Exact Permutation Tests for Non-nested Non-linear Regression Models

Richard Luger *

Department of Economics, Emory University,
Atlanta, GA 30322-2240, USA

November 2, 2004

Abstract

This paper proposes exact distribution-free permutation tests for the specification of a non-linear regression model against one or more possibly non-nested alternatives. The new tests may be validly applied to a wide class of models, including models with endogenous regressors and lag structures. These tests build on the well-known $J$ test developed by Davidson and MacKinnon (1981) and their exactness holds under broader assumptions than those underlying the conventional $J$ test. The $J$-type test statistics are used with a randomization or Monte Carlo resampling technique which yields an exact and computationally inexpensive inference procedure. A simulation experiment confirms the theoretical results and also shows the performance of the new procedure under violations of the maintained assumptions. The test procedure developed is illustrated by an application to inflation dynamics.

Keywords: Non-nested hypotheses; $J$ test; Finite-sample distribution-free test; Monte Carlo test; New Keynesian Phillips curve

JEL classification: C12; C15; C52

*Tel.: +1-404-727-0328; fax: +1-404-727-4639.
E-mail address: rluger@emory.edu (R. Luger).
1. Introduction

The $J$ test proposed by Davidson and MacKinnon (1981) is a well-known procedure for testing non-nested regression models. This procedure does not yield an exact test so that inference based on it can be misleading in finite samples. For example, Godfrey and Pesaran (1983) report simulation evidence showing that the $J$ test suffers severe size distortions in small samples, rejecting too often, especially when the number of regressors in the alternative model exceeds that in the null model. This over-rejection problem is further exacerbated as the correlations between the non-nested regressors tend to zero. Michelis (1999) derives the asymptotic distribution of the $J$ test statistic when the non-nested regressors are nearly orthogonal. In this case, the $J$ test is no longer asymptotically standard normal as in the case of non-orthogonal models.

Attempts have been made to improve the finite-sample properties of the $J$ test. Fisher and McAleer (1981) proposed a variant of the $J$ test, called the $JA$ test, which is exact in the case of linear regression models with exogenous regressors and normal errors (Godfrey 1983). More recently, simulation methods have been proposed to obtain finite-sample improvements. Fan and Li (1995) and Godfrey (1998), for example, demonstrate that the bootstrap can be used to control the empirical significance level of the $J$ and $JA$ tests. However, neither Fan and Li nor Godfrey provide a theoretical explanation of why the bootstrap apparently works so well when testing non-nested linear regression models.

Davidson and MacKinnon (2002) develop an approach that gives a precise characterization of the finite-sample distribution of the $J$ test. Based on this characterization, they propose a simulation method to obtain an exact version of the $J$ test. Their approach assumes that the error terms are normally distributed and only applies in the case of linear models with exogenous regressors. However, they present simulation evidence which shows that their bootstrap $J$ test perform extremely well even when those assumptions are not satisfied.

This paper extends the permutation principles described in Dufour and Roy (1985), McCabe (1989), and Kennedy (1995) to tests of non-nested non-linear regression models. Although Davidson and MacKinnon’s (1981) $J$ test statistic forms the basic building block of the proposed tests, their validity holds under broader assumptions than those underlying the conventional $J$ test. Here it is assumed that: (i) the variables of the null model and the regressors exclusive to the alternative model are independent, and (ii) that either of these two sets forms a collection of exchangeable random vectors. These assumptions are extended to cover situations where a null model is tested against several alternatives at
once. The proposed tests are exact under those assumptions. This framework is more
general than that in Davidson and MacKinnon (2002) since: (i) the error terms need
not be normally distributed, (ii) the models may contain endogenous regressors and lag
structures, and (iii) they may be non-linear.

The exchangeability assumption means that serial dependence is not allowed in both
the variables of the null model and the regressors exclusive to the alternative model, si-
multaneously. In this respect, the exchangeability assumption is more likely to be satisfied
by cross-sectional or survey data rather than by time-series data. Nevertheless, the null
model may contain time trends or explicit lag structures if the regressors exclusive to the
alternative model are exchangeable.

Section 2 defines the models and the assumptions under which the $J$-type permutation
tests are developed. Section 3 proposes to use the $J$-type test statistics with a random-
ization or Monte Carlo resampling technique that yields an exact and computationally
inexpensive inference procedure. Section 4 presents the results of a simulation experiment
as evidence of the finite-sample performance of the proposed test procedure. Size and
power comparisons are made with the conventional $J$ test, and the results also illustrate
the performance of the new procedure under violations of the maintained assumptions. Section
5 applies the procedure to test the so-called New Keynesian Phillips curve proposed
by Galí and Gertler (1999) against a conventional Phillips curve specification. Section 6
concludes.

2. $J$-type Permutation Tests

2.1 A single non-nested alternative

Let $\{u_t = (y_t, \tilde{x}_t, \tilde{z}_t) : t = 1, ..., T\}$ be a collection of random vectors, where $y_t$ is a
scalar, and $\tilde{x}_t$ and $\tilde{z}_t$ are vectors of conditioning variables. Regroup the common variables
which appear in both $\tilde{x}_t$ and $\tilde{z}_t$ into $w_t$ so that the observations may be rewritten as
$u_t = (y_t, x_t, w_t, z_t)$, where $x_t$ and $z_t$ represent the vectors $\tilde{x}_t$ and $\tilde{z}_t$, respectively, without
the common variables, $w_t$.

Assume the existence of a scalar-valued function, $\varepsilon_t = \varepsilon(y_t, x_t, w_t, z_t)$, such that $\{\varepsilon_t : t = 1, ..., T\}$ is a collection of independently and identically distributed (i.i.d.) random
error terms.

Suppose that one is considering the following parametric model for the conditional
expectation $E(y_t|x_t, w_t, z_t)$:

$$\{f(x_t, w_t; \beta) : t = 1, ..., T\}, \quad (1)$$
where \( f(\cdot) \) is a function with parameters \( \beta \in B; B \subseteq \mathbb{R}^p \). The null hypothesis that (1) is correctly specified is stated as

\[
H_0: \quad \varepsilon_t = y_t - f(x_t, w_t; \beta_0),
\]

so that if (1) is correctly specified, then there exists some value \( \beta_0 \in B \) such that \( \{\varepsilon_t : t = 1, \ldots, T\} \) is an i.i.d. collection of random variables. A general approach to testing the truth of \( H_0 \) is to compare (1) against some alternative parametric specification for \( E(y_t|x_t, w_t, z_t) \). Unlike model-selection procedures where one of the specifications under consideration is always selected, non-nested hypothesis testing asks whether there is any statistically significant departure from the null in the direction of the alternative hypothesis, and need not lead to a favorite model. As Davidson and MacKinnon (1981) emphasize, one need not have any faith in the alternative—it merely gives the direction in which one looks for departures from \( H_0 \). Suppose the alternative is

\[
H_1: \quad \varepsilon_t = y_t - g(x_t, w_t, z_t; \gamma_0),
\]

where \( g(\cdot) \) is a function with parameters \( \gamma \in \Gamma; \gamma \subseteq \mathbb{R}^q \). This setup allows consideration of rival models that differ in the set of conditioning variables, while \( f(x_t, w_t; \beta) \) may or may not be nested within \( g(x_t, w_t, z_t; \gamma) \).

The exactness of the test procedure proposed here rests on the following assumptions.

**Assumption 1.** Either \( \{(y_t, x_t, w_t) : t = 1, \ldots, T\} \) or \( z_1, \ldots, z_T \) is a collection of exchangeable random vectors.

**Assumption 2.** The vectors \( (y_t, x_t, w_t), t = 1, \ldots, T \), are independent of \( z_1, \ldots, z_T \).

Exchangeability in Assumption 1 means that the unconditional distribution of \( (y_1, x_1, w_1), \ldots, (y_T, x_T, w_T) \) (or \( z_1, \ldots, z_T \)) is the same as that of \( (y_{d_1}, x_{d_1}, w_{d_1}), \ldots, (y_{d_T}, x_{d_T}, w_{d_T}) \) (or \( z_{d_1}, \ldots, z_{d_T} \)) for any permutation \( d_1, \ldots, d_T \) of the integers 1, \ldots, T. Essentially, the order of exchangeable observations is irrelevant so that their joint distribution is unaffected when they are permuted. Clearly, i.i.d. observations are exchangeable—but the converse is not necessarily true. For example, jointly normal observations with identical covariances, while not i.i.d., are nevertheless exchangeable (Rao 1973, p. 196); see Galambos (1982), McCabe (1989), and Draper et al. (1983) for more on exchangeability. Data with explicit lag structures or non-stationary data cannot be exchangeable. Therefore, Assumption 1 is more likely to be satisfied by cross-sectional or survey data rather than by time-series data. Nevertheless, the model under \( H_0 \) may
contain time trends and explicit lag structures if \(z_1, ..., z_T\) are exchangeable. Section 4 presents simulation evidence on the performance of the proposed test procedure under violations of Assumptions 1 and 2.

When \(f(x_t, w_t; \beta)\) is not nested within \(g(x_t, w_t, z_t; \gamma)\), the \(J\) test statistic proposed by Davidson and MacKinnon (1981) for testing the truth of \(H_0\) against that of \(H_1\) is computed as the ordinary \(t\)-statistic for \(\delta = 0\) in the artificial compound regression:

\[
y_t = (1 - \delta)f(x_t, w_t; \beta) + \delta g(x_t, w_t, z_t; \hat{\gamma}) + \varepsilon_t,
\]

where \(\hat{\gamma} = \hat{\gamma}((y_1, x_1, w_1, z_1), ..., (y_T, x_T, w_T, z_T))\) denotes the estimate of \(\gamma\) under \(H_1\). Davidson and MacKinnon (1981) proposed an even simpler procedure, referred to as the \(C\) test, where one performs a \(t\)-test for \(\delta = 0\) in

\[
y_t - f(x_t, w_t; \hat{\beta}) = \delta \left(g(x_t, w_t, z_t; \hat{\gamma}) - f(x_t, w_t; \hat{\beta})\right) + \varepsilon_t,
\]

where \(\hat{\beta} = \hat{\beta}((y_1, x_1, w_1), ..., (y_T, x_T, w_T))\) is the estimate of \(\beta\) under \(H_0\). The numerator of the ordinary least squares estimate of \(\delta\), given by

\[
S_C = \sum_{t=1}^T \left(g(x_t, w_t, z_t; \hat{\gamma}) - f(x_t, w_t; \hat{\beta})\right) \left(y_t - f(x_t, w_t; \hat{\beta})\right),
\]

forms the basic building block of the inference procedure proposed here. Other statistics, such as Cox-type test statistics (Cox 1961, 1962), based on a comparison of sum-of-squared residuals could also be considered. However, simulations reveal that permutation tests based on such statistics tend to over-reject far more than those based on (2) when Assumption 2 is violated.

The value of \(\sum_{t=1}^T f(x_t, w_t; \hat{\beta})(y_t - f(x_t, w_t; \hat{\beta}))\) appearing in (2) is invariant to permutations of \(z_1, ..., z_T\) so that an equivalent permutation test can be based on

\[
S_J = \sum_{t=1}^T g(x_t, w_t, z_t; \hat{\gamma})\varepsilon_t^{(0)},
\]

where \(\varepsilon_t^{(0)} = y_t - f(x_t, w_t; \hat{\beta}), t = 1, ..., T\), are the residuals under \(H_0\). The following proposition establishes a permutation principle which is then used to construct a distribution-free \(J\)-type test of \(H_0\) against \(H_1\).

**Proposition 1.** Consider the statistic

\[
S_J(d^*) = \sum_{t=1}^T g(x_t, w_t, z_{d_t}; \hat{\gamma}(d^*))\varepsilon_t^{(0)},
\]
where \( \hat{\gamma}(d^*) \) denotes the estimate of \( \gamma \) in the regression

\[
y_t = g(x_t, w_t, z_{d_t}; \gamma) + \varepsilon_t,
\]

based on \( z_{d_1}, \ldots, z_{d_T} \). Letting \( P_0 \) indicate a probability computed under \( H_0 \), it then follows under Assumptions 1 and 2 that

\[
P_0 \left[ S_J = S_J(d^*) \big| (y_1, x_1, w_1), \ldots, (y_T, x_T, w_T) \right] = \frac{1}{T!}, \tag{5}
\]

for each of the \( T! \) \( S_J(d^*) \) values as \( d^* = (d_1, \ldots, d_T) \) ranges over all possible permutations of the integers \( 1, \ldots, T \).

**Proof.** Let \( \overset{d}{=} \) stand for the equality in distribution. Under Assumption 1, either

\[
(z_1, \ldots, z_T) \overset{d}{=} (z_{d_1}, \ldots, z_{d_T}) \tag{6}
\]
or

\[
(\hat{\varepsilon}_1^{(0)}, \ldots, \hat{\varepsilon}_T^{(0)}) \overset{d}{=} (\hat{\varepsilon}_1^{(0)}(d^*), \ldots, \hat{\varepsilon}_T^{(0)}(d^*)), \tag{7}
\]

where \( \hat{\varepsilon}_t^{(0)}(d^*) = y_{d_t} - f(x_{d_t}, w_{d_t}; \hat{\beta}(d^*)) \) and \( \hat{\beta}(d^*) \) denotes the estimate of \( \beta \) in the regression \( y_{d_t} = f(x_{d_t}, w_{d_t}; \beta) + \varepsilon_t \), based on \( (y_{d_1}, x_{d_1}, w_{d_1}), \ldots, (y_{d_T}, x_{d_T}, w_{d_T}) \). Note that

\[
(\hat{\varepsilon}_1^{(0)}, \ldots, \hat{\varepsilon}_T^{(0)}) = (y_1 - f(x_1, w_1; \hat{\beta}), \ldots, y_T - f(x_T, w_T; \hat{\beta})) \overset{d}{=} (y_{d_1} - f(x_{d_1}, w_{d_1}; \hat{\beta}(d^*)), \ldots, y_{d_T} - f(x_{d_T}, w_{d_T}; \hat{\beta}(d^*))
\]

which establishes (7). The proof of (7) represents a generalization of the result on functions of exchangeable random vectors found in Randles and Wolfe (1979, Theorem 11.2.3), in that here the function \( \hat{\beta}(\cdot) \) is not assumed to be symmetric in its arguments. Assumption 2 combined with either (6) or (7) implies, under \( H_0 \) and conditional on \( (y_1, x_1, w_1), \ldots, (y_T, x_T, w_T) \), that the \( T! \) possible vectors

\[
(g(x_1, w_1, z_{d_1}; \hat{\gamma}(d^*))\hat{\varepsilon}_1^{(0)}, \ldots, g(x_T, w_T, z_{d_T}; \hat{\gamma}(d^*))\hat{\varepsilon}_T^{(0)}),
\]
obtained as \( d^* = (d_1, \ldots, d_T) \) ranges over all possible permutations of the integers \( 1, \ldots, T \), are equally likely values for \( (g(x_1, w_1, z_1; \hat{\gamma})\hat{\varepsilon}_1^{(0)}, \ldots, g(x_T, w_T, z_T; \hat{\gamma})\hat{\varepsilon}_T^{(0)}) \). Hence, conditional on \( (y_1, x_1, w_1), \ldots, (y_T, x_T, w_T) \), the \( T! \) possible values of \( S_J \) represented by \( S_J(d^*) \) are also equally likely under \( H_0 \). □

Based on the equally likely property in (5), an \( \alpha \)-level \( J \)-type permutation test of \( H_0 \) can be performed in the following way. Since large absolute values of \( S_J \) are more probable
under the alternative $H_1$, the null hypothesis $H_0$ is rejected if the observed value of $|S_J|$ falls in a set $C_\alpha$ containing the $T!\alpha$ largest absolute values of the test statistic $S_J(d^*)$ that can be obtained from the class of all permutations. Consider the conditional probability of a Type I error for such a test which may be written as

$$E_0[\mathbb{I}(|S_J| \in C_\alpha) \mid (y_1, x_1, w_1), ..., (y_T, x_T, w_T)] \leq \alpha,$$

where $\mathbb{I}(\cdot)$ is the indicator function. By taking expectations on both sides of (8), such a test is seen to also have level $\alpha$ unconditionally.

Determination of the set $C_\alpha$ by direct counting would be impractical in most cases. Section 3 illustrates how a Monte Carlo resampling technique can be used to perform exact inference without the need to enumerate the entire permutation distribution.

### 2.2. Multiple non-nested alternatives

As with the $J$ test, the $J$-type permutation test can be extended to test a null model against several non-nested alternatives at once. Suppose now that the observations, $u_t$, can be written as $u_t = (y_t, \tilde{x}_t, \tilde{z}_t^{(1)}, ..., \tilde{z}_t^{(m)})$, where $y_t$ is a scalar and $\tilde{x}_t$ and $\tilde{z}_t^{(j)}$, $j = 1, ..., m$, are vectors of conditioning variables. Regroup the common variables which appear in both $\tilde{x}_t$ and $\tilde{z}_t^{(j)}$ into $w_t^{(j)}$ so that the observations may be rewritten as $u_t = (y_t, x_t, w_t^{(1)}, ..., w_t^{(m)}, z_t^{(1)}, ..., z_t^{(m)})$, where $z_t^{(j)}$ represents $\tilde{z}_t^{(j)}$ without $w_t^{(j)}$ and $x_t$ represents $\tilde{x}_t$ without any of the $w_t^{(j)}$'s. Extending the framework, assume now the existence of a scalar-valued function $\varepsilon_t = \varepsilon(y_t, x_t, w_t^{(1)}, ..., w_t^{(m)}, z_t^{(1)}, ..., z_t^{(m)})$ such that $\{\varepsilon_t : t = 1, ..., T\}$ is an i.i.d. collection of random variables.

Consider the problem of testing the non-linear model specification

$$H_0 : \varepsilon_t = y_t - f(x_t, w_t^{(1)}, ..., w_t^{(m)}; \beta_0),$$

against the $m$ alternative models

$$H_j : \varepsilon_t = y_t - g_j(x_t, w_t^{(j)}, z_t^{(j)}; \gamma_{0j}),$$

for $j = 1, ..., m$.

Assumptions 1 and 2 are extended as follows.

**Assumption 1'**. Either $\{(y_t, x_t, w_t^{(1)}, ..., w_t^{(m)}) : t = 1, ..., T\}$ or $\{(z_t^{(1)}, ..., z_t^{(m)}) : t = 1, ..., T\}$ is a collection of exchangeable random vectors.

**Assumption 2'**. The vectors $(y_s, x_s, w_s^{(1)}, ..., w_s^{(m)})$ are independent of $(z_t^{(1)}, ..., z_t^{(m)})$ for all $s, t$. 

6
When \( f(x_t, w_t^{(1)}, ..., w_t^{(m)}; \beta) \) is not nested within \( g_j(x_t, w_t^{(j)}, z_t^{(j)}; \gamma_j) \), \( j = 1, ..., m \), a joint \( J \) test is an \( F \) test of \( \delta_1 = ... = \delta_m = 0 \) in the regression

\[
y_t = \left( 1 - \sum_{j=1}^{m} \delta_j \right) f(x_t, w_t^{(1)}, ..., w_t^{(m)}; \beta) + \sum_{j=1}^{m} \delta_j g_j(x_t, w_t^{(j)}, z_t^{(j)}; \gamma_j) + \varepsilon_t,
\]

where \( \hat{\gamma}_j \) is the estimate of \( \gamma_j \) under \( H_j \), \( j = 1, ..., m \). Generalizing (3) yields the following joint \( J \)-type permutation test statistic:

\[
F_J = \hat{\varepsilon}' \hat{G} \hat{G}' \hat{\varepsilon},
\]

where \( \hat{\varepsilon} = (\hat{\varepsilon}_1^{(0)}, ..., \hat{\varepsilon}_T^{(0)})' \) and \( \hat{G} = [\hat{g}_1, ..., \hat{g}_m] \) is a \( T \times m \) matrix with columns \( \hat{g}_j = (g_j(x_1, w_1^{(j)}, z_1^{(j)}; \hat{\gamma}_j), ..., g_j(x_T, w_T^{(j)}, z_T^{(j)}; \hat{\gamma}_j))' \), \( j = 1, ..., m \).

**Proposition 2.** Consider the statistic

\[
F_J(d^*) = \hat{\varepsilon}' \hat{G}(d^*) \hat{G}(d^*)' \hat{\varepsilon},
\]

where

\[
\hat{G}(d^*) = [\hat{g}_1(d^*), ..., \hat{g}_m(d^*)],
\]

\[
\hat{g}_j(d^*) = (g_j(x_1, w_1^{(j)}, z_1^{(j)}; \hat{\gamma}_j(d^*))), ..., g_j(x_T, w_T^{(j)}, z_T^{(j)}; \hat{\gamma}_j(d^*)))',
\]

and where \( \hat{\gamma}_j(d^*) \) is the estimate of \( \gamma_j \) in the regression

\[
y_t = g_j(x_t, w_t^{(j)}, z_t^{(j)}; \gamma_j) + \varepsilon_t,
\]

based on \( (z_1^{(j)}, ..., z_T^{(j)}) \), \( j = 1, ..., m \). Under \( H_0 \) and Assumptions 1' and 2',

\[
P_0 \left[ F_J = F_J(d^*) \mid (y_1, x_1, w_1^{(1)}, ..., w_1^{(m)}), ..., (y_T, x_T, w_T^{(1)}, ..., w_T^{(m)}) \right] = \frac{1}{T!},
\]

for each of the \( T! \) \( F_J(d^*) \) values as \( d^* = (d_1, ..., d_T) \) ranges over all possible permutations of the integers 1, ..., \( T \).

This result is a straightforward generalization of Proposition 1 and, in a completely analogous fashion, it establishes a permutation principle which can be used to perform a distribution-free \( J \)-type one-sided test of \( H_0 \) against \( H_1, ..., H_m \) at once.
3. Exact Inference Based on Monte Carlo Resampling

Generation of the exact permutation distribution of the $J$-type test statistics by a complete enumeration of all possible permutations is computationally prohibitive for sample sizes typical in applied work. The computational burden of finding tail probabilities can be reduced by drawing random samples from the permutation distribution and computing the value of the $J$-type test statistic each time. The relative frequencies of these values comprise the simulated permutation distribution (Pratt and Gibbons 1981, p. 210). A simple method to obtain the desired significance level and a precise $p$-value without performing a large number of random draws, is the Monte Carlo procedure proposed by Dwass (1957). The construction of a Monte Carlo test is illustrated with the absolute value of $S_J$ as a two-sided test of $H_0$ against a single alternative. The same procedure can be applied directly with $F_J$ as a one-sided test when multiple non-nested alternatives are considered at once.

Let $|S_J(d^*_B)|$ denote the absolute value of $S_J$ computed from the original sample and $|S_J(d^*_b)|$, $b = 1, ..., B - 1$, those obtained by randomly sampling (with or without replacement) the permutation distribution; i.e., $|S_J(d^*_b)|$ is the absolute value of $S_J$ computed from $(y_1, w_1, z_{d_1}, \tilde{\varepsilon}_1^{(0)}), ..., (y_T, w_T, z_{d_T}, \tilde{\varepsilon}_T^{(0)})$ according to (4) based on a randomly chosen permutation $d^*_b = (d_1, ..., d_T)$ of the integers $1, ..., T$. It is important to emphasize that for each permutation, $\hat{\gamma}(d^*_b)$ is the estimate of $\gamma$ in the regression $y_t = g(x_t, w_t, z_{d_t}; \gamma) + \varepsilon_t$, which is based on the permuted vector $(z_{d_1}, ..., z_{d_T})$ while $((y_1, w_1, \tilde{\varepsilon}_1^{(0)}), ..., (y_T, w_T, \tilde{\varepsilon}_T^{(0)}))$ is held in the same order used to compute the original $S_J$.

Note that the permutation distribution of the $S_J$’s is discrete so that ties among the randomly sampled $S_J$ statistics have a non-zero probability of occurrence. A test with size $\alpha$ can be obtained by applying the following tie-breaking rule (Dufour 2004). Draw $B$ variates $U_i$, $i = 1, ..., B$, from a continuous uniform distribution independently of the $S_J(d^*_i)$’s and arrange the pairs $(|S_J(d^*_i)|, U_1), ..., (|S_J(d^*_B)|, U_B)$ according to the lexicographic order: $(|S_J(d^*_i)|, U_i) < (|S_J(d^*_j)|, U_j) \iff |S_J(d^*_i)| < |S_J(d^*_j)|$ or $(|S_J(d^*_i)| = |S_J(d^*_j)|$ and $U_i < U_j)$.

Let $\hat{R}_B$ denote the rank of $(|S_J(d^*_B)|, U_B)$ in the lexicographic ordering, which is easily computed as:

$$\hat{R}_B = 1 + \sum_{i=1}^{B-1} \mathbb{I}(|S_J(d^*_B)| > |S_J(d^*_i)|) + \sum_{i=1}^{B-1} \mathbb{I}(|S_J(d^*_B)| = |S_J(d^*_i)|) \times \mathbb{I}(U_B > U_i),$$

(9)

where $\mathbb{I}(\cdot)$ is again the indicator function. If $\alpha B$ is an integer, then

$$P_0 \left[ \hat{R}_B \geq B - \alpha B + 1 \right] = \alpha$$

8
so that $p_B = (B - \tilde{R}_B + 1)/B$ can be interpreted as a randomized $p$-value, which can be used to perform a test with size $\alpha$.

For a given, possibly small, number of random draws, the Monte Carlo procedure allows one to control exactly the size of the test. This feature stands in sharp contrast with bootstrap test procedures which generally are only valid asymptotically. As $B$ increases without bound, inference based on the Monte Carlo procedure becomes equivalent to that based on the exact permutation distribution. See Dufour (2004) and Dufour and Khalaf (2001) for more on the technique of Monte Carlo testing.

4. Simulation Experiment

Several simulation experiments have shown that the finite-sample distribution of the $J$ test can be very different from its asymptotic $N(0, 1)$ distribution. Godfrey and Pesaran (1983), Godfrey (1998), and Michelis (1999), for example, show that the $J$ test can be oversized. In particular, they show that the over-rejection problem tends to be especially severe when one or more of the following features is present: (i) a poor fit of the null model, (ii) low correlations between the regressors of the rival models, and (iii) the number of regressors in the alternative models exceeds that of the null model. This section reports the results of a simulation experiment in which the proposed Monte Carlo $J$ (MC $J$) test procedure is compared to the $J$ test. The simulation experiment, which is similar to Godfrey and Pesaran (1983), examines a very prominent example of conditional non-nested models: linear normal regression models with rival sets of conditioning variables.

The null model is generated as

\[ H_0 : y_t = \phi_y y_{t-1} + \sum_{j=1}^{k_0} x_{tj} + \varepsilon_t^{(0)}, \quad t = 1, \ldots, T, \]

where $|\phi_y| < 1$, the exogenous variables $x_{tj}$ are $IIN(0, 1)$, and the error terms $\varepsilon_t^{(0)}$ are $IIN(0, \sigma_0^2)$ independent of $x_{tj}$ for all $t, j$.

Under the alternative, the data are generated according to

\[ H_1 : y_t = \phi_y y_{t-1} + \sum_{j=1}^{k_1} z_{tj} + \varepsilon_t^{(1)}, \quad t = 1, \ldots, T, \]

where $\varepsilon_t^{(1)}$ are $IIN(0, \sigma_1^2)$. The exogenous variables under the alternative are given by

\[ z_{tj} = \phi_z z_{t-1} + \lambda x_{tj} + v_{tj}, \quad j = 1, \ldots, \min(k_0, k_1), \]
and, if \( k_1 > k_0 \),
\[
z_{tj} = \phi_z z_{t-1} + v_{tj}, \quad j = k_0 + 1, \ldots, k_1,
\]
where \(|\phi_z| < 1 \) and \( v_{tj} \sim IN(0,1) \) for \( t = 1, \ldots, T \). The value of \( \lambda \) is set by choosing the value of the correlation coefficient \( \rho \) between \( x_{tj} \) and \( z_{tj} \) as \( \lambda = \rho / \sqrt{(1 - \rho^2 - \phi_z^2)} \), for \( j = 1, \ldots, \min(k_0, k_1) \); the other \( x_{tj} \)'s and \( z_{tj} \)'s are uncorrelated. The initial values \( z_{0j}, j = 1, \ldots, k_1 \), are random draws from the stationary normal distribution which has mean zero and variance \((\lambda^2 + 1)/(1 - \phi_z^2)\).

The variances \( \sigma_0^2 \) and \( \sigma_1^2 \) are determined by specifying the population coefficient of determination \( R^2 \) of the true model so that under \( H_0 \):
\[
\sigma_0^2 = k_0(1 - R^2)/(R^2 - \phi_y^2),
\]
while under \( H_1 \):
\[
\sigma_1^2 = k_1(1 + \lambda^2)(1 - R^2)/((R^2 - \phi_y^2)(1 - \phi_z^2)), \quad \text{if} \ k_0 \geq k_1,
\]
\[
= (k_0 \lambda^2 + k_1)(1 - R^2)/((R^2 - \phi_y^2)(1 - \phi_z^2)), \quad \text{if} \ k_0 < k_1,
\]
where in each case \( R^2 > \phi_y^2 \). The initial value \( y_0 \) is a random draw from the stationary normal distribution which has mean zero and variance: \((k_0 + \sigma_0^2)/(1 - \phi_y^2)\) under \( H_0 \), \((k_1(1 + \lambda^2 + \sigma_1^2))/(1 - \phi_y^2)(1 - \phi_z^2)\) under \( H_1 \) when \( k_0 \geq k_1 \), and \((k_0 \lambda^2 + k_1 + \sigma_1^2)/(1 - \phi_y^2)(1 - \phi_z^2)\) under \( H_1 \) when \( k_0 < k_1 \).

In this setup, Assumption 1 is satisfied only when \( \phi_y = 0 \) or \( \phi_z = 0 \), and Assumption 2 is satisfied only when \( \rho = 0 \). In any other case, the proposed test procedure is no longer exact.

Let \( Y = (y_1, \ldots, y_T)' \), \( X = [W, X_1, \ldots, X_{k_0}] \), \( W = (w_1, \ldots, w_T)' \), where \( w_t = y_{t-1} \), \( X_j = (x_{1j}, \ldots, x_{Tj})' \) for \( j = 1, \ldots, k_0 \), and \( Z = [W, Z_1, \ldots, Z_{k_1}] \), \( Z_j = (z_{1j}, \ldots, z_{Tj})' \) for \( j = 1, \ldots, k_1 \). When the two rival models are linear, the \( J \) test statistic becomes
\[
J = \frac{(P_Z Y)'\hat{\epsilon}}{\hat{\sigma}(Y'P_Z M_X P_Z Y)^{1/2}},
\]
where \( \hat{\epsilon} = M_X Y = (\hat{\epsilon}_1^{(0)}, \ldots, \hat{\epsilon}_T^{(0)})' \), with \( M_X = I - X(X'X)^{-1}X' \) and \( P_Z = Z(Z'Z)^{-1}Z' \), and where \( \hat{\sigma} \) is the estimate of the standard deviation of the error in the artificial compound regression. For each replication performed under the various data-generating configurations considered, the \( J \) statistic is compared with the critical value from the \( N(0, 1) \) distribution for a nominal 5 per cent level two-sided test. The MC \( J \) test, based on the absolute value of \( S_J = (P_Z Y)'\hat{\epsilon} \), is implemented with \( B = 100 \) so that the null is rejected at the 5 per
cent level when $\hat{R}_{100}$ computed according to (9) is greater or equal to 96. The simulated statistics are computed as the absolute value of

$$S_J(d^*_b) = \left(Z(d^*_b)(Z(d^*_b)'Z(d^*_b))^{-1}Z(d^*_b)'Y\right)'\hat{\epsilon},$$

where $Z(d^*_b) = [W, Z_1(d^*_b), ..., Z_q(d^*_b)]$ with $Z_j(d^*_b) = (z_{d_1j}, ..., z_{d_Tj})'$, $j = 1, ..., k_1$, for each randomly drawn permutation $d^*_b = (d_1, ..., d_T)$ of the integers 1, ..., $T$ for $b = 1, ..., 99$.

Table 1 reports the empirical rejection rates under $H_0$ with $\phi_z = 0$. In this case, $\{(z_{t1}, ..., z_{tk_1}) : t = 1, ..., T\}$ is an exchangeable collection so that Assumption 1 is satisfied. Assumption 2, however, is only satisfied when $\rho = 0$. The proposed test procedure is therefore theoretically exact only in the left-most column of entries for the MC $J$ test in Table 1. As expected, the empirical rejection rates closely correspond to the nominal level in that case. The empirical size of the $J$ test in many cases is considerably more than the nominal level. All else equal, the $J$ test appears oversized when: $\rho$ is close to zero, $k_1$ exceeds $k_0$, $R^2$ is low, and $\phi_y$ is high. In some of those cases, the $J$ test has rejection rates of over 40 per cent—more than eight times the nominal level. On the other hand, there is a closer agreement between the nominal and empirical rejection probabilities for the MC $J$ test. Compare the rejection rates of the MC $J$ test when $\rho = 0$. As expected, the MC $J$ test is not affected by different combinations of $(k_0, k_1)$, values of $R^2$, and values of $\phi_y$ since each of these features are implicitly accounted for by the test procedure. Assumption 2 is violated as soon as $\rho \neq 0$, in which cases the exogenous regressors of the rival models are no longer independent. The MC $J$ test appears somewhat undersized in many cases when $\rho$ is high, which of course may result in a power loss. The robustness of the MC $J$ test to non-zero values of $\rho$ is remarkable.

Table 2 repeats the size experiment setting the value of the autoregressive coefficient $\phi_z$ at 0.95. Note that each of the $k_1$ series of exogenous regressors under the alternative model have the same autocorrelation structure. Compared to the $J$ test, the MC $J$ behaves remarkably well even under this gross violation of the exchangeability assumption. While
in many cases the $J$ test has rejection rates of over 60 per cent, those of the MC $J$ test never exceed 15 per cent. Of course, these results depend on the sample size. Unlike asymptotic tests whose size distortions decrease as the sample size grows, the MC $J$ test is likely to perform better in smaller samples for given non-zero values of $\rho$ and $\phi_z$. In this respect, the MC $J$ test and the $J$ test can be seen as complements.

The power of the $J$ test increases with the population coefficient of determination, $R^2$, and the sample size, $T$. In order to have a contrast between the power functions of each test, Tables 3 and 4 report percentage power estimates for lower values of $R^2$ and $\phi_y$. The power results for the conventional $J$ test are based on size-corrected critical values.

Table 3 reports the empirical rejection rates under $H_1$ when $\phi_z = 0$ based on critical values for a nominal 5 per cent level test. One immediately notices that the tests have no power whenever $\phi_y$ is strictly greater than $R^2$. As Godfrey and Pesaran (1983, p. 149) note, this phenomenon is due to the fact that the regressors specific to the null are of little consequence in these cases. The same reason explains the overall decrease in power seen as $\phi_z$ is increased from 0 in Table 3 to 0.95 in Table 4. When comparing the $J$ and MC $J$ tests, the picture that emerges from Table 3 is one of good power with relatively small differences between the two tests, despite the small sample size. In Table 4, the $J$ test is seen to be more powerful than the MC $J$ test. However, it is important to emphasize that the size-corrected $J$ test is not a feasible test in practice. It is merely used here as a theoretical benchmark to which the MC $J$ test may be compared.

5. Application to Inflation Dynamics

In conventional specifications of the Phillips curve, inflation is related to unemployment in the short run. According to this conventional view, movements in the unemployment rate have predictive power for future changes in inflation. This inflation-unemployment tradeoff is fundamental to understanding the effects of monetary policy and other events that can influence the aggregate demand for goods and services. Galí and Gertler (1999) suggest a so-called New Keynesian Phillips curve (NKPC), which has two distinguishing features. First, the inflation process is a function of the expected future inflation rate, and second, it is a measure of real marginal cost rather than the unemployment rate that is the relevant determinant of inflation. These features are a consequence of their framework in
which imperfectly competitive firms optimally set prices in anticipation of future demand and factor costs, subject to costly price adjustment constraints.

The reduced-form NKPC advocated by Galí and Gertler is

$$\pi_t = \omega_f E_t \pi_{t+1} + \omega_b \pi_{t-1} + \omega_x x_t,$$

where $\pi_t$ is inflation at time $t$, $E_t \pi_{t+1}$ is the time-$t$ expectation of next period inflation, and $x_t$ is a measure of real marginal cost. Assuming rational expectations, the inflation equation in (10) can be rewritten as

$$\pi_{t+1} = \frac{1}{\omega_f} \pi_t - \frac{\omega_b}{\omega_f} \pi_{t-1} - \frac{\omega_x}{\omega_f} x_t + \varepsilon_{t+1},$$

where the implied forecast error, $\varepsilon_{t+1}$, is orthogonal to information timed $t$ and earlier. The proposed test procedure can be used to test the null hypothesis that the NKPC is correctly specified, formally stated as

$$H_0 : \pi_t = \phi_1 \pi_{t-1} + \phi_2 \pi_{t-2} + \beta x_{t-1} + \varepsilon_t,$$

against a non-nested alternative of the form:

$$H_1 : \pi_t = \phi_1 \pi_{t-1} + \phi_2 \pi_{t-2} + \gamma u_{t-1} + \varepsilon_t,$$

where inflation is related to the (lagged) unemployment rate, $u_{t-1}$, as in conventional specifications of the Phillips curve. These formulations, involving only lagged variables on the right-hand side, are especially useful for forecasting purposes (Stock and Watson 1999). Model parameters are estimated by ordinary least squares. A virtue of the proposed test procedure is that its exactness does not depend on the choice of estimation method. This is particularly attractive in light of the weak identification issues raised by Ma (2002) and Nason and Smith (2003) with typical generalized-method-of-moments estimation of the structural parameters of the NKPC.

The data is quarterly for the U.S. over the period 1960:1-1999:4. Inflation is measured as $100 \log(P_t/P_{t-4})$, where $P_t$ is the GDP implicit price deflator. Real marginal cost is measured as $100 \log(s_t - \bar{s})$, where $s_t$ is the U.S. Bureau of Labor Statistics measure of nominal unit labor costs divided by the non-farm business implicit price deflator (which is proportional to the share of labor income in nominal output for the non-farm business sector). These data definitions are similar to those in Galí and Gertler (1999). The unemployment rate is defined as the deviation of the civilian unemployment rate from its sample mean. The data used here are available on the Federal Reserve Bank of St. Louis Web site. There are 154 observations once the first two lags of inflation are allowed for.
Ordinary least squares yields
\[ H_0 : \hat{\pi}_t = 1.56 \pi_{t-1} - 0.57 \pi_{t-2} + 0.015 x_{t-1}, \quad R^2 = 0.982, \]
and
\[ H_1 : \hat{\pi}_t = 1.46 \pi_{t-1} - 0.47 \pi_{t-2} - 0.062 u_{t-1}, \quad R^2 = 0.981, \]
where the numbers in parentheses are usual standard errors. The first-order autocorrelations are 0.92 for the real marginal cost and 0.98 for the unemployment rate; the two series have a contemporaneous correlation of 0.18.

The \( J \) test statistic takes value 3.48 with a \( p \)-value of 0.00024. The MC \( J \) test with \( B = 1000 \) yields a \( p \)-value of 0.0020. In this case, both tests strongly reject the null model with real marginal cost at conventional levels. Given the similarities of the two models, one might expect the tests to yield similar conclusions once the roles of \( H_0 \) and \( H_1 \) are reversed. When the models are reversed and the unemployment rate appears under the null, the \( p \)-values change dramatically. In that case, the \( J \) test statistic takes value 1.78 (\( p \)-value of 0.037) and the MC \( J \) test has a \( p \)-value of 0.20. While the result of the \( J \) test might be construed as a (marginal) rejection, the MC \( J \) test clearly indicates a non-rejection. These results do not depend on whether \( \pi_t \) or \( \Delta \pi_t \) appears as regressand.

Although Galí and Gertler’s NKPC is appealing on theoretical grounds, the empirical application of the proposed test procedure shows that the real marginal cost fails to describe the inflation process. Rather, the data support the conventional view that unemployment is the relevant determinant of inflation. The results in Kurmann (2003) also cast considerable doubt on the role of real marginal cost in a New Keynesian inflation equation.

6. Conclusion

This paper has shown how finite-sample distribution-free tests for the specification of a non-linear regression model against one or more possibly non-nested alternatives can be obtained by first considering permutation versions of the well-known \( J \) test of Davidson and MacKinnon (1981), and then implementing these permutation \( J \)-type tests with a Monte Carlo resampling technique. The proposed test procedure is exact under Assumptions 1 and 2 (or Assumptions 1’ and 2’ in cases involving multiple alternatives). Under those assumptions, the MC \( J \) test offers a solution to the potential over-rejection problem of the \( J \) test. In particular, the MC \( J \) test is not affected by the number of regressors in each model, the goodness of fit, and the degree of persistence in the null model.
Another advantage of the permutation tests described in this paper is that their validity does not depend on the estimation method. This feature can be particularly attractive in problems which involve parameters that are locally almost unidentified such as regression models with ratios of parameters and instrumental variables regressions with weak instruments. The permutation principle also yields test procedures that are invariant to departures from standard assumptions, such as those of normality and the existence of moments, required for the validity of many parametric test methods.

Acknowledgements

This paper was written while the author was affiliated with the Bank of Canada. The views expressed in this paper are those of the author. No responsibility for them should be attributed to the Bank of Canada. The author thanks two anonymous referees and especially Jean-Marie Dufour for several helpful comments.
References


16


<table>
<thead>
<tr>
<th>$(k_0, k_1)$</th>
<th>$R^2$</th>
<th>$\phi_y$</th>
<th>$\rho$</th>
<th>0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.4</th>
<th>0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2,2)</td>
<td>0.65</td>
<td>0.5</td>
<td>13.38</td>
<td>J</td>
<td>MC J</td>
<td>13.22</td>
<td>5.12</td>
<td>12.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8</td>
<td>18.16</td>
<td>J</td>
<td>MC J</td>
<td>18.00</td>
<td>5.10</td>
<td>17.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.90</td>
<td>9.00</td>
<td>J</td>
<td>MC J</td>
<td>8.82</td>
<td>5.22</td>
<td>8.30</td>
</tr>
<tr>
<td>(2,4)</td>
<td>0.65</td>
<td>0.5</td>
<td>29.10</td>
<td>J</td>
<td>MC J</td>
<td>28.72</td>
<td>5.44</td>
<td>27.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8</td>
<td>48.36</td>
<td>J</td>
<td>MC J</td>
<td>47.80</td>
<td>5.26</td>
<td>47.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.90</td>
<td>13.96</td>
<td>J</td>
<td>MC J</td>
<td>13.66</td>
<td>5.20</td>
<td>12.46</td>
</tr>
<tr>
<td>(4,2)</td>
<td>0.65</td>
<td>0.5</td>
<td>12.76</td>
<td>J</td>
<td>MC J</td>
<td>12.76</td>
<td>5.28</td>
<td>12.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8</td>
<td>17.48</td>
<td>J</td>
<td>MC J</td>
<td>17.44</td>
<td>5.18</td>
<td>17.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.90</td>
<td>8.84</td>
<td>J</td>
<td>MC J</td>
<td>8.46</td>
<td>5.04</td>
<td>8.20</td>
</tr>
<tr>
<td>(4,4)</td>
<td>0.65</td>
<td>0.5</td>
<td>25.88</td>
<td>J</td>
<td>MC J</td>
<td>25.66</td>
<td>4.58</td>
<td>24.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8</td>
<td>43.06</td>
<td>J</td>
<td>MC J</td>
<td>42.68</td>
<td>5.16</td>
<td>41.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.90</td>
<td>11.68</td>
<td>J</td>
<td>MC J</td>
<td>12.28</td>
<td>4.50</td>
<td>11.78</td>
</tr>
</tbody>
</table>

Note: The MC $J$ test is theoretically exact only when $\rho = 0$; cases where $\rho \neq 0$ represent departures from Assumption 2. Nominal level is 5 per cent. Results based on 5,000 replications.
Table 2
Empirical Rejection Probabilities (in Percentages) under $H_0$ with $\phi_z = 0.95$

<table>
<thead>
<tr>
<th>$(k_0, k_1)$</th>
<th>$R^2$</th>
<th>$\phi_y$</th>
<th>0</th>
<th></th>
<th>0.1</th>
<th></th>
<th>0.2</th>
<th></th>
<th>0.4</th>
<th></th>
<th>0.7</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$J$</td>
<td>MC $J$</td>
<td>$J$</td>
<td>MC $J$</td>
<td>$J$</td>
<td>MC $J$</td>
<td>$J$</td>
<td>MC $J$</td>
<td>$J$</td>
<td>MC $J$</td>
</tr>
<tr>
<td>(2,2)</td>
<td>0.65</td>
<td>0.5</td>
<td>14.92</td>
<td>7.22</td>
<td>14.76</td>
<td>7.70</td>
<td>14.80</td>
<td>8.70</td>
<td>8.32</td>
<td>8.18</td>
<td>7.90</td>
<td>8.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>30.26</td>
<td>10.70</td>
<td>30.58</td>
<td>10.70</td>
<td>31.06</td>
<td>9.40</td>
<td>27.52</td>
<td>7.56</td>
<td>25.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.90</td>
<td>9.44</td>
<td>6.70</td>
<td>9.26</td>
<td>7.68</td>
<td>9.04</td>
<td>9.92</td>
<td>3.84</td>
<td>12.82</td>
<td>3.38</td>
</tr>
<tr>
<td>(2,4)</td>
<td>0.65</td>
<td>0.5</td>
<td>37.22</td>
<td>9.42</td>
<td>36.46</td>
<td>9.66</td>
<td>35.94</td>
<td>10.02</td>
<td>32.62</td>
<td>9.16</td>
<td>31.52</td>
<td>9.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>69.30</td>
<td>13.12</td>
<td>69.50</td>
<td>13.32</td>
<td>70.26</td>
<td>12.38</td>
<td>73.46</td>
<td>10.22</td>
<td>71.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.90</td>
<td>16.04</td>
<td>7.82</td>
<td>15.72</td>
<td>7.94</td>
<td>15.90</td>
<td>9.80</td>
<td>10.84</td>
<td>9.80</td>
<td>10.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>29.80</td>
<td>10.90</td>
<td>29.46</td>
<td>11.42</td>
<td>30.94</td>
<td>13.60</td>
<td>28.36</td>
<td>13.42</td>
<td>25.36</td>
</tr>
<tr>
<td>(4,2)</td>
<td>0.65</td>
<td>0.5</td>
<td>14.60</td>
<td>7.00</td>
<td>14.02</td>
<td>6.86</td>
<td>13.30</td>
<td>6.48</td>
<td>9.34</td>
<td>6.42</td>
<td>9.22</td>
<td>6.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>27.10</td>
<td>10.40</td>
<td>26.82</td>
<td>9.98</td>
<td>26.28</td>
<td>8.26</td>
<td>23.52</td>
<td>6.54</td>
<td>22.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.90</td>
<td>9.48</td>
<td>6.20</td>
<td>9.12</td>
<td>6.82</td>
<td>8.76</td>
<td>7.22</td>
<td>4.76</td>
<td>8.40</td>
<td>4.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>13.12</td>
<td>8.24</td>
<td>12.68</td>
<td>8.64</td>
<td>12.36</td>
<td>8.78</td>
<td>8.26</td>
<td>9.88</td>
<td>7.32</td>
</tr>
<tr>
<td>(4,4)</td>
<td>0.65</td>
<td>0.5</td>
<td>32.64</td>
<td>8.80</td>
<td>32.24</td>
<td>8.70</td>
<td>32.10</td>
<td>8.14</td>
<td>25.78</td>
<td>5.90</td>
<td>25.18</td>
<td>7.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>64.08</td>
<td>13.70</td>
<td>64.44</td>
<td>13.10</td>
<td>64.28</td>
<td>10.40</td>
<td>66.40</td>
<td>6.26</td>
<td>65.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.90</td>
<td>15.36</td>
<td>7.68</td>
<td>14.90</td>
<td>7.98</td>
<td>14.74</td>
<td>7.84</td>
<td>7.96</td>
<td>7.94</td>
<td>7.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>26.62</td>
<td>10.64</td>
<td>26.16</td>
<td>10.72</td>
<td>27.16</td>
<td>10.78</td>
<td>21.46</td>
<td>9.18</td>
<td>19.26</td>
</tr>
</tbody>
</table>

Note: Entries correspond to various departures from Assumptions 1 and 2; see text for details. Nominal level is 5 per cent. Results based on 5,000 replications.
<table>
<thead>
<tr>
<th>((k_0, k_1))</th>
<th>(R^2)</th>
<th>(\phi_y)</th>
<th>0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.4</th>
<th>0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(J)</td>
<td>MC(J)</td>
<td>(J)</td>
<td>MC(J)</td>
<td>(J)</td>
</tr>
<tr>
<td>(2,2)</td>
<td>0.5</td>
<td>0.5</td>
<td>67.38</td>
<td>64.34</td>
<td>67.28</td>
<td>63.68</td>
<td>67.16</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>7.72</td>
<td>7.74</td>
<td>7.80</td>
<td>7.34</td>
<td>7.90</td>
<td>6.78</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>97.56</td>
<td>96.30</td>
<td>97.68</td>
<td>96.18</td>
<td>97.50</td>
<td>95.76</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>81.92</td>
<td>78.52</td>
<td>81.48</td>
<td>78.24</td>
<td>81.08</td>
<td>76.60</td>
</tr>
<tr>
<td>(2,4)</td>
<td>0.5</td>
<td>0.5</td>
<td>58.60</td>
<td>50.38</td>
<td>57.14</td>
<td>50.14</td>
<td>57.18</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>6.76</td>
<td>6.90</td>
<td>6.60</td>
<td>6.72</td>
<td>6.58</td>
<td>6.10</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>96.20</td>
<td>90.90</td>
<td>96.28</td>
<td>90.86</td>
<td>96.20</td>
<td>90.40</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>73.38</td>
<td>63.88</td>
<td>73.08</td>
<td>63.64</td>
<td>73.04</td>
<td>62.90</td>
</tr>
<tr>
<td>(4,2)</td>
<td>0.5</td>
<td>0.5</td>
<td>60.88</td>
<td>62.34</td>
<td>61.48</td>
<td>61.40</td>
<td>61.50</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>6.82</td>
<td>7.22</td>
<td>7.00</td>
<td>7.08</td>
<td>7.24</td>
<td>6.54</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>96.20</td>
<td>96.36</td>
<td>96.32</td>
<td>95.92</td>
<td>96.14</td>
<td>95.54</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>74.36</td>
<td>76.24</td>
<td>75.58</td>
<td>75.60</td>
<td>75.76</td>
<td>74.50</td>
</tr>
<tr>
<td>(4,4)</td>
<td>0.5</td>
<td>0.5</td>
<td>53.74</td>
<td>48.44</td>
<td>53.72</td>
<td>47.44</td>
<td>53.74</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>6.64</td>
<td>6.46</td>
<td>6.70</td>
<td>6.32</td>
<td>6.64</td>
<td>5.86</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>94.58</td>
<td>88.98</td>
<td>94.68</td>
<td>88.26</td>
<td>94.60</td>
<td>87.20</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>68.80</td>
<td>61.50</td>
<td>68.78</td>
<td>60.86</td>
<td>68.36</td>
<td>59.24</td>
</tr>
</tbody>
</table>

Note: The \(J\) test is compared to size-corrected critical values. Nominal level is 5 per cent. Results based on 5,000 replications.
Table 4

Empirical Rejection Probabilities (in Percentages) under $H_1$ with $\phi_z = 0.95$

<table>
<thead>
<tr>
<th>$(k_0, k_1)$</th>
<th>$R^2$</th>
<th>$\phi_y$</th>
<th>$0$</th>
<th>$0.1$</th>
<th>$0.2$</th>
<th>$0.4$</th>
<th>$0.7$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$J$</td>
<td>MC $J$</td>
<td>$J$</td>
<td>MC $J$</td>
<td>$J$</td>
</tr>
<tr>
<td>$(2,2)$</td>
<td>0.5</td>
<td>0.5</td>
<td>40.00</td>
<td>29.06</td>
<td>39.20</td>
<td>27.96</td>
<td>38.10</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>5.90</td>
<td>8.94</td>
<td>5.22</td>
<td>8.72</td>
<td>5.56</td>
<td>8.02</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>5.90</td>
<td>8.94</td>
<td>5.22</td>
<td>8.72</td>
<td>5.56</td>
<td>8.02</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>5.90</td>
<td>8.94</td>
<td>5.22</td>
<td>8.72</td>
<td>5.56</td>
<td>8.02</td>
</tr>
<tr>
<td>$(2,4)$</td>
<td>0.5</td>
<td>0.5</td>
<td>29.84</td>
<td>23.42</td>
<td>30.88</td>
<td>22.88</td>
<td>33.00</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>5.34</td>
<td>12.98</td>
<td>5.90</td>
<td>12.84</td>
<td>5.96</td>
<td>11.40</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>5.34</td>
<td>12.98</td>
<td>5.90</td>
<td>12.84</td>
<td>5.96</td>
<td>11.40</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>5.34</td>
<td>12.98</td>
<td>5.90</td>
<td>12.84</td>
<td>5.96</td>
<td>11.40</td>
</tr>
<tr>
<td>$(4,2)$</td>
<td>0.5</td>
<td>0.5</td>
<td>35.04</td>
<td>27.16</td>
<td>33.48</td>
<td>26.20</td>
<td>34.86</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>5.82</td>
<td>8.30</td>
<td>5.82</td>
<td>8.20</td>
<td>6.04</td>
<td>7.76</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>5.82</td>
<td>8.30</td>
<td>5.82</td>
<td>8.20</td>
<td>6.04</td>
<td>7.76</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>5.82</td>
<td>8.30</td>
<td>5.82</td>
<td>8.20</td>
<td>6.04</td>
<td>7.76</td>
</tr>
<tr>
<td>$(4,4)$</td>
<td>0.5</td>
<td>0.5</td>
<td>27.42</td>
<td>21.10</td>
<td>28.06</td>
<td>19.30</td>
<td>25.84</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>3.54</td>
<td>10.04</td>
<td>4.12</td>
<td>9.50</td>
<td>4.60</td>
<td>7.94</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>3.54</td>
<td>10.04</td>
<td>4.12</td>
<td>9.50</td>
<td>4.60</td>
<td>7.94</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>3.54</td>
<td>10.04</td>
<td>4.12</td>
<td>9.50</td>
<td>4.60</td>
<td>7.94</td>
</tr>
</tbody>
</table>

Note: The $J$ test is compared to size-corrected critical values. Nominal level is 5 per cent. Results based on 5,000 replications.