Econometrica, Vol. 60, No. 1 (January, 1992), 145–157

A NEW FORM OF THE INFORMATION MATRIX TEST

BY RUSSELL DAVIDSON AND JAMES G. MACKINNON[1]

We develop a new form of the information matrix test for a wide variety of statistical models. Chesher (1984) showed that the implicit alternative of this test is a model with random parameter variation, a fact which we exploit by constructing the test against an explicit alternative of this type. The new test is computed using a double-length artificial regression, instead of the more conventional outer-product-of-the-gradient regression, which, although easy to use, is known to give test statistics with distributions very far from the asymptotic nominal distribution even in rather large samples. The new form on the other hand performs remarkably well, at least for the case of univariate regression models. Some approximate finite-sample distributions are calculated for this case, and lend support to the use of the new form of the test.

KEYWORDS: Information matrix test, specification test, double-length regression, artificial regression, stochastic expansion, Monte Carlo experiments.

1. INTRODUCTION

THE ORIGINAL FORM of White's (1982) information matrix (IM) test was not particularly easy to compute. Later, Chesher (1983) and Lancaster (1984) showed how to compute an asymptotically equivalent version by means of a certain artificial regression,[2] a variant of the outer-product-of-the-gradient (OPG) regression popularized by Godfrey and Wickens (1981) as a way to calculate Lagrange multiplier (LM) tests. Unfortunately, all available evidence (Davidson and MacKinnon (1984b, 1985), Bera and McKenzie (1986), Godfrey, McAleer, and McKenzie (1988)) suggests that tests based on the OPG regression are excessively prone to reject true null hypotheses in finite samples. This seems to be particularly true of the OPG form of the IM test, for which extremely poor finite-sample performance has recently been reported by Taylor (1987), Kennan and Neumann (1988), Orme (1990a, 1990b), and Chesher and Spady (1991).

In this paper, we develop a new form of the IM test based on an artificial double-length regression (DLR) proposed by Davidson and MacKinnon (1984a). Although the new test is somewhat less generally applicable than the OPG variant, it does not share the poor finite-sample properties of the latter. Several other variants of the IM test have recently been proposed. Chesher and Spady (1991) propose a variant based on corrections derived from Edgeworth-style expansions. Orme (1990a) proposes a variant that is more widely applicable than the one we propose in this paper, but is also harder to implement. Orme

[1] This research was supported, in part, by grants from the Social Sciences and Humanities Research Council of Canada. We discussed an earlier version of the paper with Andrew Chesher, and have also benefited from comments by three referees and a co-editor, and discussions with seminar participants at Oxford University, the Econometric Study Group, and INSEE (Paris).

[2] Chesher called it a "pseudo regression." Since it has the properties of an artificial regression, as defined in Davidson and MacKinnon (1990), we prefer the latter terminology.

(1990b) suggests using double- and even triple-length regressions to compute an efficient-score version of the test for a more restricted set of models than that considered here.

In Section 2 we present both the OPG and DLR forms of the IM test. Section 3 contains Monte Carlo evidence confirming the poor performance of OPG and suggesting that DLR performs very much better, especially for models with more than a very small number of parameters. The efficient score form of the test, which is readily available in the simple case dealt with, is also studied, as a benchmark for the other two. In Section 4, we perform stochastic expansions which reveal some of the sources of the differences in finite-sample behavior of all three forms of the test.

2. REGRESSION-BASED FORMS OF THE IM TEST

Consider a statistical model characterized by a loglikelihood function of the following form:

$$l(y,\boldsymbol{\theta}) = \sum_{t=1}^{n} l_t(y_t,\boldsymbol{\theta}) \tag{1}$$

where y denotes an $n \times 1$ vector of observations $y_t, t = 1,\ldots,n$, on a dependent variable, and $\boldsymbol{\theta}$ is a k-vector of parameters. The contribution from observation t to the loglikelihood function, l_t, depends in general on exogenous or predetermined variables that vary across the n observations.

The OPG form of the IM test requires one to compute the explained sum of squares from an artificial regression in which the regressand is a vector of ones. For each component $\theta_i, i = 1,\ldots,k$, of the parameter vector $\boldsymbol{\theta}$, there is a regressor with typical element

$$\tilde{l}_{ti} \equiv \frac{\partial l_t(\tilde{\boldsymbol{\theta}})}{\partial \theta_i}, \quad \text{for} \quad t = 1,\ldots,n, \qquad i = 1,\ldots,k,$$

where $\boldsymbol{\theta}$ is the maximum likelihood estimate from (1). In addition, there are at most $\frac{1}{2}k(k+1)$ test regressors, fewer if the model structure leads to exact collinearities, with typical elements given by

$$\tilde{l}^*_{tij} + \tilde{l}_{ti}\tilde{l}_{tj}, \quad i = 1,\ldots,k, \quad j = 1,\ldots,i, \tag{2}$$

where $\tilde{l}^*_{tij} \equiv (\partial^2 l_t/\partial\theta_i\,\partial\theta_j)(\tilde{\boldsymbol{\theta}})$. The test statistic is then asymptotically distributed as χ^2 with as many degrees of freedom as there are test regressors.

Models for which a double-length regression can be used for testing purposes must take the following form:

$$f_t(y_t,\boldsymbol{\theta}) = \varepsilon_t, \quad t = 1,\ldots,n, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0},\mathbf{I}_n). \tag{3}$$

The requirement that ε_t be normally distributed may seem restrictive. However, at least for regression models, the IM test tests for nonnormality of the errors, so that the test would have very limited interest in a more robust context. Since the functions f are arbitrary, any continuous distribution may be transformed to satisfy (3). However, models with discrete, truncated, or censored data cannot be dealt with in this framework, and this is certainly a significant limitation.

For the class of models (3), the contribution to the loglikelihood from observation t is

$$l_t = -\tfrac{1}{2}\log(2\pi) - \tfrac{1}{2}f_t^2 + k_t,$$

where $k_t \equiv \log|f_t'|$ is the Jacobian contribution to l_t. We may write typical elements of the gradient and Hessian of l_t as follows:

$$(4) \qquad l_{ti} = -f_t F_{ti} + K_{ti}, \quad l_{tij}^* = -F_{ti}F_{tj} - f_t F_{tij}^* + K_{tij}^*,$$

where $F_{ti} \equiv \partial f_t/\partial\theta_i$ and $K_{ti} \equiv \partial k_t/\partial\theta_i$, and the starred quantities with three subscripts are defined just like l_{tij}^*. From (2) and (4) we see that a typical direction tested by the IM test applied to model (3) is given by

$$(5) \qquad -F_{ti}F_{tj} - f_t F_{tij}^* + K_{tij}^* + f_t^2 F_{ti}F_{tj} - f_t(F_{ti}K_{tj} + K_{ti}F_{tj}) + K_{ti}K_{tj}.$$

Davidson and MacKinnon (1984a) (see also Davidson and MacKinnon (1988)) propose a double-length regression to perform tests on model (3). It has $2n$ "observations." The regressand is $\tilde{f}_t$ for "observation" t and one for "observation" $t+n$, and the regressor corresponding to parameter θ_i is $-\tilde{F}_{ti}$ for "observation" t and $\tilde{K}_{ti}$ for "observation" $t+n$, where the tildes imply function evaluations at restricted ML estimates $\tilde{\boldsymbol{\theta}}$. The test statistic is then the explained sum of squares from this artificial regression, exactly as for the OPG test regression.

It is clear from (2) that the IM test is testing whether

$$\sum_{t=1}^{n} \left(l_{tij}^*(\tilde{\boldsymbol{\theta}}) + l_{ti}(\tilde{\boldsymbol{\theta}}) l_{tj}(\tilde{\boldsymbol{\theta}}) \right)$$

is significantly different from zero, that is, whether the information matrix equality is significantly violated in the sample. But Chesher (1984) showed that the test has another interpretation, namely as a test for random heterogeneity of the parameters $\boldsymbol{\theta}$. In order to obtain the DLR form of the IM test, an explicit alternative hypothesis in the form (3) is needed. Chesher's result suggests the use of the following model:

$$(6) \qquad f_t(y_t, \boldsymbol{\theta} + \boldsymbol{\eta}_t) = \varepsilon_t, \quad \varepsilon_t \sim N(0,1), \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, 2\boldsymbol{\Omega}),$$

with the k-dimensional random vector $\boldsymbol{\eta}_t$ being distributed independently of ε_t and of $\boldsymbol{\eta}_s$, $s \neq t$. We can expand (6) in a Taylor expansion to second order in $\boldsymbol{\eta}_t$:

$$(7) \qquad f_t(y_t, \boldsymbol{\theta}) \cong \varepsilon_t - \boldsymbol{F}_t\boldsymbol{\eta}_t - \tfrac{1}{2}\boldsymbol{\eta}_t^T \boldsymbol{F}_t^* \boldsymbol{\eta}_t,$$

where $\boldsymbol{F}_t$ is a row vector with typical element F_{ti} and $\boldsymbol{F}_t^*$ is a $k \times k$ matrix with typical element F_{tij}^*. The expectation of the right-hand side of (7) is $-\operatorname{tr}(\boldsymbol{\Omega}\boldsymbol{F}_t^*)$, and the variance, to first order in $\boldsymbol{\Omega}$, is $2\operatorname{tr}(\boldsymbol{\Omega}\boldsymbol{F}_t^T\boldsymbol{F}_t)$. Thus, locally in the neighborhood of $\mathbf{\Omega}=\mathbf{0}$, model (6) is equivalent to

$$(8) \qquad q_t(y_t,\boldsymbol{\theta},\mathbf{\Omega}) \equiv \frac{f_t(y_t,\boldsymbol{\theta}) + \operatorname{tr}(\mathbf{\Omega}\boldsymbol{F}_t^*)}{\left(1+2\operatorname{tr}(\mathbf{\Omega}\boldsymbol{F}_t^T\boldsymbol{F}_t)\right)^{1/2}} = \boldsymbol{\nu}_t,$$

for some new vector of disturbances $\boldsymbol{\nu}$, distributed approximately as $N(\mathbf{0},\mathbf{I}_n)$.

It may seem that a model like (8) is an inappropriate one on which to base a test, because the null hypothesis (i.e., $\mathbf{\Omega}=\mathbf{0}$) is on the boundary of the set of permissible values for a covariance matrix, and because the specification of parameter heterogeneity may appear to cause inadmissible parameter values, like a negative variance. In fact these are not problems, since the analysis is purely local: alternative (8) need not be a well-specified model for Lagrange multiplier tests based on it to be valid. See Chant (1974) on testing hypotheses on the boundary of the parameter space, and Godfrey (1981) on locally equivalent alternative hypotheses.

The function $q_t(\cdot)$ defined by (8) is now the function from which the double-length regression to test the hypothesis $\mathbf{\Omega}=\mathbf{0}$ will be formed. This requires us to find the analogues of F_{ti} and K_{ti} for the function (8), evaluated at $\mathbf{\Omega}=\mathbf{0}$. The analogue of F_{ti} is clearly

$$Q_{tij} \equiv \frac{\partial q_t}{\partial \Omega_{ij}} = F^*_{tij} - f_t F_{ti} F_{tj}, \quad i=1,\ldots,k, \quad j=1,\ldots,i,$$

while the analogue of K_{ti} can be shown to be

$$R_{tij} \equiv \frac{\partial \log(q_t')}{\partial \Omega_{ij}} = K^*_{tij} + K_{ti}K_{tj} - F_{ti}F_{tj} - f_t(K_{ti}F_{tj} + F_{ti}K_{tj}).$$

It is now easy to verify from (4) that the score vectors associated with the directions $\mathbf{\Omega}$, namely $-q_t Q_{tij} + R_{tij}$, are identical to the directions (5) tested by the IM test.

3. FINITE-SAMPLE PERFORMANCE OF THE TESTS

In this section we describe a fairly simple Monte Carlo experiment designed to study the finite-sample performance of the OPG and DLR forms of the IM test for linear regression models. We considered models of the form

$$y_t = \beta_1 + \sum_{i=2}^{k} \beta_i X_{ti} + u_t, \quad u_t \sim N(0,\sigma^2),$$

where k was 2, 3, 5, 6, 8, or 10. The X_{ti}'s were normal random variates, independent across observations, and equicorrelated with correlation coefficient one half. It is easy to see that all forms of the IM test studied are invariant to both $\boldsymbol{\beta}$ and σ, and so those parameters were chosen arbitrarily. We studied the behavior of the tests under the null hypothesis only; the OPG form rejected so often under the null that any meaningful comparison of power would have been impossible.

For linear regression models it is quite easy to compute the so-called efficient-score (ES) form of the IM test. In this form of the test the covariance matrix that is used is the exact information matrix associated with the alternative hypothesis, evaluated at the restricted parameter estimates, which here are simply the OLS estimates of the linear regression. It can be shown that the ES test statistic is the sum of three separate statistics, which test for heteroskedasticity, skewness, and kurtosis respectively; see Hall (1987). These three test statistics may be written as

$$\frac{1}{2}h_2^T Z(Z^T Z)^{-1} Z^T h_2, \quad \frac{1}{6}h_3^T X(X^T X)^{-1} X^T h_3, \quad \text{and} \tag{9}$$

$$\frac{1}{24n}\left(\sum_{t=1}^{n}(e_t^4-3)\right)^2.$$

Here e_t denotes the standardized residual $\tilde{u}_t/\tilde{\sigma}$, h_2 and h_3 denote n-vectors with typical elements (e_t^2-1) and e_t^3 respectively, and Z denotes a matrix that consists of all the nonredundant squares and cross-products of the columns of X.

Results for samples of 100, 200, 500, and 1000, based on 5000 replications, are presented in Table I. All the tests display a tendency to over-reject which grows with the number of regressors (and hence the number of degrees of freedom for the test), and (eventually) decreases with the sample size, n. The OPG test, as expected, always rejects the null far too often. At the nominal 5% level it never rejects less than 16% of the time (for $k=2$ and $n=1000$), and in the worst case ($k=10$, $n=200$) it rejects an amazing 99.9% of the time! The improvement in the performance of OPG as n increases is quite slow, so that its distribution is never acceptably close to its asymptotic distribution even when $n=1000$, and for the larger values of k it is very far away indeed.

The increasing tendency to over-reject as k rises is an important and rather alarming result, since in applied work the number of parameters is often much greater than ten. The fact that we considered relatively large values of k may in part explain the fact that OPG tests performed even less well in our experiments than other authors have reported. The substantial deterioration as k increases suggests that, in practice, the OPG variant of the IM test will often perform even worse than it did in the table.

Moderately large numbers of degrees of freedom did not prevent the DLR form from performing remarkably well. For $k \leqslant 4$, rejection frequencies were always reasonably close to the nominal levels even for samples of size 100. For $k=6$, DLR did over-reject noticeably for $n=100$, but not for the larger sample sizes. Only for the rather large number of degrees of freedom associated with $k=8$ and $k=10$, did the DLR test over-reject noticeably for samples of 200 or more.

Because the efficient-score test uses the least noisy covariance matrix estimate possible, it might have been expected to outperform both the others. In fact, it is not uniformly superior to the DLR test. It appears that ES copes better than

TABLE I

PERFORMANCE OF OPG AND DLR TESTS UNDER THE NULL

						Rejection Frequencies at Nominal Levels		
k	d.f.	n	Test	Mean	S.D.	10%	5%	1%
2	5	100	OPG	13.43^{++}	9.74^{++}	58.4^{++}	49.3^{++}	34.1^{++}
			DLR	4.91^{-}	2.95^{--}	8.7^{-}	3.9^{--}	0.5^{-}
			ES	4.60^{--}	4.59^{++}	8.2^{--}	5.5	2.5^{++}
		200	OPG	10.54^{++}	8.40^{++}	44.5^{++}	36.0^{++}	22.1^{++}
			DLR	4.89^{-}	3.01^{--}	8.6^{-}	4.0^{--}	0.7^{-}
			ES	4.89	4.56^{++}	9.2	6.0^{+}	2.6^{++}
		500	OPG	8.13^{++}	6.53^{++}	31.7^{++}	23.1^{++}	11.9^{++}
			DLR	5.00	3.10	9.7	4.5	0.8
			ES	5.06	3.92^{++}	10.4	6.5^{++}	2.5^{++}
		1000	OPG	6.86^{++}	5.44^{++}	23.2^{++}	16.3^{++}	7.7^{++}
			DLR	4.97	3.17	10.0	4.8	0.9
			ES	4.98	3.49^{++}	10.2	6.0^{+}	1.7^{++}
3	9	100	OPG	23.55^{++}	11.36^{++}	76.4^{++}	68.2^{++}	50.7^{++}
			DLR	9.08	3.98^{--}	9.3	4.2^{-}	0.7^{-}
			ES	8.35^{--}	6.34^{++}	9.8	6.5^{++}	3.0^{++}
		200	OPG	19.65^{++}	11.28^{++}	60.8^{++}	51.2^{++}	34.2^{++}
			DLR	8.80^{--}	3.99^{--}	8.5^{--}	3.5^{--}	0.6^{-}
			ES	8.54^{--}	5.75^{++}	9.6	6.2^{++}	2.9^{++}
		500	OPG	15.40^{++}	9.27^{++}	43.7^{++}	34.3^{++}	20.1^{++}
			DLR	8.98	4.08^{-}	9.5	4.3^{-}	1.0
			ES	8.93	4.92^{++}	10.7	6.4^{++}	2.4^{++}
		1000	OPG	12.92^{++}	7.51^{++}	32.1^{++}	23.6^{++}	11.8^{++}
			DLR	9.05	4.15^{-}	9.8	4.5	0.7^{-}
			ES	9.06	4.70^{++}	10.8	6.1^{++}	1.8^{++}
4	14	100	OPG	35.19^{++}	12.11^{++}	89.1^{++}	82.0^{++}	65.7^{++}
			DLR	14.50^{++}	5.08^{--}	10.6	5.0	0.9
			ES	13.16^{--}	8.32^{++}	11.0^{+}	7.8^{++}	4.4^{++}
		200	OPG	31.09^{++}	13.30^{++}	76.2^{++}	66.9^{++}	49.5^{++}
			DLR	13.88	5.06^{--}	8.9^{-}	4.5	0.7^{-}
			ES	13.35^{--}	7.59^{++}	10.9^{+}	7.7^{++}	3.6^{++}
		500	OPG	24.08^{++}	11.40^{++}	53.6^{++}	43.9^{++}	26.8^{++}
			DLR	13.73^{--}	5.12^{-}	8.5^{--}	4.4	0.7
			ES	13.69^{--}	6.50^{++}	11.0^{+}	7.0^{++}	2.6^{++}
		1000	OPG	20.52^{++}	9.62^{++}	39.6^{++}	29.9^{++}	16.2^{++}
			DLR	13.97	5.34	10.0	5.0	1.2
			ES	14.04	6.23^{++}	11.3^{+}	6.9^{++}	2.8^{++}
6	27	100	OPG	59.99^{++}	10.85^{++}	98.4^{++}	96.4^{++}	87.3^{++}
			DLR	29.76^{++}	7.36^{--}	17.1^{++}	8.8^{++}	1.7^{++}
			ES	25.19^{--}	12.33^{++}	12.1^{++}	8.8^{++}	4.9^{++}
		200	OPG	61.61^{++}	16.49^{++}	95.8^{++}	91.9^{++}	80.3^{++}
			DLR	28.10^{++}	7.09^{--}	11.8^{++}	5.5	0.9
			ES	26.06^{--}	11.31^{++}	11.9^{++}	8.5^{++}	4.7^{++}
		500	OPG	51.07^{++}	16.69^{++}	80.6^{++}	72.5^{++}	54.5^{++}
			DLR	27.10	7.22	10.0	4.7	1.0
			ES	26.63^{-}	9.85^{++}	12.7^{++}	7.7^{++}	3.3^{++}
		1000	OPG	42.81^{++}	14.34^{++}	62.6^{++}	52.3^{++}	33.8^{++}
			DLR	26.94	7.22	9.5	4.5	1.0
			ES	26.85	8.71^{++}	11.8^{++}	7.3^{++}	2.7^{++}
8	44	100	OPG	80.81^{++}	6.96^{--}	99.9^{++}	99.3^{++}	94.4^{++}
			DLR	52.13^{++}	9.35	31.4^{++}	18.7^{++}	4.6^{++}
			ES	41.53^{--}	16.50^{++}	13.3^{++}	9.8^{++}	5.6^{++}
		200	OPG	96.27^{++}	17.13^{++}	99.4^{++}	98.6^{++}	95.4^{++}
			DLR	47.55^{++}	9.37	17.7^{++}	9.2^{++}	1.8^{++}
			ES	42.60^{--}	14.65^{++}	14.0^{++}	9.6^{++}	5.4^{++}

TABLE I—Continued.

k	d.f.	n	Test	Mean	S.D.	Rejection Frequencies at Nominal Levels 10%	5%	1%
		500	OPG	88.05^{++}	20.88^{++}	95.6^{++}	92.4^{++}	82.1^{++}
			DLR	44.95^{++}	9.17^{-}	11.4^{++}	5.7^{+}	1.1
			ES	43.46^{-}	12.65^{++}	13.6^{++}	9.3^{++}	3.9^{++}
		1000	OPG	74.41^{++}	19.20^{++}	82.9^{++}	76.1^{++}	58.0^{++}
			DLR	44.06	9.20^{-}	9.8	4.5	0.9
			ES	43.58^{-}	11.26^{++}	11.6^{++}	7.4^{++}	2.7^{++}
10	65	100	OPG	94.25^{++}	2.80^{--}	99.98^{++}	99.3^{++}	55.4^{++}
			DLR	83.45^{++}	10.70^{--}	63.1^{++}	44.5^{++}	15.7^{++}
			ES	61.30^{--}	21.98^{++}	13.2^{++}	10.2^{++}	5.9^{++}
		200	OPG	129.69^{++}	14.68^{++}	99.96^{++}	99.9^{++}	99.2^{++}
			DLR	72.98^{++}	11.50	26.6^{++}	15.3^{++}	4.0^{++}
			ES	63.06^{--}	19.69^{++}	14.4^{++}	10.9^{++}	6.3^{++}
		500	OPG	134.73^{++}	24.78^{++}	99.7^{++}	99.1^{++}	96.3^{++}
			DLR	67.68^{++}	11.27	14.0^{++}	7.4^{++}	1.5^{++}
			ES	64.12^{--}	16.01^{++}	13.8^{++}	9.5^{++}	4.8^{++}
		1000	OPG	117.41^{++}	23.96^{++}	95.8^{++}	92.7^{++}	83.2^{++}
			DLR	65.82^{++}	11.09^{-}	10.3	5.1	0.9
			ES	64.48^{-}	14.28^{++}	13.0^{++}	8.0^{++}	3.5^{++}

Notes: All results are based on 5000 replications.

$^{+}$ and $^{-}$ indicate that a quantity is significantly larger or smaller than it should be asymptotically at the .05 level, on a two-tail test.

$^{++}$ and $^{--}$ indicate that a quantity is significantly larger or smaller than it should be asymptotically at the .001 level, on a two-tail test.

DLR with large numbers of degrees of freedom in small samples, but that the improvement of DLR with sample size is more rapid. For $k = 10$ for instance, ES is much better than DLR for $n = 100$, but worse for $n = 1000$. In fact, DLR shows no sign of yielding faulty inference for $n = 1000$ for any of the sizes of test considered. Nevertheless, it should clearly be used with caution if the number of degrees of freedom is large relative to the sample size.

Table I indicates that the OPG form of the IM test is so far from its asymptotic distribution in finite samples that it should probably never be used unless the sample size is truly enormous. In contrast, the DLR form can probably be used with confidence in samples of only a few hundred, provided the number of degrees of freedom is not very large. Even in circumstances in which the ES test can be easily computed, as here, it does not seem to be uniformly more reliable than DLR. These results are of course subject to the usual qualifications which attend any evidence from sampling experiments. But although it is quite possible that a different experimental design might have made the DLR test perform less well, it seems very unlikely that the OPG test would have performed acceptably in any experiment.

4. ASYMPTOTIC EXPANSIONS OF SOME TEST STATISTICS

The Monte Carlo results of the previous section are suggestive, but provide little insight into the reasons for the very different finite-sample performance of the different versions of the IM test. In this section, we shall attempt to provide such insight by use of stochastic expansions. We consider a very simple model, and limit our attention to one of the directions of the IM test, namely the kurtosis direction. This limitation makes it possible to consider the test statistic in a convenient form distributed asymptotically under the null as $N(0, 1)$. We deal with five different tests for kurtosis: two based on the OPG, two based on the DLR, and the efficient-score ES test.

The model we treat has a regression function that is simply a constant:

$$y_t = \mu + u_t, \quad u_t \sim N(0, \sigma^2), \quad E(u_t u_s) = 0, \quad t \neq s. \tag{10}$$

Chesher and Spady (1991) have shown that the ES form of the kurtosis test is in fact invariant to order n^{-1} to changes in the regression function, although the other forms we consider are unlikely to share this property. The efficient-score form of the test for kurtosis of the errors in the model (10) is based on the standardized residuals after estimation by maximum likelihood:

$$\tilde{f}_t \equiv f_t(y_t, \tilde{\mu}, \tilde{\sigma}) \equiv \frac{y_t - \tilde{\mu}}{\tilde{\sigma}}.$$

The actual ES test statistic (see (9)) is then

$$(24n)^{-1/2} \sum_{t=1}^{n} (\tilde{f}_t^4 - 3). \tag{11}$$

For the purposes of the DLR test, we rewrite model (10) as

$$f_t(y_t, \mu, \sigma) \equiv \frac{y_t - \mu}{\sigma} = \varepsilon_t, \quad \text{with} \qquad \varepsilon_t \sim N(0, 1).$$

In the notation of Section 2, we find that $k_t = \log|f_t'| = -\log\sigma$, and so the DLR for testing in the kurtosis direction can be expressed as follows:

$$\begin{aligned}
&\text{regressand:} && (\tilde{f}_t, 1);\\
&\text{regressors:} && (1, 0) \text{ for } \mu,\\
& && (\tilde{f}_t, -1) \text{ for } \sigma, \quad .\\
(12)\quad & && (\tilde{f}_t^3 - 3\tilde{f}_t, 3\tilde{f}_t^2 - 3) \text{ for the test direction.}
\end{aligned}$$

The test regressor (12) evidently tests in the direction corresponding to kurtosis of the error terms ε_t. However, if one simply derives the artificial regression for the IM test in the usual way, (12) is not one of the regressors. Instead, it is obtained by subtracting the regressor corresponding to σ from the regressor corresponding to $\sigma \times \sigma$. The result is that, under the null hypothesis of correct specification (with normal errors), the direction (12) is asymptotically orthogonal to the directions of all the regressors in the test regression. There are thus two possible DLR test statistics: one, which we call DLR1, where all regressors are retained, and a second, DLR2, in which the only regressor is the test regressor.

TABLE II

THE TABLE LISTS, TO ORDER n^{-1}, THE PERTURBATIONS OF THE CUMULANTS OF THE STATISTICS CONSIDERED AWAY FROM THOSE OF THE STANDARD NORMAL DISTRIBUTION, FOR WHICH THE ONLY NONVANISHING CUMULANT IS THE SECOND, EQUAL TO UNITY. ONLY THE FIRST FOUR CUMULANTS ARE AFFECTED TO THE ORDER GIVEN.

	Order of Cumulant			
Statistic[a]	1st (mean)	2nd (variance)	3rd (skewness)	4th (kurtosis)
$h_4/\sqrt{24}$	0	0	$3\sqrt{24}\,n^{-1/2}$	$636n^{-1}$
ES	$-\frac{1}{4}\sqrt{24}\,n^{-1/2}$	$-15n^{-1}$	$3\sqrt{24}\,n^{-1/2}$	$540n^{-1}$
DLR1	$-\frac{5}{8}\sqrt{24}\,n^{-1/2}$	$-\frac{319}{8}n^{-1}$	$\frac{3}{4}\sqrt{24}\,n^{-1/2}$	$\frac{261}{4}n^{-1}$
DLR2	$-\frac{5}{8}\sqrt{24}\,n^{-1/2}$	$-\frac{333}{8}n^{-1}$	$\frac{3}{4}\sqrt{24}\,n^{-1/2}$	$\frac{237}{4}n^{-1}$
OPG1	$-\frac{7}{4}\sqrt{24}\,n^{-1/2}$	$519n^{-1}$	$-6\sqrt{24}\,n^{-1/2}$	$1698n^{-1}$
OPG2	$-\frac{7}{4}\sqrt{24}\,n^{-1/2}$	$399n^{-1}$	$-6\sqrt{24}\,n^{-1/2}$	$1314n^{-1}$

[a] ES is the efficient score test statistic (11), DLR1 and DLR2 are the statistics from the DLR artificial regression including or excluding respectively the regressors other than the test column, and similarly for OPG1 and OPG2.

The OPG artificial regression can be characterized as follows:

regressand: 1;

regressors: $\tilde{f}_t$ for μ; $\tilde{f}_t^2 - 1$ for σ^2; $\tilde{f}_t^4 - 6\tilde{f}_t^2 + 3$ for the test column.

Again there are two forms of the test statistic, OPG1 and OPG2, in which the regressors other than the test regressor are or are not retained in the test regression.

In the Appendix, some details are given of the stochastic expansions used in order to derive the results presented in Table II. This table shows, to order n^{-1}, the perturbations of the cumulants of the five different test statistics away from those of the standard normal distribution. In addition, for reference purposes, these perturbations are shown for a random variable denoted as $h_4/\sqrt{24}$, which is given by formula (11) using the actual errors f_t instead of the residuals $\tilde{f}_t$. This random variable, unobservable in practice, has precisely zero mean and unit variance. It can be seen that the perturbations are nonzero only for the first four cumulants, which correspond respectively to bias, variance, skewness, and kurtosis.

The random variable $h_4/\sqrt{24}$, although standardized to have zero mean and unit variance, is normal only asymptotically; in fact it has skewness of $3\sqrt{24}\,n^{-1/2}$ and kurtosis of $636n^{-1}$. Comparison with ES shows that the use of residuals rather than the real "errors" induces no change in the skewness, and actually lowers the excess kurtosis. In contrast, the bias and reduction of the variance evident for ES are directly attributable to the use of residuals.

The bias of DLR1 is slightly greater than that of ES, but the skewness and kurtosis are smaller. The variance is still further below one. Note that ES is *not* necessarily superior to DLR1, which corrects more of the nonnormality in $h_4/\sqrt{24}$. The presence of bias does not alter the fact that the uncentered second moment of DLR1 is still less than unity: it is in fact $1-(61/2)n^{-1}$.

DLR2, calculated with only the test column in the artificial regression, is very similar to DLR1. There is no difference to order $n^{-1/2}$, and there are further reductions to the even-order cumulants. In terms of size properties, there seems to be little reason to prefer one version of the statistic over the other. Power, on the other hand, may well be substantially reduced by using the numerically smaller statistic if the null hypothesis of correct model specification is not true.

The two forms of test statistic calculated from the OPG regression differ only at order n^{-1}. Everything is much worse than for the DLR test. The bias is greater; the variance is substantially greater, not smaller, than one; the skewness is increased by a factor of eight times and is now of the same sign as the bias; and the excess kurtosis is massive. All of this is due to the use of the evidently inefficient OPG estimate of the information matrix, which fails to use information about the model that DLR and ES exploit. However, although ES exploits more of this information than DLR, it does not always yield better inference.

5. CONCLUSION

We have developed a new way to compute the information matrix test for a wide variety of statistical and econometric models. The new form of the test is based on a double-length artificial regression that can be used in order to compute an LM test against an *explicit* alternative of random parameter variation. Unlike the familiar OPG form of the IM test, the new DLR form performs reasonably well, better even than an efficient score test in some cases.

The complexity of the calculations limited the scope of our study based on stochastic expansions. Even so, it serves to make clear the features of the tests which are responsible for some of the conclusions of the Monte Carlo experiments. The chief cause of the remarkably good behavior of the DLR test seems to be the substantial variance reduction along with relatively small excess kurtosis. This more than offsets bias and skewness, and leads the test to under-reject somewhat. But the Monte Carlo evidence strongly suggests that the trade-off goes the other way with more degrees of freedom. In the context of one-degree-of-freedom tests, it was not possible to explain why an increase in the number of model parameters leads to deterioration in the behavior of all the forms of IM tests we studied, especially OPG. This will be the object of further research.

The DLR form of the IM test, while greatly superior to the OPG form, is not perfect. Investigators may prefer to employ individual tests for heteroskedasticity, skewness, kurtosis, and so on, rather than a single IM test. Such tests, with fewer degrees of freedom, have more power than an IM test against the specific alternatives they are designed to test against, and may be more informative about what is wrong with the model.

G.R.E.Q.E.-E.H.E.S.S., Centre de la Vieille Charité, 2 rue de la Charité, 13002 Marseille, France

and

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

Manuscript received July 1989; final revision received May, 1991.

APPENDIX

The maximum likelihood estimates for model (10) can be written in terms of the error process $\{f_t\}_{t=1}^n$, assumed distributed as $N(0, \mathbf{I}_n)$, as follows:

$$\text{(A1)} \qquad \bar{\mu} = \frac{1}{n}\sum_{t=1}^{n} y_t = \mu + \frac{\sigma}{n}\sum_{t=1}^{n} f_t$$

and

$$\text{(A2)} \qquad \bar{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{\mu})^2 = \frac{\sigma^2}{n}\sum_{t=1}^{n}\left(f_t - \frac{1}{n}\sum_{s=1}^{n} f_s\right)^2.$$

Introduce the convenient notation:

$$h_i \equiv n^{-1/2}\sum_{t=1}^{n} He_i(f_t),$$

where $He_i(\cdot)$ denotes the Hermite polynomial of degree i; see, for example, Abramowitz and Stegun (1965). Standard properties of these polynomials lead to the results:

$$E(h_i) = 0; \quad \mathrm{Var}(h_i) = i!; \quad \mathrm{cov}(h_i, h_j) = 0, \quad i \neq j$$

The h_i are evidently asymptotically normal. The standardized residuals $\bar{f}_t$ of (10), by use of (A1) and (A2), are, up to order $n^{-3/2}$:

$$\bar{f}_t = f_t - n^{-1/2}\left(h_1 + \tfrac{1}{2}h_2 f_t\right) + \tfrac{1}{8}n^{-1}\left(4h_1^2 f_t + 3h_2^2 f_t + 4h_1 h_2\right)$$
$$-(1/16)n^{-3/2}\left(8h_1^3 + 6h_1 h_2^2 + 12h_1^2 h_2 f_t + 5h_2^3 f_t\right).$$

From this we find that the numerator of all the test statistics considered is

$$n^{-1/2}\sum_{t=1}^{n}\left(\bar{f}_t^4 - 3\right) = h_4 - n^{-1/2}\left(4h_1 h_3 + 2h_2 h_4 + 3h_2^2\right)$$
$$+ n^{-1}\left(2h_1^2 h_4 + 12h_1^2 h_2 + 3h_2^2 h_4 + 6h_2^3 + 8h_1 h_2 h_3\right) + O(n^{-3/2})$$

Our analysis is carried out in terms of the cumulant-generating functions of the statistics studied For the standardized random variable $h_4/\sqrt{24}$, we obtain

$$\begin{aligned}
\log E\left(\exp\left(ith_4/\sqrt{24}\right)\right) &= \log E\left(\exp\left(it(24n)^{-1/2}\sum_{t=1}^{n} He_4(x_t)\right)\right)\\
&= n\log E\Big(1+it(24n)^{-1/2}He_4(x)-\frac{t^2}{2}(24n)^{-1}(He_4(x))^2\\
&\qquad -\frac{it^3}{6}(24n)^{-3/2}(He_4(x))^3\\
&\qquad +\frac{t^4}{24}(24n)^{-2}(He_4(x))^4+O(n^{-5/2})\Big)\\
&= -\tfrac{1}{2}t^2-\tfrac{1}{2}\sqrt{24}\,n^{-1/2}it^3+(53/2)n^{-1}t^4+O(n^{-3/2}).
\end{aligned}$$

The above result establishes the first row of Table II, which shows, to order n^{-1}, the perturbations of the cumulants of $h_4/\sqrt{24}$ away from those of the standard normal distribution. (Recall that the cumulant-generating function is so called because, in its Taylor expansion in powers of t, the coefficient of $(it)^k/k!$ is the kth cumulant of the distribution.)

Subsequent rows of Table II show these perturbations for the statistics in which we are actually interested. These can all be expanded to take the form:

$$h_4/\sqrt{24}+n^{-1/2}A+n^{-1}B,$$

with leading term $h_4/\sqrt{24}$, and their characteristic functions can therefore be expanded around that of $h_4/\sqrt{24}$, which we denote as $\phi(t)$, as follows:

$$\begin{aligned}
\text{(A3)}\qquad \log\phi(t)+(\phi(t))^{-1}\Big(& itn^{-1/2}E\left(A\exp\left(ith_4/\sqrt{24}\right)\right)+itn^{-1}E\left(B\exp\left(ith_4\sqrt{24}\right)\right)\\
&-\tfrac{1}{2}t^2n^{-1}\Big(E\left(A^2\exp\left(ith_4\sqrt{24}\right)\right)\\
&\qquad+(\phi(t))^{-1}\left(E\left(A\exp\left(ith_4\sqrt{24}\right)\right)\right)^2\Big)\Big).
\end{aligned}$$

The results of Table II follow from (A3) and the following relations:

$$\begin{aligned}
E\left(h_i\exp\left(ith_4\sqrt{24}\right)\right) &= \phi(t)\left(\delta_{i4}\sqrt{24}\,it-n^{-1/2}(t^2/48)E\left(He_i(x)(He_4(x))^2\right)\right),\\
E\left(h_ih_j\exp\left(ith_4\sqrt{24}\right)\right) &= \phi(t)\Big(-24t^2\delta_{i4}\delta_{j4}-(1/48)\sqrt{24}\,n^{-1/2}it^3\\
&\qquad\times E\left(He_i(x)He_j(x)He_4(x)\right)(\delta_{i4}+\delta_{j4})\\
&\qquad+E\left(He_i(x)He_j(x)\right)\delta_{ij}\\
&\qquad+it\sqrt{24}\,n^{-1/2}E\left(He_i(x)He_j(x)He_4(x)\right)\Big);\\
E\left(h_ih_jh_k\exp\left(ith_4\sqrt{24}\right)\right) &= \phi(t)\Big(24\sqrt{24}\,(-it^3)\delta_{i4}\delta_{j4}\delta_{k4}\\
&\qquad+\sqrt{24}\,itE\left(He_i(x)He_j(x)\right)\delta_{ij}\delta_{k4}\\
&\qquad+2\text{ further terms got by permuting }(i,j,k)\Big);\\
E\left(h_ih_jh_kh_l\exp\left(ith_4\sqrt{24}\right)\right) &= \phi(t)\Big((24)^2t^4\delta_{i4}\delta_{j4}\delta_{k4}\delta_{l4}\\
&\qquad-24t^2E\left(He_i(x)He_j(x)\right)\delta_{ij}\delta_{k4}\delta_{l4}\\
&\qquad+5\text{ further terms}\\
&\qquad+E\left(He_i(x)He_j(x)\right)E\left(He_k(x)He_l(x)\right)\delta_{ij}\delta_{kl}\\
&\qquad\qquad+2\text{ further terms}\Big).
\end{aligned}$$

REFERENCES

ABRAMOWITZ, M., AND I. A. STEGUN (1965): *Handbook of Mathematical Functions*. New York: Dover.

BERA, A. K., AND C. R. MCKENZIE (1986): "Alternative Forms and Properties of the Score Test," *Journal of Applied Statistics*, 13, 13–25.

CHANT, D. (1974): "On Asymptotic Tests of Composite Hypotheses in Nonstandard Conditions," *Biometrika*, 61, 291–298.

CHESHER, A. (1983): "The Information Matrix Test: Simplified Calculation via a Score Test Interpretation," *Economics Letters*, 13, 45–48.

——— (1984): "Testing for Neglected Heterogeneity," *Econometrica*, 52, 865–872.

CHESHER, A., AND R. SPADY (1991): "Asymptotic Expansions of the Information Matrix Test Statistic," *Econometrica*, 59, 787–815.

DAVIDSON, R., AND J. G. MACKINNON (1984a): "Model Specification Tests Based on Artificial Linear Regressions," *International Economic Review*, 25, 485–502.

——— (1984b): "Convenient Specification Tests for Logit and Probit Models," *Journal of Econometrics*, 25, 241–262.

——— (1985): "Testing Linear and Loglinear Regressions Against Box-Cox Alternatives," *Canadian Journal of Economics*, 25, 499–517.

——— (1988): "Double-Length Artificial Regressions," *Oxford Bulletin of Economics and Statistics*, 50, 203–217.

——— (1990): "Specification Tests Based on Artificial Regressions," *Journal of the American Statistical Association*, 85, 220–227.

GODFREY, L. G. (1981): "On the Invariance of the Lagrange Multiplier Test with Respect to Certain Changes in the Alternative Hypothesis," *Econometrica*, 49, 1443–1455.

GODFREY, L. G., M. MCALEER, AND C. R. MCKENZIE (1988): "Variable Addition and Lagrange Multiplier Tests for Linear and Logarithmic Regression Models," *Review of Economics and Statistics*, 70, 492–503.

GODFREY, L. G., AND M. R. WICKENS (1981): "Testing Linear and Log-Linear Regressions for Functional Form," *Review of Economic Studies*, 48, 487–496.

HALL, A. (1987): "The Information Matrix Test for the Linear Model," *Review of Economic Studies*, 54, 257–263.

KENNAN, J., AND G. R. NEUMANN (1988): "Why Does the Information Matrix Test Reject Too Often? A Diagnosis of Some Monte Carlo Symptoms," Hoover Institution, Stanford University, Working Papers in Economics E-88-10.

LANCASTER, T. (1984): "The Covariance Matrix of the Information Matrix Test," *Econometrica*, 52, 1051–1053.

ORME, C. (1990a): "The Small Sample Performance of the Information Matrix Test," *Journal of Econometrics*, 46, 309–331.

——— (1990b): "Double and Triple Length Regressions for the Information Matrix Test and Other Conditional Moment Tests," University of York, mimeo.

TAYLOR, L. W. (1987): "The Size Bias of White's Information Matrix Test," *Economics Letters*, 24, 63–67.

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.