value is to estimate both H_1 and H_2 by OLS and calculate the J statistic, which we denote by $\hat{J}$. Estimation of H_1 also yields unbiased parameter estimates $\hat{\boldsymbol{\beta}}$ and s^2. These estimates provide a bootstrap DGP, or data-generating process, which in the parametric case can be written as

$$\boldsymbol{y}^* = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{u}^*, \quad \boldsymbol{u}^* \sim N(\mathbf{0}, s^2\mathbf{I}), \tag{15}$$

where a star is used to denote random quantities generated by the bootstrap DGP. Then B bootstrap samples are drawn from the DGP (15); ideally, B should be a reasonably large number and should be chosen so that $\alpha(B+1)$ is an integer for all levels α of interest. For each bootstrap sample, a bootstrap test statistic is computed in exactly the same way as $\hat{J}$ was computed from the original data. We denote the bootstrap statistics by J_j^*, $j = 1, \ldots, B$. Finally, the bootstrap P value is computed by the formula

$$\hat{p}^*(\hat{J}) = \frac{1}{B}\sum_{j=1}^{B} I(J_j^* \geq \hat{J}), \tag{16}$$

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

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 true P value depends on the unknown true distribution of $\hat{J}$, while the bootstrap P value (16) is based on the distribution of the bootstrap statistics J_j^*, which depends on the bootstrap DGP (15). These two distributions will differ whenever a test statistic is not pivotal and the parameter estimates used in the bootstrap DGP differ from the true values of the parameters. However, provided the test statistic is asymptotically pivotal, the bootstrap distribution will converge to the true one as the sample size increases. In consequence, as Beran (1988) showed, the bootstrap P value will converge to the true P value at a rate faster than does the asymptotic P value.

Except when $\boldsymbol{\theta} = \mathbf{0}$, the J statistic is asymptotically pivotal. Nevertheless, as is clear from Figures 1 through 7, it is far from being pivotal for small and moderate values of $\|\boldsymbol{\theta}\|$. This suggests that the bootstrap J test might be expected to perform poorly. However, a more detailed analysis leads us to conclude that the bootstrap J test will actually perform remarkably well except when $\|\boldsymbol{\theta}\|$ is very close to 0. The key to this analysis is a result that was proved in Davidson and MacKinnon (1999).

For a given sample size, the rejection probability function, or RPF, $R(\alpha, \boldsymbol{\theta})$ of the J test is defined as the true rejection probability, under the DGP characterized by $\boldsymbol{\theta}$, for the asymptotic test based on J at nominal level α. This is the function graphed in Figures 1, 2, 5, and 6. Similarly, the critical value function, or CVF, $Q(\alpha, \boldsymbol{\theta})$ is defined to be the nominal level of the asymptotic test that yields a true rejection probability of α. By construction, $R(Q(\alpha, \boldsymbol{\theta}), \boldsymbol{\theta}) = \alpha$.

By assumption, the data used to compute the statistic $\hat{J}$ and estimate the parameters of the bootstrap DGP were generated by a DGP characterized by the true parameter vector $\boldsymbol{\theta}_0$, which depends on $\boldsymbol{\beta}_0$ and σ_0^2 only through the ratio $\boldsymbol{\beta}_0/\sigma_0$; recall that the regressors $\boldsymbol{X}$ and $\boldsymbol{Z}$ are treated as exogenous and therefore fixed. The parameter estimates $\hat{\boldsymbol{\theta}}$ which characterize the bootstrap DGP are asymptotically independent of $\hat{J}$, because they were obtained under the null hypothesis. The difference between the true and nominal rejection probabilities of a test will be referred to as the error in rejection probabilities, or ERP. According to the result of Davidson and MacKinnon (1999), whenever a test statistic is asymptotically independent of the bootstrap DGP, the ERP of a bootstrap test is approximately given by

$$\alpha - E_{\boldsymbol{\theta}_0}\Big(R\big(Q(\alpha, \boldsymbol{\theta}_0), \hat{\boldsymbol{\theta}}\big)\Big), \tag{17}$$

where $E_{\boldsymbol{\theta}_0}$ means that we are taking expectations under the DGP characterized by $\boldsymbol{\theta}_0$. The test will overreject or underreject depending on whether the sign of (17) is positive or negative. By studying this expression, we can learn a great deal about the performance of the bootstrap J test. Notice that (17) will equal 0 if the test statistic is pivotal, because, in that special case, the rejection probability for the critical value $Q(\alpha, \boldsymbol{\theta}_0)$ will be the same for all $\hat{\boldsymbol{\theta}}$.

By performing a second-order Taylor expansion of (17), it is easy to obtain an approximate expression for the ERP of the bootstrap J test:

$$\begin{aligned} &- \sum_{i=1}^{s_5} \frac{\partial R}{\partial \theta_i}\big(Q(\alpha, \boldsymbol{\theta}_0), \boldsymbol{\theta}\big)\, E(\hat{\theta}_i - \theta_i^0) \\ -\tfrac{1}{2} &\sum_{i=1}^{s_5}\sum_{j=1}^{s_5} \frac{\partial^2 R}{\partial \theta_i \partial \theta_j}\big(Q(\alpha, \boldsymbol{\theta}_0), \boldsymbol{\theta}\big)\, E\big((\hat{\theta}_i - \theta_i^0)(\hat{\theta}_j - \theta_j^0)\big), \end{aligned} \tag{18}$$

where θ_i^0 is the i^{th} element of $\boldsymbol{\theta}_0$, and the derivatives are evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. If it is supposed that $\boldsymbol{\theta} = O(N^{1/2})$, it can be shown that the two terms in (18) are of order $N^{-3/2}$ and that higher-order terms in the Taylor series are of lower order.

The expectations that appear in expression (18) can be evaluated analytically. It can be shown (see the Appendix) that the expectations of a $\chi^2(m)$ variable raised to the powers $-1/2$ and -1 are

$$\frac{\Gamma\big((m-1)/2\big)}{2^{1/2}\,\Gamma(m/2)} \qquad \text{and} \qquad \frac{1}{m-2}, \tag{19}$$

respectively. Thus the expectation of expression (11) is

$$E(\hat{\boldsymbol{\theta}}) = \left(\frac{N - s_5}{2}\right)^{1/2} \frac{\Gamma\big((N - s_5 - 1)/2\big)}{\Gamma\big((N - s_5)/2\big)}\, \boldsymbol{\theta} \equiv b\boldsymbol{\theta}, \tag{20}$$

and its mean squared error matrix is

$$E\big((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top\big) = \frac{N - s_5}{N - s_5 - 2} \boldsymbol{\Delta}^2 \boldsymbol{D}^2 + \frac{2\big((N - s_5)(1 - b) + (2b - 1)\big)}{N - s_5 - 2}\, \boldsymbol{\theta}\boldsymbol{\theta}^\top. \tag{21}$$

The result (20) shows that the bias of $\hat{\boldsymbol{\theta}}$ is positive, very small, and proportional to $\boldsymbol{\theta}$ itself. For $N - s_5 \geq 14$, the relative error in the estimate of each component of $\boldsymbol{\theta}$ is less than 2%, and for $N - s_5 \geq 27$, it is less than 1%. This implies that the first term in (18) will always be extremely small, even when the RPF is quite steep, as it can be when $\|\boldsymbol{\theta}\|$ is small but not too close to 0 (see Figures 1, 2, 5, and 6). Thus how well the bootstrap test performs will largely depend on the second term in (18), which depends on the second derivatives of the RPF and the second moments of $\hat{\boldsymbol{\theta}}$.

Expression (21) can easily be evaluated for various values of $\boldsymbol{\theta}$, $N - s_5$, and the d_i. It turns out to be fairly small, especially when $\|\boldsymbol{\theta}\|$ is close to 0. In all the cases we examined, the standard errors of the θ_i were between 0 and 1. These standard errors are quite small relative to the scale on which the rejection probabilities vary with the θ_i; see Figures 1, 5, and 6. Thus, even though the concavity of the RPF near $\|\boldsymbol{\theta}\| = 0$ implies that the bootstrap test will overreject in that region, and the convexity of the RPF for larger values of $\|\boldsymbol{\theta}\|$ implies that the bootstrap test will underreject in some other regions, the magnitude of these errors should be small.

It can be seen from (21) that, for fixed N and $\boldsymbol{\theta}$, and for $d_1 = d_2 = d$, the mean squared error of $\hat{\boldsymbol{\theta}}$ is maximized at $d = 2^{-1/2}$. However, it can be seen from Figure 4 that the distortion of the asymptotic test is greatest for small values of d. Monte Carlo experimentation shows that the distortion of the bootstrap rejection probability is greatest for d around 0.55. It is for that reason that we set $d = 0.55$ in the baseline case, in order to observe the worst case for the bootstrap test.

We have performed a large number of Monte Carlo experiments on the performance of the bootstrap J test. For replication j, realizations $\hat{J}_j$ and $\hat{\boldsymbol{\theta}}_j$ are generated using (6) and (11) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, B bootstrap statistics are generated using (6) with $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_j$, and then a bootstrap P value is computed by means of formula (16). Of course, this procedure is available only in the context of a Monte Carlo experiment and could never be used in practice. Each experiment used 500,000 replications for each of 41 different values of N, with $B = 399$. In applied work, it would be wise to use a larger value of B, but experience shows that Monte Carlo experiments with $B = 399$ yield reliable results, because simulation errors in the bootstrap P values tend to offset each other over many replications.

In our experiments, a total of 200 million J statistics were computed for each of 41 sample sizes. This was feasible because (6) makes it possible to compute J statistics for many sample sizes at once. In fact, each experiment for the baseline case took only about three and a quarter hours on a 450 MHz. Pentium II computer running Debian GNU/Linux. The extremely large number of replications ensures that experimental error is very small. It is essential that it be small, because, as we are about to see, the bootstrap tests perform extraordinarily well. There is one disadvantage of this procedure, however: 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.

In Figure 9, rejection probabilities for the bootstrap J test as a function of N are shown for the baseline case ($s_4 = 4$, $s_5 = 2$, $d_1 = d_2 = 0.55$, and $\alpha = 0.05$) for various values of $\boldsymbol{\theta}$. As in Figure 7, the parameters are being held constant. Therefore, $N^{-1/2}\boldsymbol{\theta}$ rather than $\boldsymbol{\theta}$ is held fixed as N is varied. Except when $\theta_i = 0$ for all i, the bootstrap test performs extremely well. For the two largest values of θ_i, which are not all that large, the test performs almost perfectly for all sample sizes. Figure 10 also shows rejection probabilities for the bootstrap J test in the baseline case, but this time $N^{-1/2}\theta_i$ is on the horizontal axis. This figure shows clearly that the test overrejects for very small values of $N^{-1/2}\|\boldsymbol{\theta}\|$ and underrejects for somewhat larger values.

When $\theta_i = 0$ for all i, the test modestly overrejects for all sample sizes, and there is no real indication that this overrejection will go away as N becomes large. Thus, as theory predicts, the bootstrap fails in this case. The extent of this bootstrap failure in the asymptotic limit when $N \to \infty$ can very easily be estimated by Monte Carlo. Expression (14) gives the asymptotic distribution of the J statistic, and the limit of (11) for $N \to \infty$ and $\boldsymbol{\theta} = \boldsymbol{0}$ is $\hat{\boldsymbol{\theta}} = \boldsymbol{\Delta D v}_3$. On the basis of 250,000 replications, with $B = 399$, the limit as $N \to \infty$ of the rejection probability of the bootstrap test at nominal level 0.05 is 0.0565. This value is entirely compatible with what we see from Figures 9 and 10. Thus, although the theoretical prediction of bootstrap failure is borne out, the actual extent of it is very small by conventional standards.

Figure 11 shows results of another set of experiments, this time for a more extreme case, in which $s_4 = 5$ and $s_5 = 4$. Note that N ranges from 15 to 55, rather than from 10 to 50 as it did in Figure 9, because the smallest value of N for which it is possible to calculate the test statistic is 14. When $\theta_i = 0$ for all i, and very small values of the θ_i, the test performs noticeably worse than in the baseline case, but for larger values it once again performs almost perfectly.

The experiments summarized in Figures 9, 10, and 11 are only 12 out of more than 100 similar ones that we have performed. Except when $\|\boldsymbol{\theta}\|$ was small, the bootstrap J test always worked essentially perfectly.

6. Simulation Results for Semiparametric Bootstrap Tests

All of the simulation results presented so far are for models with exogenous regressors and normally distributed error terms. In this section, we relax both of these assumptions. We consider a pair of models of the form

$$H_1: \quad y_t = \boldsymbol{X}_t\boldsymbol{\beta} + \delta_1 y_{t-1} + u_t \tag{22}$$

$$H_2: \quad y_t = \boldsymbol{Z}_t\boldsymbol{\gamma} + \delta_2 y_{t-1} + v_t, \tag{23}$$

where the error terms are IID but not normally distributed. The first elements of $\boldsymbol{X}_t$ and $\boldsymbol{Z}_t$ are unity, and their dimensions are $k-1$ and $l-1$, respectively.

The J statistic is calculated in the usual way, but it is no longer appropriate to use a parametric bootstrap. Instead, we use a semiparametric bootstrap, in which the bootstrap errors are obtained by resampling the rescaled residuals. This means that, in each bootstrap sample, every bootstrap error term u_t^* is equal to each of the n rescaled residuals $(n/(n-k))^{1/2}\,\hat{u}_t$ with probability $1/n$. Then the y_t^* are generated dynamically using the equation

$$y_t^* = \boldsymbol{X}_t\hat{\boldsymbol{\beta}} + \hat{\delta}_1 y_{t-1}^* + u_t^*.$$

Here $\hat{\boldsymbol{\beta}}$ is the vector of OLS estimates of $\boldsymbol{\beta}$ from (22), $\hat{\delta}_1$ is usually the OLS estimate of δ_1, and the actual value y_0 is used for y_0^*. In order to ensure stationarity, values of $\hat{\delta}_1$ greater than 0.99 are replaced by 0.99, and values less than -0.99 are replaced by -0.99. This replacement was very rarely needed.

It is extremely important to use rescaled residuals rather than ordinary residuals. The ordinary residuals tend to be too small, and using them is therefore like using too small a value of σ^2 in the parametric bootstrap. This would effectively cause $\hat{\boldsymbol{\theta}}$ to be biased upwards, and our theory predicts that such a bias would cause the bootstrap test to overreject. That is precisely what we found in a few experiments where we did not rescale the residuals. For example, in the base case (described below) with $n = 20$ and $\sigma = 1$, the rejection frequency for the bootstrap test at the .05 level was 0.0613 with ordinary residuals and 0.0515 with rescaled residuals. Since these two experiments both had 100,000 replications and used the same sequence of random numbers, the estimated difference of .0098 between the two rejection frequencies is very accurate.

In all the experiments, the components of $\boldsymbol{X}_t$, except for the constant term, were distributed as $N(0, 1)$. Each component of $\boldsymbol{Z}_t$ was also normally distributed and correlated with the corresponding component of $\boldsymbol{X}_t$, with correlation ρ. When $k > l$ or $l > k$, any extra components of $\boldsymbol{X}_t$ or $\boldsymbol{Z}_t$ were uncorrelated with everything else. We experimented with the choice of δ_1, σ, ρ, k, and l. We found that the values of δ_1 and ρ had relatively little effect on the performance of the bootstrap J test, and we settled on base-case values of $\delta_1 = 0.8$ and $\rho = 0.5$. The choice of k and l mattered considerably more, and we settled on $k = 5$ and $l = 7$ for the base case. This implies that the H_2 model has 5 regressors that are not in the H_1 model, which will cause both the asymptotic and bootstrap J tests to perform relatively poorly. We used the Student's t distribution with 4 degrees of freedom, rescaled to have variance σ^2, to generate the error terms. Using the normal distribution or the Student's t with a different number of degrees of freedom had very little effect on the results.

These experiments were far more expensive than the ones described in the previous section, because we were unable to make use of the formula (6). In order to reduce computational costs, we used the procedure described in Davidson and MacKinnon (1999) to estimate the performance of the bootstrap test without actually computing the latter. For all sample sizes, we used 500,000 replications to compute an estimated rejection frequency for the bootstrap test. In addition, for sample sizes divisible by 10, we used 100,000 replications and calculated actual rejection frequencies for the bootstrap test based on $B = 399$. Except in a few cases for $n = 10$, the estimated bootstrap rejection frequencies never differed from the actual ones by more than could be explained by experimental error.

Results for the base case with $\sigma = 1$, $\sigma = 4$, and $\sigma = 16$ for $n = 10, 11, \ldots, 70$ are shown in Figure 12. For $\sigma = 1$, the bootstrap test appears to work perfectly for $n > 30$. For the larger values of σ, it overrejects slightly, but its performance improves as n increases, just as the theory predicts. Figure 12 bears a considerable resemblance to Figures 9 and 11. The rejection frequency curves are more jagged because there is no dependence among the experimental results for different values of n, but the basic shape is quite similar. The curve for $\sigma = 16$ is flatter than the curves for small values of θ_i in the earlier figures. This suggests that, in this case, the presence of the lagged dependent variable may actually be causing the bootstrap J test to overreject less severely for very small values of n.

Although Figure 12 strongly suggests that our theoretical results remain useful when the errors are nonnormal and there is a lagged dependent variable, it deals with just three cases. We performed a substantial number of experiments for other cases and could present similar figures for them. However, in order to obtain a better sense of how the performance of the bootstrap J test varies with all the parameters of the model, it is more enlightening to take a different approach, which we now describe.

We performed 18,000 experiments, each with 10,000 replications and $B = 199$, in which the parameters were chosen at random. There were three values of n: 15, 30, and 60, each value being used for 6000 experiments. The $\boldsymbol{X}$ and $\boldsymbol{Z}$ matrices were chosen as described above for each experiment. The elements of $\boldsymbol{\beta}$ were uniform $(0.02, 2.00)$, δ_1 was uniform $(-0.99, 0.99)$, σ^2 was uniform $(0.1, 20)$, and ρ was uniform $(-0.95, 0.95)$. The error terms followed the $t(c)$ distribution, where $c = 4, 5, 6, 7, 8, 9,$ and 10 with equal probability. Similarly, the parameters k and l were equal to 5, 6, 7, 8, and 9 with equal probability.

For each experiment, we recorded the proportion of the time that the asymptotic and bootstrap J tests rejected the null hypothesis at the .05 level. We also recorded the parameter values and the value of $\|\boldsymbol{\theta}\|$, which, because of the lagged dependent variable, was actually an average over the 10,000 replications. In all 18,000 experiments, the highest rejection frequency for the bootstrap test was 0.0836, compared with 0.8815 for the asymptotic test. Of course, since these are maxima over a great many cases, both figures are inflated by experimental error. Thus we did not find a single case in which the bootstrap J test could really be said to perform poorly.

The theory of Section 4 suggests that the bootstrap test will perform well whenever $\|\boldsymbol{\theta}\|$ is reasonably large, but the test will probably overreject somewhat when it is small. Since the relationship between rejection frequencies and $\|\boldsymbol{\theta}\|$ must be nonlinear, we estimated it using Nadaraya-Watson kernel regression.[1] The results of this

[1] We used the program N by Jeff Racine, with an Epanechnikov kernel and the default, fixed, bandwidth.

estimation, for $n = 15$, $n = 30$, and $n = 60$ separately, are shown in Figure 13. These results are striking. For small values of $\|\boldsymbol{\theta}\|$, there is indeed a modest tendency to overreject, but for values greater than about 3, the rejection frequency is very nearly 0.05, as our theory says it should be when $\|\boldsymbol{\theta}\|$ is large.

Of course, we do not claim that the nonparametric regression results shown in Figure 13 tell the whole story. As the theory predicts, rejection frequencies appear to depend on more than just $\|\boldsymbol{\theta}\|$, at least when $\|\boldsymbol{\theta}\|$ is not large. The conditional variances from the nonparametric regressions are therefore larger for small values of $\|\boldsymbol{\theta}\|$ than for larger values, presumably because other features of the model, such as k, l, and the parameter values, matter as well.

We also performed a substantial number of additional experiments, similar to those just described but with error terms generated from the $\chi^2(c)$ distribution, recentered to have mean 0 and rescaled to have variance σ^2, where c was chosen randomly in the same way as before. The object of these experiments was to see whether skewness of the error terms would affect the results. It did not: The results of these experiments were very similar to the ones shown in Figure 13.

The results of this and the preceding section suggest that, in the rare event that the bootstrap J test yields a P value near the level of the test, it may be useful to compute $\|\hat{\boldsymbol{\theta}}\|$, the sample analog of $\|\boldsymbol{\theta}\|$, in which $\hat{\boldsymbol{\beta}}$ and s^2 replace $\boldsymbol{\beta}$ and σ^2 in expression (8). If $\|\hat{\boldsymbol{\theta}}\|$ is large, it is almost certainly safe to accept the bootstrap P value. However, if it is small and the P value is just a little smaller than the level of the test, it may not be safe to do so. Of course, when $\|\hat{\boldsymbol{\theta}}\|$ is small, the J test may be of only modest interest, because there may not be enough information in the sample to discriminate between the two models.

The suggestion made in the previous paragraph appears to ignore the fact that $\|\hat{\boldsymbol{\theta}}\|$ is a random quantity. Indeed, expression (11) suggests that $\|\hat{\boldsymbol{\theta}}\|$ will often be quite variable. In order to see whether this randomness would invalidate this suggestion, we generated a further three million J statistics and their bootstrap P values, using the same scheme as was used to generate the data that underlie Figure 13. There were one million J statistics for each of $n = 15$, 30, and 60. Overall, 5.70% of the bootstrap tests rejected at the .05 level when $n = 15$, 5.39% did so when $n = 30$, and 5.17% did so when $n = 60$. When we looked at the rejection frequencies conditional on $\|\hat{\boldsymbol{\theta}}\| > 3$, the rejection percentages fell to 4.96%, 5.02%, and 5.03%, respectively. Despite the often considerable randomness of $\|\hat{\boldsymbol{\theta}}\|$, conditioning on $\|\boldsymbol{\theta}\| > 3$ instead of $\|\hat{\boldsymbol{\theta}}\| > 3$ produced almost identical results.

7. Effects of Bootstrapping on Power

It is natural to worry that bootstrapping the J test, or indeed any test, may adversely affect its power. However, such worries are generally unfounded. When a test rejects much less frequently under the null hypothesis after it is bootstrapped, as is often the case for the J test, it will generally also reject less frequently under every alternative hypothesis. But this does not mean that the bootstrap test is less powerful in any meaningful sense. In order to compare power, we must hold test size constant. When this is done, Davidson and MacKinnon (1997) showed that there is no power loss at all when the underlying test statistic is pivotal (assuming that B is very large), and any difference between the size-corrected power of the two tests will be of low order in the sample size when the underlying test statistic is not pivotal.

An easy way to compute and display size-corrected power was suggested by Davidson and MacKinnon (1998). The idea is to perform two matched experiments, for one of which the alternative is true, and for one of which the null is true. We then graph the rejection frequencies under the alternative, on the vertical axis, against the rejection frequencies under the null, on the horizontal axis, for a large number of possible critical values. This yields a "size-power curve" which would ideally have the shape of a Γ. The further the curve is above the $45°$ line, the more powerful is the test.

In the case of a nonpivotal statistic, the results of this procedure will depend on precisely what parameter values are used to generate the data under the null hypothesis. In Davidson and MacKinnon (1997), we suggested using the parameter values associated with the "pseudo-true null," that is, the parameter values which make the null hypothesis as close as possible, in the sense of the Kullback-Leibler Information Criterion, to the alternative against which power is being computed.

We have performed a few Monte Carlo experiments to verify that the general results on the power of bootstrap tests alluded to above do indeed apply to the J test. We used models similar to those of the last section, but with no lagged dependent variables, since their presence would have made it harder to calculate the pseudo-true nulls. For the alternative, we generated the data from the H_2 model, equation (23), with $\delta_2 = 0$ and all of the γ_i equal to 1. For the null, we generated the data from the H_1 model, equation (22), with $\delta_1 = 0$ and the β_i chosen so that, on average, the H_1 model fits the H_2 data as well as possible.

We performed four experiments, each with 100,000 replications, $k = l = 4$, $\sigma = 2$ or $\sigma = 4$, and $n = 10$ or $n = 40$. For the bootstrap tests, we used $B = 999$ in order to minimize the power loss associated with using too small a number of bootstrap samples. Results are shown in Figure 14. It is clear that, as the theory predicts, bootstrapping the J test has very little effect on size-corrected power. If anything, the bootstrap tests seem to be slightly more powerful than the asymptotic tests, although this may be an artifact of the experimental design.

8. Conclusion

Most Monte Carlo experiments on the performance of hypothesis tests are not very conclusive. They often suffer from excessive experimental error, and they inevitably deal with only a tiny subset of all the possible DGPs. In contrast, except for the thousands of experiments with random parameters discussed in Section 6, which allow us to deal with a very large number of DGPs, our experiments utilize very large numbers of replications. In the case of the experiments of Section 5, our theoretical results made this feasible. In the case of the experiments of Section 6, we were able to use a previous theoretical result to avoid actually computing bootstrap tests for most sample sizes.

The principal reason that our results are quite conclusive is that they are based on a detailed theory of the finite-sample distribution of the J test. This theory shows that the value of a parameter that we call $\|\boldsymbol{\theta}\|$ is crucial. Based on this theory, we were able to identify cases in which the bootstrap J test can be expected to work particularly badly, and we made these the focus of our experiments. That the test nevertheless works extraordinarily well, albeit somewhat less well in extreme cases where $\|\boldsymbol{\theta}\|$ is very small, provides very strong evidence that the bootstrap J test is a reliable procedure in general.

Although our theoretical results were developed for the case in which the error terms are normally distributed and the regressors are exogenous, there are, as we explained in Section 3, good reasons to believe that they apply more generally. In Section 6, we provided a great deal of simulation-based evidence that they in fact do so. The theoretical results also did not deal with the case in which the null hypothesis is false, but general results on bootstrap tests, which are confirmed by the simulations of Section 7, suggest that bootstrapping the J test will have little effect on its size-corrected power.

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, 1997, The power of bootstrap tests, Queen's Institute for Economic Research Discussion Paper 937, revised.

Davidson, R. and J. G. MacKinnon, 1998, Graphical methods for investigating the size and power of hypothesis tests, The Manchester School 66, 1–26.

Davidson, R. and J. G. MacKinnon, 1999, The size distortion of bootstrap tests, Econometric Theory 15, 361–376.

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., 1998, Tests of non-nested regression models: Some results on small sample behaviour and the bootstrap, Journal of Econometrics 84, 59–74.

Godfrey, L. G., 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 (Springer-Verlag, New York).

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., 1999, The distributions of the J and Cox non-nested tests in regression models with weakly correlated regressors, Journal of Econometrics 93, 369–401.

Pesaran, M. H., 1974, On the general problem of model selection, Review of Economic Studies 41, 153–171.

Appendix

Proof of Theorem 1:

We introduce arbitrary orthonormal bases of the spaces S_3, S_5, and S_6, all of which are of dimension s_5. We can represent these bases by three $n \times s_5$ matrices, $\boldsymbol{Z}_3$, $\boldsymbol{Z}_5$, and $\boldsymbol{Z}_6$, the columns of which span their respective spaces. For $i = 3, 5, 6$, orthonormality implies that the matrices satisfy $\boldsymbol{Z}_i^\top \boldsymbol{Z}_i = \mathbf{I}$ and $\boldsymbol{Z}_i \boldsymbol{Z}_i^\top = \boldsymbol{P}_i$.

Since $s_3 = s_5$, $\boldsymbol{Z}_3^\top \boldsymbol{Z}_5$ is a square matrix. We can express its singular value decomposition as follows:

$$\boldsymbol{Z}_3^\top \boldsymbol{Z}_5 = \boldsymbol{U}_3 \boldsymbol{D} \boldsymbol{U}_5^\top, \tag{24}$$

where $\boldsymbol{U}_3$ and $\boldsymbol{U}_5$ are $s_5 \times s_5$ orthogonal matrices, and $\boldsymbol{D}$ is an $s_5 \times s_5$ diagonal matrix. We denote the diagonal elements of $\boldsymbol{D}$ by d_i, $i = 1, \ldots, s_5$. The d_i^2 are the eigenvalues of the positive definite matrix

$$(\boldsymbol{Z}_3^\top \boldsymbol{Z}_5)^\top (\boldsymbol{Z}_3^\top \boldsymbol{Z}_5) = \boldsymbol{Z}_5^\top \boldsymbol{P}_3 \boldsymbol{Z}_5,$$

and so the d_i are just the canonical correlations between $\boldsymbol{Z}_3$ and $\boldsymbol{Z}_5$. As such, they are independent of the particular choice of orthonormal basis. Define new orthonormal bases in terms of the matrices $\boldsymbol{W}_i = \boldsymbol{Z}_i \boldsymbol{U}_i$ for $i = 3, 5$. Then $\boldsymbol{W}_i^\top \boldsymbol{W}_i = \mathbf{I}$, $\boldsymbol{W}_i \boldsymbol{W}_i^\top = \boldsymbol{P}_i$, and

$$\boldsymbol{W}_3^\top \boldsymbol{W}_5 = \boldsymbol{D}. \tag{25}$$

Similarly to (24), we can write

$$\boldsymbol{Z}_6^\top \boldsymbol{W}_5 = \boldsymbol{U}_6 \boldsymbol{\Delta} \boldsymbol{V}_5^\top, \tag{26}$$

where $\boldsymbol{U}_6$ and $\boldsymbol{V}_5$ are orthogonal matrices, and $\boldsymbol{\Delta}$ is diagonal with diagonal elements δ_i, $i = 1, \ldots, s_5$, with the δ_i^2 the eigenvalues of $\boldsymbol{W}_5^\top \boldsymbol{P}_6 \boldsymbol{W}_5$. Let $\boldsymbol{W}_6 = \boldsymbol{Z}_6 \boldsymbol{U}_6$. Then

$$\boldsymbol{W}_6^\top \boldsymbol{W}_5 = \boldsymbol{\Delta} \boldsymbol{V}_5^\top. \tag{27}$$

Note that every vector in S_5 is in $S_3 \oplus S_6$, since S_3 and S_6 are the only spaces in S_0 to which S_5 is not orthogonal. Thus $(\boldsymbol{P}_3 + \boldsymbol{P}_6)\boldsymbol{W}_5 = \boldsymbol{W}_5$, and so

$$\begin{aligned}\mathbf{I} = \boldsymbol{W}_5^\top \boldsymbol{W}_5 &= \boldsymbol{W}_5^\top \boldsymbol{P}_3 \boldsymbol{W}_5 + \boldsymbol{W}_5^\top \boldsymbol{P}_6 \boldsymbol{W}_5 = \boldsymbol{W}_5^\top \boldsymbol{W}_3 \boldsymbol{W}_3^\top \boldsymbol{W}_5 + \boldsymbol{W}_5^\top \boldsymbol{W}_6 \boldsymbol{W}_6^\top \boldsymbol{W}_5 \\ &= \boldsymbol{D}^2 + \boldsymbol{V}_5 \boldsymbol{\Delta}^2 \boldsymbol{V}_5^\top, \end{aligned} \tag{28}$$

by (25) and (27). Since $\mathbf{I}$, $\boldsymbol{D}$, and $\boldsymbol{\Delta}$ are all diagonal matrices, the orthogonal matrix $\boldsymbol{V}_5$ must be a permutation matrix that alters only the ordering of the elements of $\boldsymbol{\Delta}$. Thus, reordering the rows and columns of $\boldsymbol{U}_6$ if necessary, we can choose $\boldsymbol{V}_5 = \mathbf{I}$. It follows that $\boldsymbol{\Delta}^2 = \mathbf{I} - \boldsymbol{D}^2$, or, equivalently, $\delta_i^2 = 1 - d_i^2$ for each i.

Under the assumptions of model H_1, $\boldsymbol{u} \sim N(0, \sigma^2 \mathbf{I})$. Because of the multivariate normality of $\boldsymbol{u}$, the projections $\boldsymbol{P}_i \boldsymbol{u}$, $i = 1, 2, 3, 4, 6$, and $\boldsymbol{M}_0 \boldsymbol{u}$ are all mutually independent. In order to express the J statistic in terms of these quantities, it is convenient to make the following definitions of mutually independent standard normal variables:

$$\begin{aligned} \boldsymbol{v}_3 &= \sigma^{-1} \boldsymbol{W}_3^\top \boldsymbol{u} \sim N(\mathbf{0}, \mathbf{I}_{s_5}) \\ \boldsymbol{v}_4 &= \sigma^{-1} \boldsymbol{W}_4^\top \boldsymbol{u} \sim N(\mathbf{0}, \mathbf{I}_{s_4}) \\ \boldsymbol{v}_6 &= \sigma^{-1} \boldsymbol{W}_6^\top \boldsymbol{u} \sim N(\mathbf{0}, \mathbf{I}_{s_5}), \end{aligned} \tag{29}$$

and a χ^2 variable independent of $\boldsymbol{v}_3$, $\boldsymbol{v}_4$, and $\boldsymbol{v}_6$:

$$V^2 = \sigma^{-2} \boldsymbol{u}^\top \boldsymbol{M}_0 \boldsymbol{u} \sim \chi^2(n - s_0).$$

From (4) and (5) it can be seen that

$$\boldsymbol{M}_{\boldsymbol{X}} = \boldsymbol{M}_0 + \boldsymbol{P}_4 + \boldsymbol{P}_6. \tag{30}$$

Thus

$$\boldsymbol{M}_{\boldsymbol{X}} \boldsymbol{u} = \boldsymbol{M}_0 \boldsymbol{u} + \boldsymbol{P}_4 \boldsymbol{u} + \boldsymbol{P}_6 \boldsymbol{u}, \quad \text{and} \quad \boldsymbol{u}^\top \boldsymbol{M}_{\boldsymbol{X}} \boldsymbol{u} = \sigma^2 \big(V^2 + \|\boldsymbol{v}_4\|^2 + \|\boldsymbol{v}_6\|^2\big). \tag{31}$$

Further,

$$\boldsymbol{M}_{\boldsymbol{X}} \boldsymbol{P}_{\boldsymbol{Z}} = (\boldsymbol{M}_0 + \boldsymbol{P}_4 + \boldsymbol{P}_6)(\boldsymbol{P}_1 + \boldsymbol{P}_4 + \boldsymbol{P}_5) = \boldsymbol{P}_4 + \boldsymbol{P}_6 \boldsymbol{P}_5. \tag{32}$$

Now $\boldsymbol{M_X P_Z y} = \boldsymbol{M_X P_Z X\beta} + \boldsymbol{M_X P_Z u}$. Denote the deterministic vector $\boldsymbol{M_X P_Z X\beta}$ by $\boldsymbol{r}$. Since $\boldsymbol{P}_4\boldsymbol{X} = \boldsymbol{0}$, we have that

$$\boldsymbol{r} = \boldsymbol{P}_6\boldsymbol{P}_5\boldsymbol{X\beta}. \tag{33}$$

For the other term in $\boldsymbol{M_X P_Z y}$, we have that $\boldsymbol{M_X P_Z u} = \boldsymbol{P}_4\boldsymbol{u} + \boldsymbol{P}_6\boldsymbol{P}_5\boldsymbol{u}$. Using the results that $\boldsymbol{P}_5 = \boldsymbol{P}_5(\boldsymbol{P}_3 + \boldsymbol{P}_6)$ and $\boldsymbol{V}_5 = \mathbf{I}$, we have from (25), (27), and (29) that

$$\begin{aligned}\boldsymbol{P}_6\boldsymbol{P}_5\boldsymbol{u} &= \boldsymbol{W}_6\boldsymbol{W}_6^\top\boldsymbol{W}_5\boldsymbol{W}_5^\top(\boldsymbol{W}_3\boldsymbol{W}_3^\top\boldsymbol{u} + \boldsymbol{W}_6\boldsymbol{W}_6^\top\boldsymbol{u})\\ &= \boldsymbol{W}_6\boldsymbol{\Delta}(\boldsymbol{D}\boldsymbol{W}_3^\top\boldsymbol{u} + \boldsymbol{\Delta}\boldsymbol{W}_6^\top\boldsymbol{u}) = \sigma\big(\boldsymbol{W}_6\boldsymbol{\Delta}(\boldsymbol{D}\boldsymbol{v}_3 + \boldsymbol{\Delta}\boldsymbol{v}_6)\big).\end{aligned}$$

We define the vector $\boldsymbol{v}_5$ by (7):

$$\boldsymbol{v}_5 = \boldsymbol{\Delta}(\boldsymbol{D}\boldsymbol{v}_3 + \boldsymbol{\Delta}\boldsymbol{v}_6),$$

so that $\boldsymbol{P}_6\boldsymbol{P}_5\boldsymbol{u} = \sigma\boldsymbol{W}_6\boldsymbol{v}_5$. Then $\boldsymbol{M_X P_Z y} = \boldsymbol{r} + \boldsymbol{P}_4\boldsymbol{u} + \sigma\boldsymbol{W}_6\boldsymbol{v}_5$.

We can now obtain suitable expressions for the other two scalar products on which J depends. Let the vector $\boldsymbol{\theta}$ be defined by

$$\boldsymbol{\theta} = \sigma^{-1}\boldsymbol{W}_6^\top\boldsymbol{r}. \tag{34}$$

Then, since $\boldsymbol{r} = \boldsymbol{P}_6\boldsymbol{r} = \sigma\boldsymbol{W}_6\boldsymbol{\theta}$,

$$\boldsymbol{y}^\top\boldsymbol{M_X P_Z y} = \boldsymbol{u}^\top\big(\boldsymbol{P}_4\boldsymbol{u} + \sigma\boldsymbol{W}_6(\boldsymbol{v}_5 + \boldsymbol{\theta})\big) = \sigma^2\big(\|\boldsymbol{v}_4\|^2 + \boldsymbol{v}_6^\top(\boldsymbol{v}_5 + \boldsymbol{\theta})\big). \tag{35}$$

Note that $\boldsymbol{r}^\top\boldsymbol{r} = \sigma^2\|\boldsymbol{\theta}\|^2$. From (32) and (33), we have

$$\boldsymbol{r}^\top\boldsymbol{r} = \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{P}_5\boldsymbol{P}_6\boldsymbol{P}_5\boldsymbol{X\beta} = \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{P_Z M_X P_Z X\beta},$$

from which (8) follows at once. Similarly,

$$\begin{aligned}\boldsymbol{y}^\top\boldsymbol{P_Z M_X P_Z y} &= \big(\boldsymbol{u}^\top\boldsymbol{P}_4 + \sigma(\boldsymbol{v}_5^\top + \boldsymbol{\theta}^\top)\boldsymbol{W}_6^\top\big)\big(\boldsymbol{P}_4\boldsymbol{u} + \sigma\boldsymbol{W}_6(\boldsymbol{v}_5 + \boldsymbol{\theta})\big)\\ &= \sigma^2\big(\|\boldsymbol{v}_4\|^2 + \|\boldsymbol{\theta}\|^2 + 2\boldsymbol{\theta}^\top\boldsymbol{v}_5 + \|\boldsymbol{v}_5\|^2\big).\end{aligned} \tag{36}$$

Finally, then, we can substitute (31), (35), and (36) in (3) to obtain (6).

Proof of (11):

Since $\boldsymbol{\theta}$ is as given in (34), with $\boldsymbol{r}$ given by (33), the estimate $\hat{\boldsymbol{\theta}}$ is obtained by replacing σ in (34) and $\boldsymbol{\beta}$ in (33) by estimates obtained by OLS estimation of H_1. We have

$$\hat\sigma^2 = \frac{\boldsymbol{y}^\top\boldsymbol{M_X y}}{n-k} = \frac{\boldsymbol{u}^\top\boldsymbol{M_X u}}{N-s_5} = \frac{\sigma^2(V^2 + \|\boldsymbol{v}_4\|^2 + \|\boldsymbol{v}_6\|^2)}{N-s_5}$$

where the last equality follows from (31). Further,

$$\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X\beta} + \boldsymbol{P_X u} = \boldsymbol{X\beta} + (\boldsymbol{P}_1 + \boldsymbol{P}_2 + \boldsymbol{P}_3)\boldsymbol{u}.$$

Thus, from (34),

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \hat{\sigma}^{-1}\boldsymbol{W}_6^\top \boldsymbol{P}_5\big(\boldsymbol{X\beta} + (\boldsymbol{P}_1 + \boldsymbol{P}_2 + \boldsymbol{P}_3)\boldsymbol{u}\big) \\ &= \left(\frac{V^2 + \|\boldsymbol{v}_4\|^2 + \|\boldsymbol{v}_6\|^2}{N - s_5}\right)^{-1/2} (\boldsymbol{\theta} + \sigma^{-1}\boldsymbol{W}_6^\top \boldsymbol{P}_5\boldsymbol{P}_3\boldsymbol{u}). \end{aligned} \tag{37}$$

Now

$$\begin{aligned}\sigma^{-1}\boldsymbol{W}_6^\top \boldsymbol{P}_5\boldsymbol{P}_3\boldsymbol{u} &= \sigma^{-1}\boldsymbol{W}_6^\top \boldsymbol{W}_5\boldsymbol{W}_5^\top \boldsymbol{W}_3\boldsymbol{W}_3^\top \boldsymbol{u} \\ &= \sigma^{-1}\boldsymbol{\Delta D W}_3^\top \boldsymbol{u} \\ &= \boldsymbol{\Delta D v}_3, \end{aligned} \tag{38}$$

where we used (25), (27), and (29). Substituting (38) into (37) gives (11).

Proof of (19):

The density of the $\chi^2(m)$ distribution is

$$f_m(x) = \frac{1}{2^{m/2}\Gamma(m/2)}\, x^{m/2-1}e^{-x/2}.$$

The expectation of 1 over the square root of a $\chi^2(m)$ variable is therefore

$$\int_0^\infty x^{-1/2}\, f_m(x)\, dx = \frac{1}{2^{m/2}\Gamma(m/2)} \int_0^\infty x^{-1/2}x^{m/2-1}e^{-x/2}\, dx. \tag{39}$$

The definition of the Gamma function is

$$\Gamma(z) = \int_0^\infty y^{z-1}e^{-y}\, dy,$$

and so the expectation (39) becomes

$$\frac{2^{(m-1)/2}}{2^{m/2}\Gamma(m/2)}\Gamma\big((m-1)/2\big) = \frac{\Gamma\big((m-1)/2\big)}{2^{1/2}\Gamma(m/2)}.$$

The expectation of the reciprocal of a $\chi^2(m)$ variable is

$$\int_0^\infty x^{-1}\, f_m(x)\, dx = \frac{1}{2^{m/2}\Gamma(m/2)} \int_0^\infty x^{m/2-2}e^{-x/2}\, dx = \frac{\Gamma\big((m-2)/2\big)}{2\Gamma(m/2)}. \tag{40}$$

Since for any argument z, $\Gamma(z+1) = z\Gamma(z)$, the expectation (40) reduces to $1/(m-2)$.

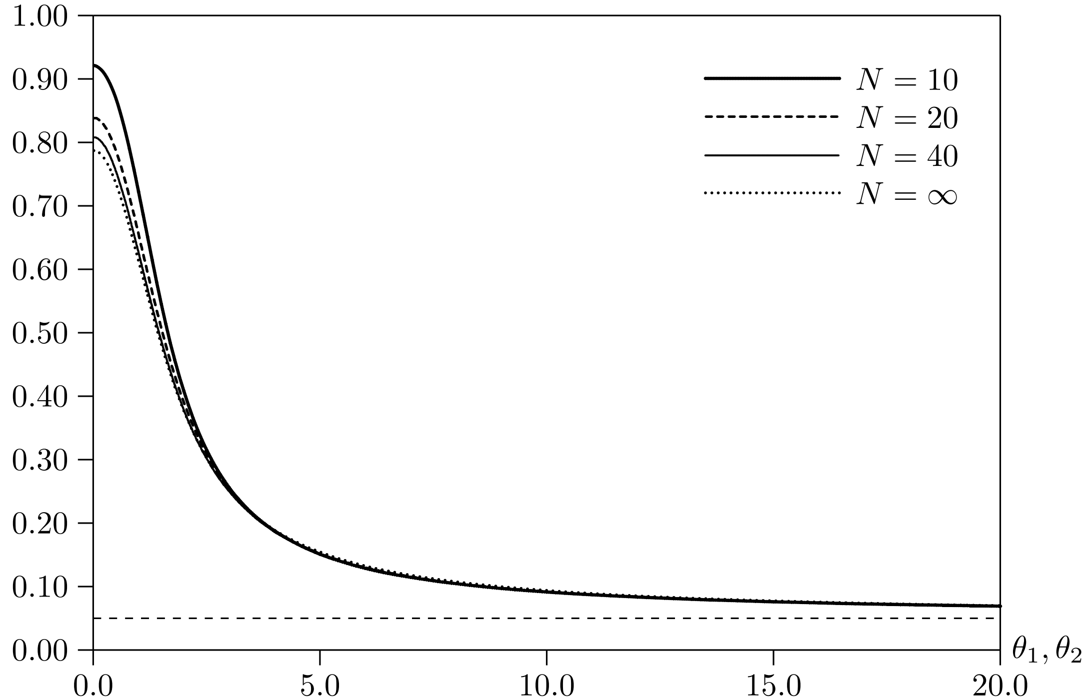

Figure 1. Rejection Frequencies at Nominal Level .05, $\theta_1 = \theta_2$

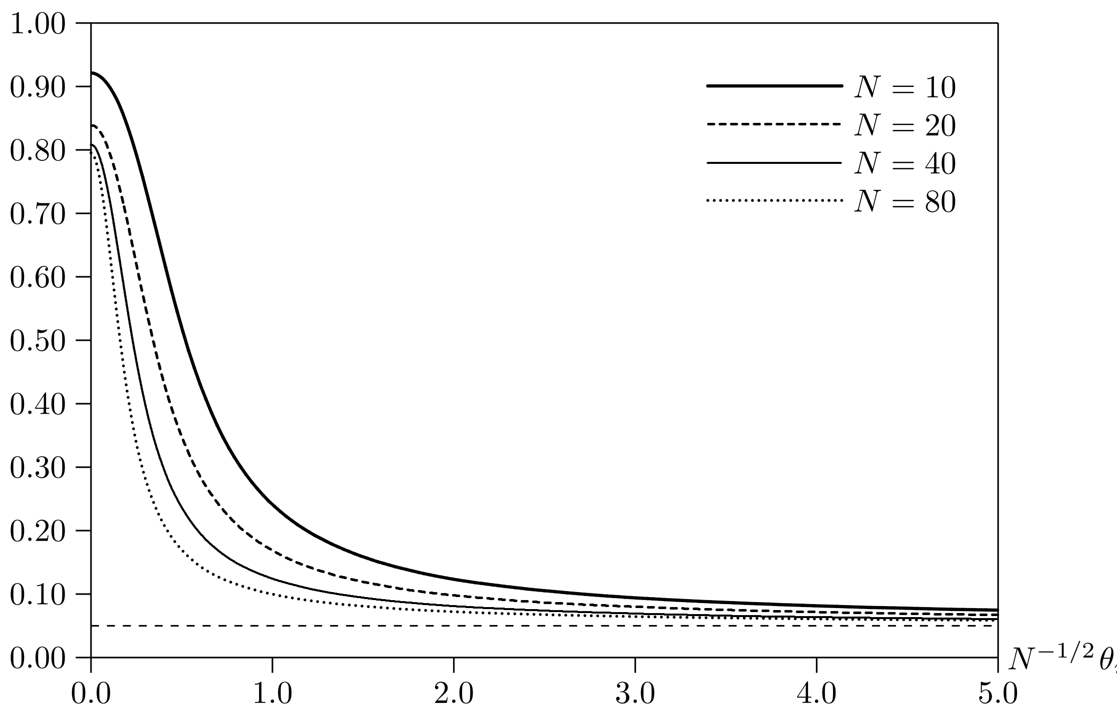

Figure 2. Rejection Frequencies at Nominal Level .05, $\theta_1 = \theta_2$

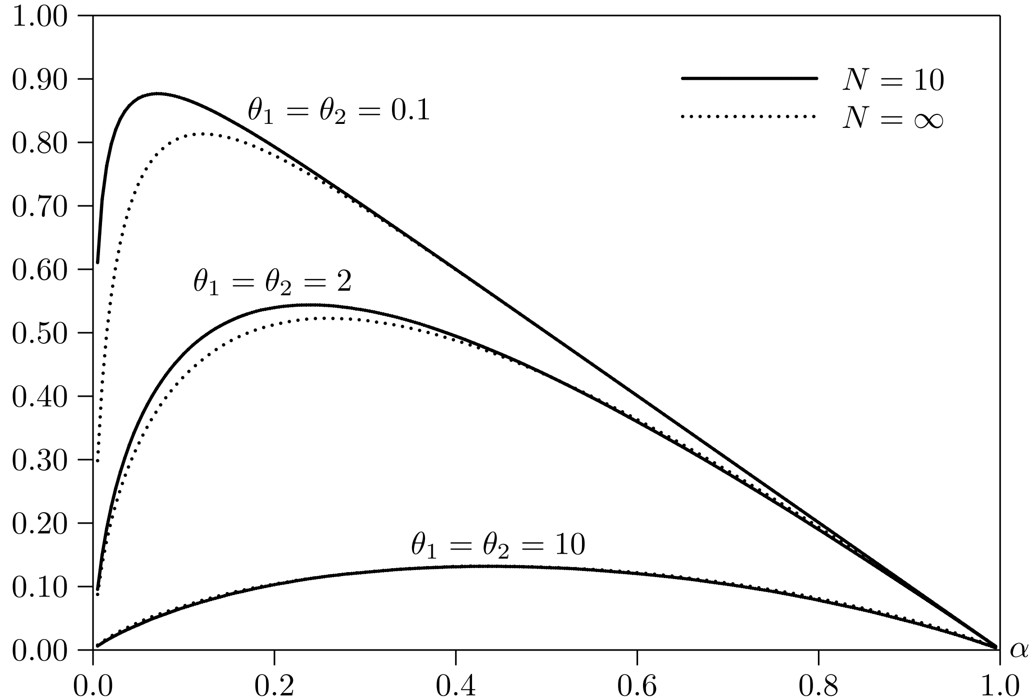

Figure 3. *P* Value Discrepancy Plots

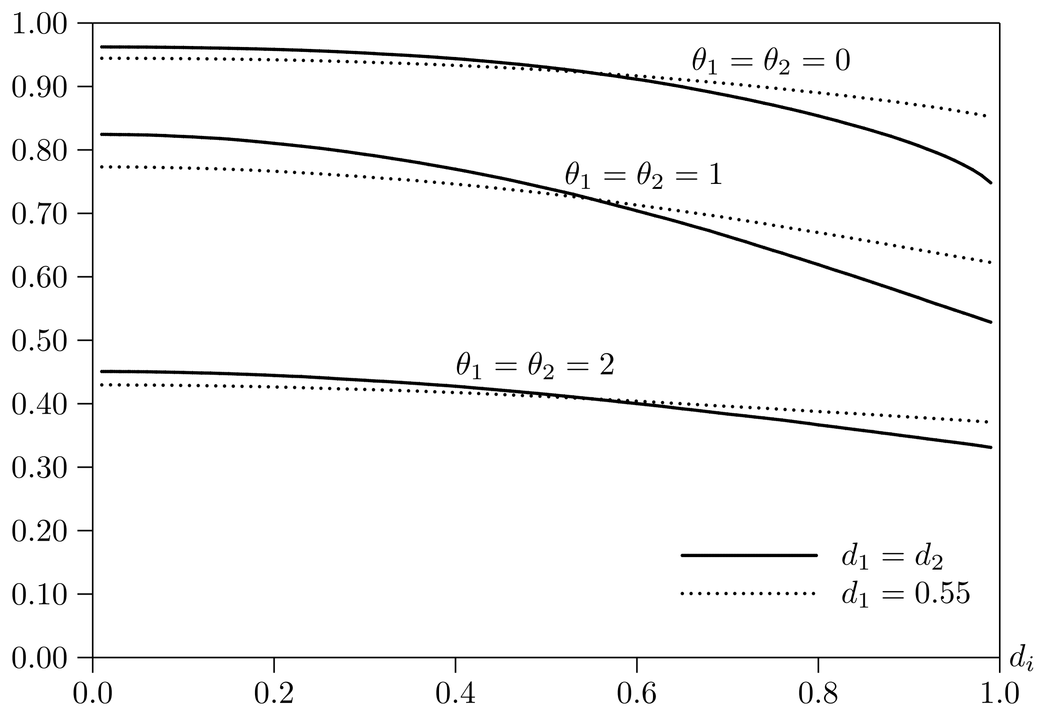

Figure 4. Dependence of Rejection Frequencies on Canonical Correlations

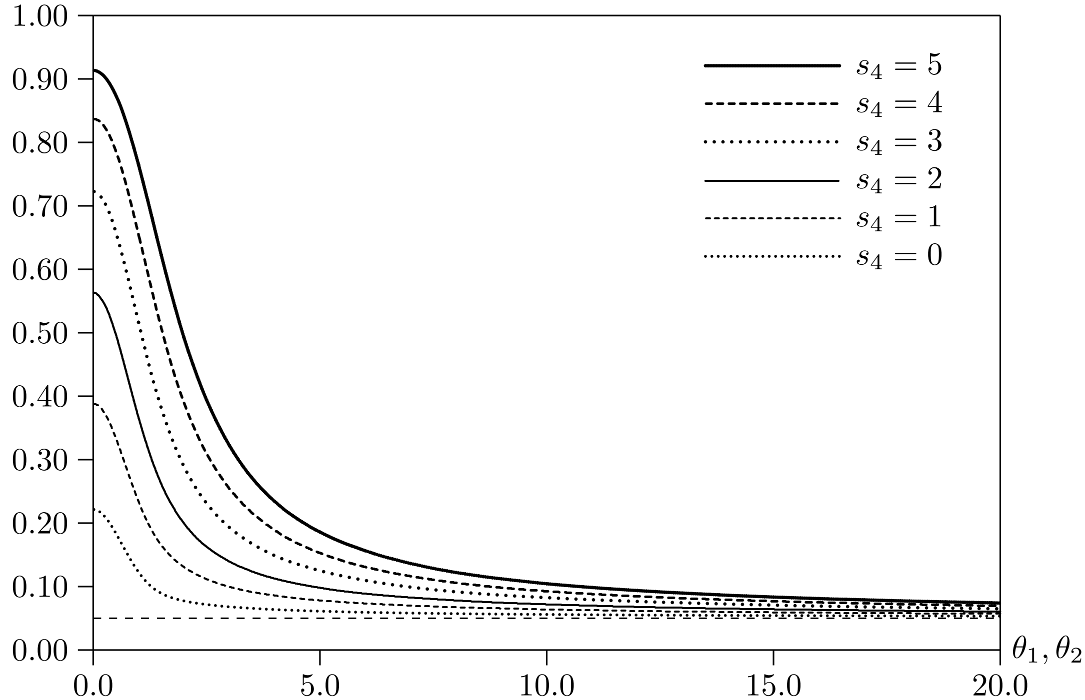

Figure 5. Rejection Frequencies at Nominal Level .05, $\theta_1 = \theta_2$, $N = 20$

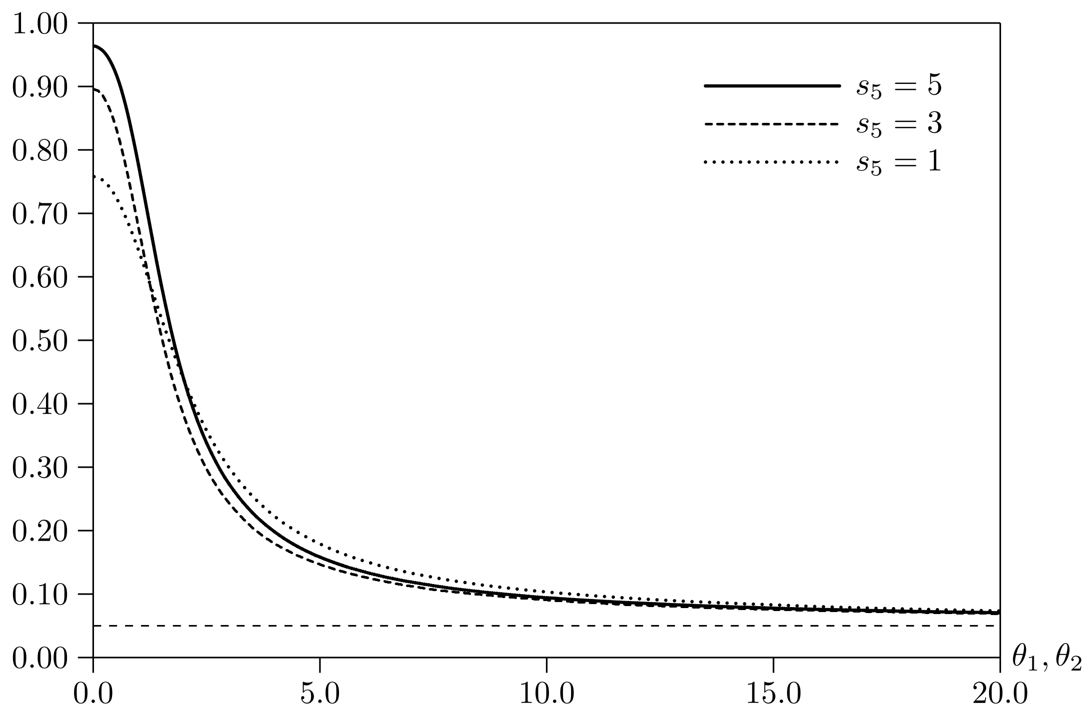

Figure 6. Rejection Frequencies at Nominal Level .05, $\theta_1 = \theta_2$, $N = 20$

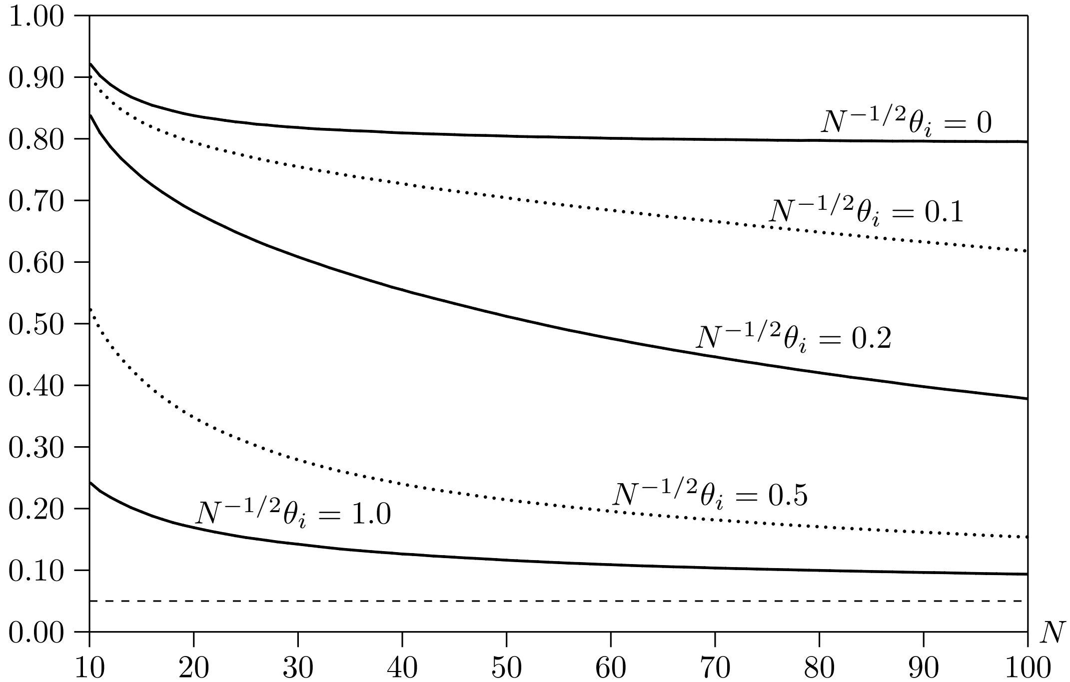

Figure 7. Rejection Frequencies as a Function of *N*

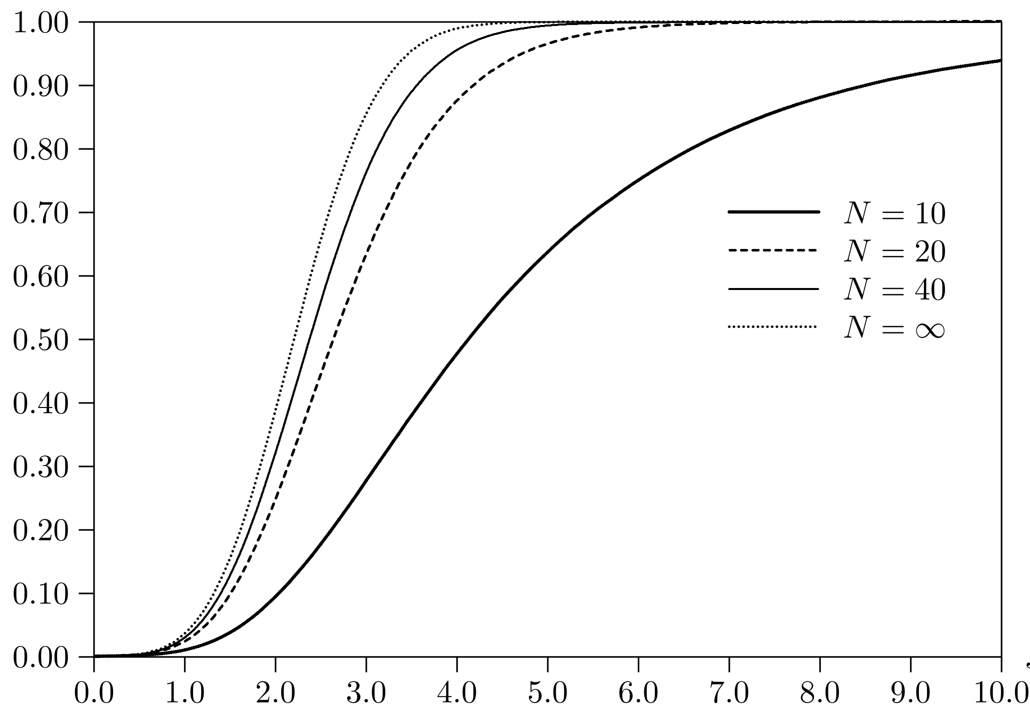

Figure 8. CDF of J for $\theta = 0$

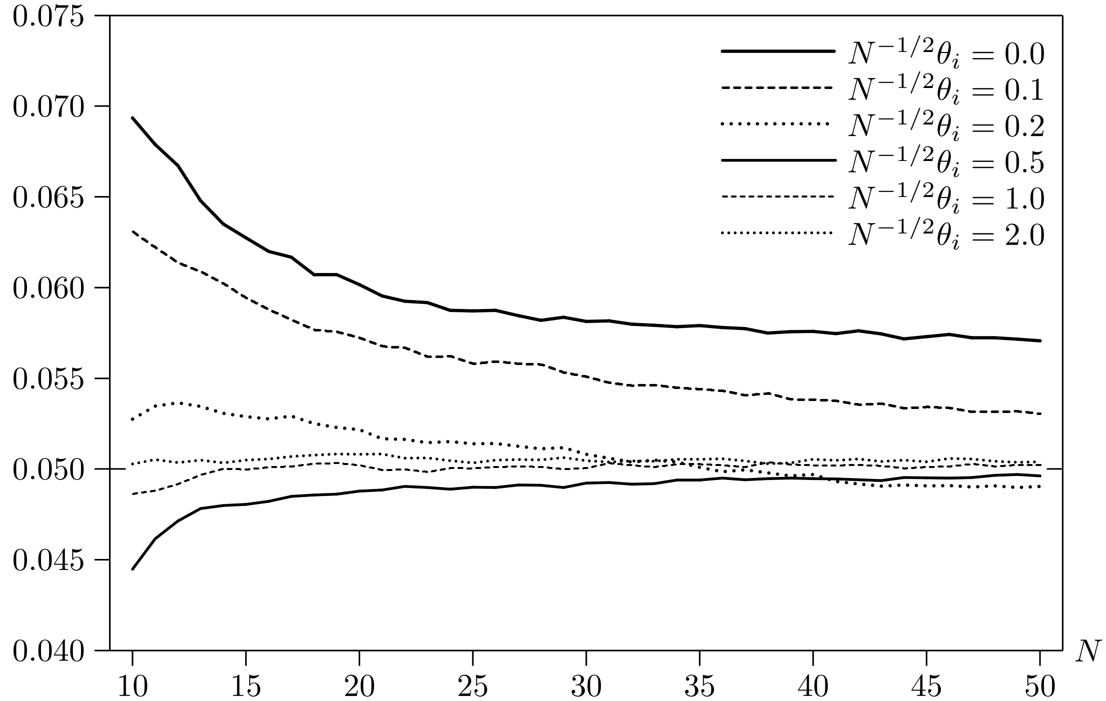

Figure 9. Rejection Frequencies for Bootstrap Tests: Base Case

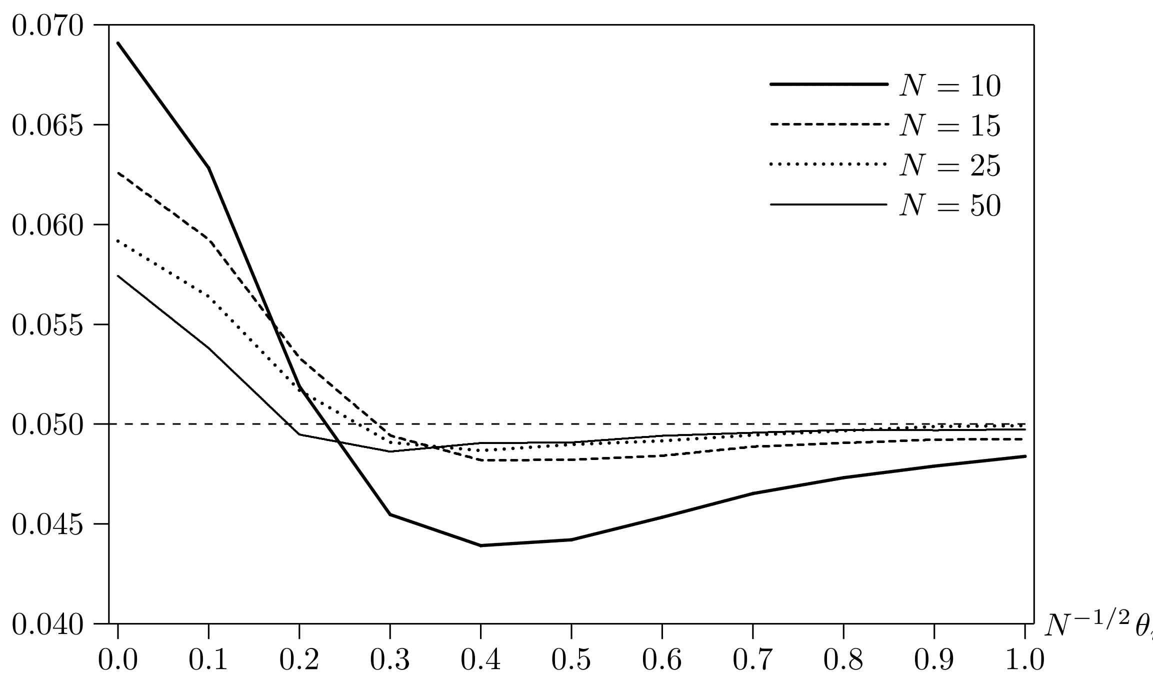

Figure 10. Rejection Frequencies for Bootstrap Tests, Base Case

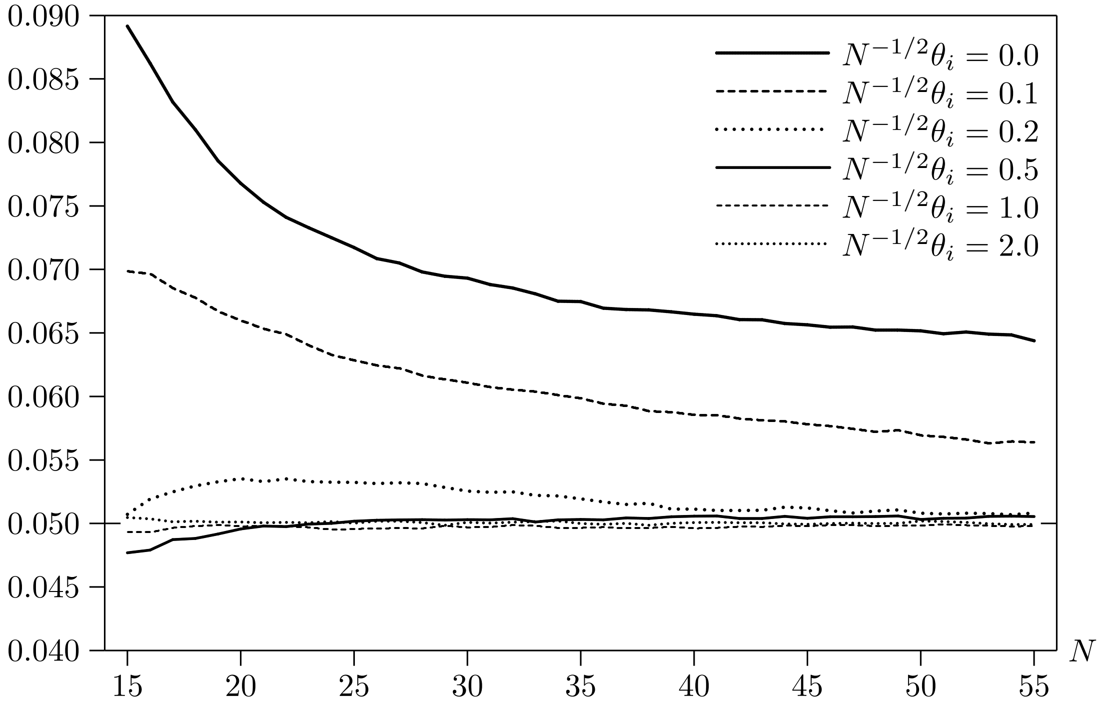

Figure 11. Rejection Frequencies for Bootstrap Tests: Extreme Case

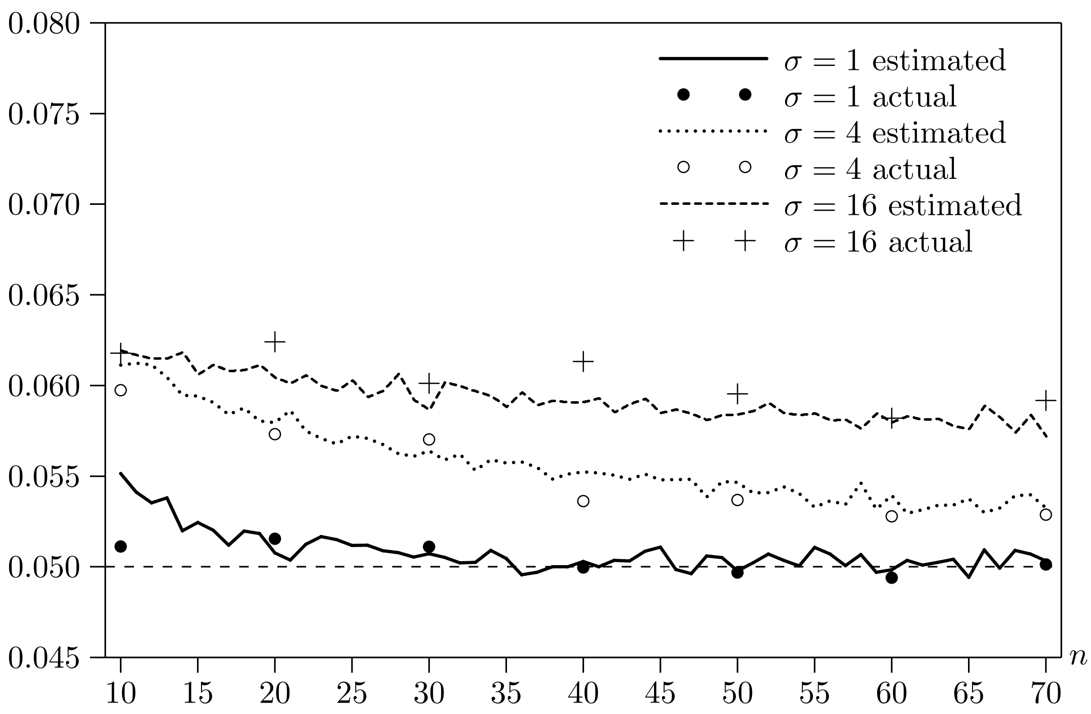

Figure 12. Estimated Rejection Frequencies for Semiparametric Bootstrap Tests

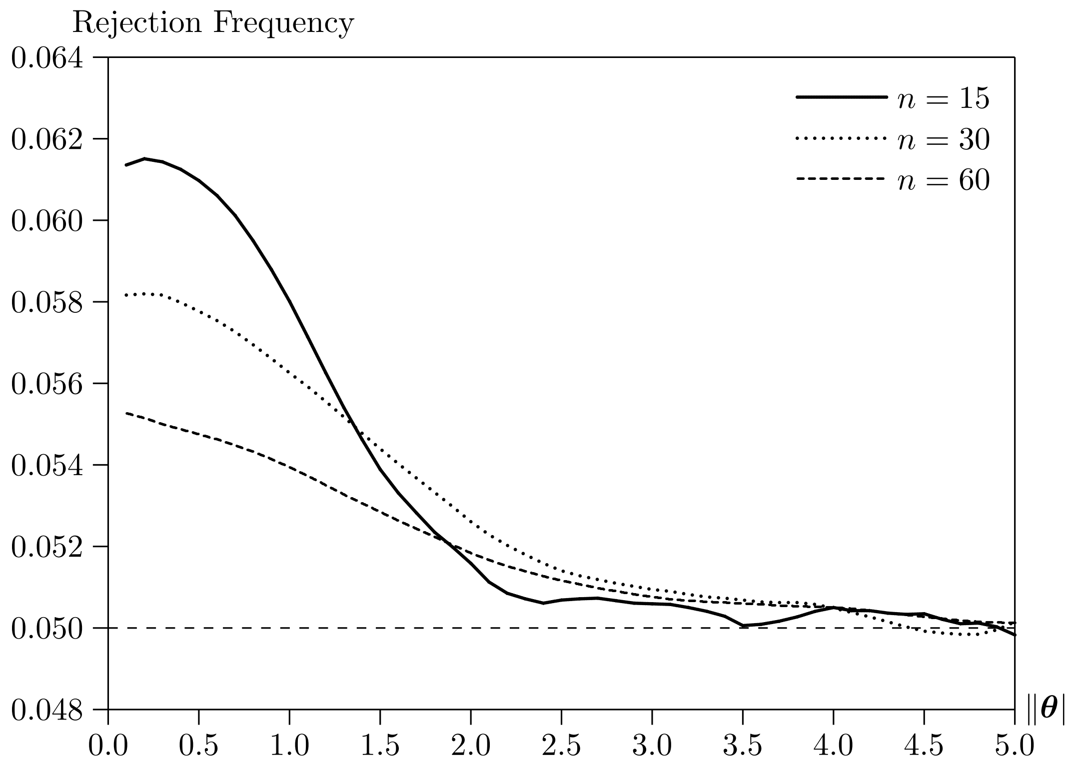

Figure 13. The Relationship between Rejection Frequency and $\|\boldsymbol{\theta}\|$

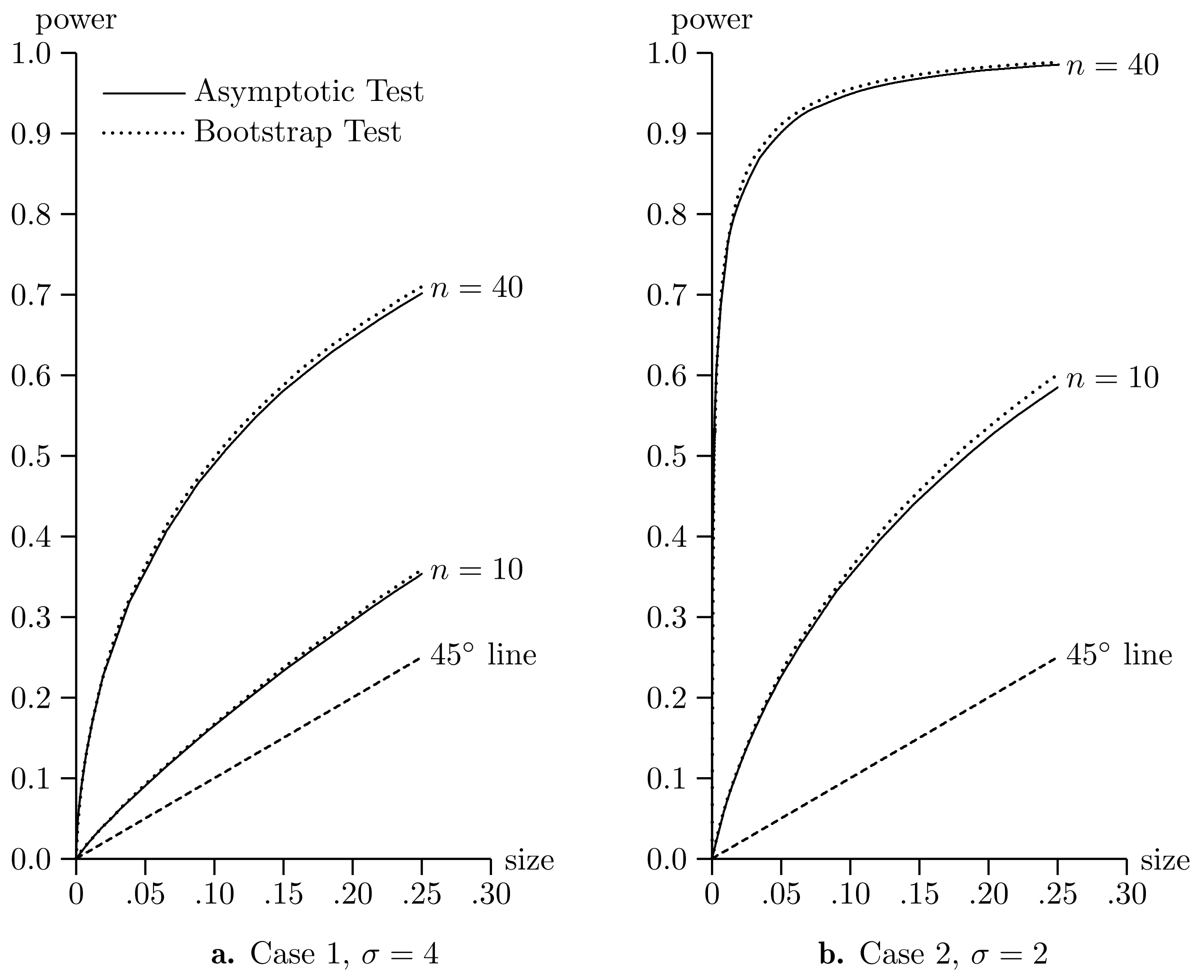

Figure 14. Size-Power Curves for Asymptotic and Bootstrap J Tests