Open access • Journal Article • DOI:10.1080/02664768600000002

Alternative forms and properties of the score test — Source link

Anil K. Bera, Colin McKenzie

Institutions: University of Illinois at Urbana–Champaign, Australian National University

Published on: 01 Jan 1986 - Journal of Applied Statistics (Routledge)

Topics: Score test, Score and Test (assessment)

B24 E 2-3

H or V	JUST	FONT	SLOT	TITLE
H	CC	1W	22	
			21	LTY
			20	WORKING
			19	PAPER
H	CC	1W	8	1983-84
			7	NO.991-1009
H	CC	1W		330 B385<"CV"> no.991-1009 cop.2
H	CC	7W		<IMPRINT> U. of ILL. LIBRARY URBANA

THE HECKMAN BINDERY, INC.
North Manchester, Indiana

DMF

BINDING COPY

PERIODICAL ☐ CUSTOM ☐ STANDARD ☐ ECONOMY ☐ THESIS | NO. VOLS THIS TITLE | LEAD ATTACH
BOOK ☐ CUSTOM ☐ MUSIC ☐ ECONOMY | AUTH 1ST ☐

ACCOUNT: 66672 001 | LIBRARY | NEW | RUB OR SAMPLE | TITLE I.D.: 6672 | COLOR FOIL: WHT | MATERIAL: 488

ACCOUNT NAME: UNIV OF ILLINOIS

ACCOUNT INTERNAL I.D.: B01912400 | ISSN.

I.D. #2: STX4 | NOTES | BINDING FREQUENCY | WHEEL: 1 | SYS. I.D.: 3 | 39276

COLLATING: 35

ADDITIONAL INSTRUCTIONS
Dept=STX4 Lot=201 Item=121 HNM=(2Y=
ICR2STICR MARK BY # B4 91)

SEP. SHEETS | PTS. BD. PAPER | TAPE STUBS | CLOTH EXT | FILLER: GUM | STUB

POCKETS: PAPER | BUCK | CLOTH | SPECIAL PREP | LEAF ATTACH

INSERT MAT | ACCOUNT LOT NO: 201 | JOB NO: HV365

PRODUCT TYPE: 11 | ACCOUNT PIECE NO: 121 | PIECE NO: 41

HEIGHT: 11 | GROUP CARD: 1 | VOL. THIS TITLE: 37

COVER SIZE: X

001247943

BEBR

FACULTY WORKING
PAPER NO. 1005

Alternative Forms and Properties of the Score Test

Anil K. Bera
Colin R. McKenzie

College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois, Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 1005

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

January 1984

Alternative Forms and Properties
of the Score Test

Anil K. Bera, Assistant Professor
Department of Economics

Colin R. McKenzie
Australian National University and
Economic Planning Agency, Tokyo

Revised: December 1983

ALTERNATIVE FORMS AND PROPERTIES
OF THE SCORE TEST

Anil K. Bera
University of Illinois

and

Colin R. McKenzie
Australian National University and Economic Planning Agency, Tokyo

ABSTRACT: Two different issues relating to the score(S) test are investigated. Firstly, we study the finite sample properties of a number of asymptotically equivalent forms of the S test. From our simulation results we observe that these forms can behave very differently in finite samples. Secondly, we investigate the power properties of the S test and find that it compares favorably to those of the likelihood ratio (LR) test although the former does not use information about the precise forms of the alternatives.

KEYWORDS: Comparison of power, finite sample properties, heteroscedasticity, likelihood ratio test, non-normality, score test, serial correlation, simulated critical value, simulation study.

1. INTRODUCTION

In the statistical literature three basic principles are available for hypothesis testing, namely the likelihood ratio (LR), Wald (W) and score (S) principles. In most cases, the LR and W tests are used. However, when the alternative hypothesis is complicated, which arises very frequently in econometric model specification tests, these two tests tend to be unpopular because both require estimation under the alternative hypothesis. Estimation under the alternative hypothesis sometimes may prove to be difficult while test statistics based on the S principle can usually be calculated easily because it requires estimation only under the null hypothesis. Thus S tests can examine the validity of a null model against a general alternative model without estimating the latter and hence provide handy tools to tackle the complex problem of model specification.

An uneasy feature of the S test is that a number of asymptotically equivalent forms of the test can easily be developed. Some of them can be discarded on analytical or computational grounds. Still there remains a choice among the several asymptotically equivalent versions of the test which could have very different finite sample properties. The question is which one should be used in what situation. We attempt to provide a partial answer to this question in the next section through a simulation study. One interesting property of the LM test is its invariance to a class of alternative hypotheses. In other words, it does not utilize the precise information regarding the alternative. In Section 3, we investigate whether this affects the power properties of the S test. In general, the power is not affected. In the last section we present some concluding remarks.

2. A COMPARATIVE STUDY OF ALTERNATIVE FORMS OF THE S STATISTIC

Let $\ell_i(\theta)$ denote the log-density function for the i-th observation, where θ is a p x 1 parameter vector. Say we have N independent observations. Then the log-likelihood function is $\ell \equiv \ell(\theta) = \Sigma_{i=1}^{N} \ell_i(\theta)$. The hypothesis to be tested is $H_o : h(\theta) = 0$ where $h(\theta)$ is an r x 1 vector function of θ. It is assumed that $H \equiv H(\theta) = \partial h(\theta)/\partial\theta$ has full column rank, i.e., rank (H) = r. The S statistic (or the Lagrange multiplier statistic as is usually known in the econometrics literature), originally suggested by Rao (1948), for testing H_o can be written as

$$LM = \tilde{d}'\tilde{\underline{I}}^{-1}\tilde{d} \tag{1}$$

where $d \equiv d(\theta) = \partial\ell/\partial\theta$ is the efficient score vector, $\underline{I} \equiv \underline{I}(\theta) = E[-\partial^2\ell/\partial\theta\partial\theta'] = V[d(\theta)]$ is the information matrix and "~" indicates that the quantities have been evaluated at the restricted maximum likelihood estimate (MLE) of θ, say $\tilde{\theta}$. Under H_o, S is asymptotically distributed as a χ^2 with r degrees of freedom (χ^2_r).

In the last few years, a number of different forms of the S statistic have been developed that differ only in their choice of an estimator for the information matrix $\underline{I}$.[†] Let A be a positive definite matrix such that

---

[†]Engle (1982) suggests calculating the variance of $d(\tilde{\theta})$ directly instead of obtaining the information matrix explicitly. Although in most cases this would be a better approach to follow, it does not remedy our difficulties since the variance can be calculated in a number of ways which are all asymptotically equivalent.

$$\text{plim } [\tilde{A}^{-1} \{-\frac{\partial^2 \ell(\tilde{\theta})}{\partial\theta\partial\theta'}\}] = I \qquad (2)$$

then all the statistics of the form $\tilde{d}'\tilde{A}^{-1}\tilde{d}$ will be asymptotically equivalent. Given their asymptotic equivalence, choices have been based primarily on computational convenience. However, in many situations more than one form of the S statistic can be calculated easily, so that computational simplicity does not offer much guidance. Also, the asymptotic equivalence of tests is not necessarily indicative of their small sample behavior. Some forms may have finite sample distributions that are much closer to the asymptotic χ^2 under the null hypothesis. There may also be substantial differences in power. In this section we study these problems, and we consider the following versions of the S statistic suggested in the literature.

Whenever it is possible to take expectations of $[d(\theta)d'(\theta)]$ or $\partial^2\ell(\theta)/\partial\theta\partial\theta'$, we can use

$$A_1 = E[d(\theta)d'(\theta)]$$

or

$$A_2 = - E [\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta'}]$$

in place of A. If the specified probability model is correct then $A_1 = A_2$ [see White (1982)]. When either A_1 or A_2 is used we will denote the statistic by S(WE); "WE" stands for "with expectation."

In certain situations it may be difficult to take expectations, for example, when testing linear and log-linear models using the Godfrey and Wickens (1981a) approach. In such cases we can use either

$$A_3 = -\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta'}$$

or

$$A_4 = GG'$$

where G' has typical (i,j)-th element as $\partial \ell_i(\theta)/\partial\theta_j$, (i=1, 2,..., N; j=1, 2,..., p). One problem with $\tilde{A}_3$ is that it may not be positive definite in small samples and this may result in negative values of the calculated test statistics. When $\tilde{A}_4$ is used, the LM statistic can be calculated as $N.R^2$ where R^2 is the uncentred coefficient of determination from the regression of a vector of ones on $\tilde{G}'$ [Godfrey and Wickens (1981a, p. 490)]. Bera (1982a) argued that this statistic can be written in Hotelling's T^2 form and for finite samples it is approximately distributed as an $Nr/(N-r+1)$ multiple of $F(r, N-r+1)$ under H_o. For future reference we will denote this form as S(WOE) to mean "without expectation."

Since $\tilde{A}_1$ and $\tilde{A}_2$ contain fewer stochastic terms than does $\tilde{A}_4$ (or $\tilde{A}_3$), S(WE) can be expected to behave better in finite samples and its convergence to χ^2 under H_o may also be faster compared to S(WOE). Efron and Hinkley (1978) argued that $\tilde{A}_3$ (or $\tilde{A}_4$) is "closer to the data" than $\tilde{A}_2$ and should be favored as a variance approximation of $d(\tilde{\theta})$. Also, in the event of possible misspecification (such as testing heteroscedasticity under normality when normality assumption is in fact violated), we can expect $\tilde{A}_3$ and $\tilde{A}_4$ to be more robust. Therefore, intuitively it is not clear which of the two forms S(WE) and S(WOE) should be used in practice when both can be calculated easily.

Many hybrid forms of the S statistic can be constructed by selecting elements from either $\tilde{A}_1$ or $\tilde{A}_4$, i.e., by taking expectations of certain elements whenever possible, otherwise simply choosing elements from $\tilde{A}_4$. This technique was used by Breusch (1978) in deriving a S statistic for testing serial independence in the regression model. However, the information matrix obtained in this manner is not necessarily positive definite.

Recently Davidson and MacKinnon (1981) have proposed a new version of the S statistic. Let the underlying model be written as

$$f_i(y_i, y_{-i}, \theta) = \varepsilon_i$$

where y_i is the dependent variable, y_{-i}'s are its lagged values, $\varepsilon_i \sim N(0,1)$ and the Jacobian matrix of the transformation from y's to ε's is lower triangular. Davidson and MacKinnon (1981) suggested that the following matrix can be used in place of $\underline{I}$

$$A_5 = FF' + JJ'$$

where F' and J' have typical (i,j)-th elements as $\partial f_i/\partial\theta_j$ and $\partial(\ln|\partial f_i/\partial y_i|)/\partial\theta_j$ respectively. It is shown in Davidson and MacKinnon (1981) that $E(A_5) = E(A_4)$. Therefore, A_5 provides a valid estimator for $\underline{I}$. The S statistic using A_5 will be denoted as S(DM). Davidson and MacKinnon (1983) argued that A_5 provides a much more efficient estimate of $\underline{I}$ than does A_4 and the S test based on A_5 can be expected to behave better in small samples.

All the above versions are based on the assumption that the underlying probability model is correctly specified. When this assumption fails, the above statistics will not have the correct size even asymptotically. To overcome this problem, White (1982) proposed using

$$A_6 = A_3^{-1} A_4 \bar{A}_3^{-1}$$

and the operational form of his statistics is

$$S(W) = \tilde{d}' \tilde{\bar{A}}_3^{-1} \tilde{H}[\tilde{H}' \tilde{A}_6 \tilde{H}]^{-1} \tilde{H}' \tilde{\bar{A}}_3^{-1} \tilde{d}.$$

An interesting feature of this statistic is that it is based not just on the score vector $\tilde{d}(\theta)$, but on $[\tilde{H}' \tilde{\bar{A}}_3^{-1} \tilde{H}]^{-1} \tilde{H}' \tilde{\bar{A}}_3^{-1} \tilde{d}$ which is the estimate of the Lagrange multiplier vector. Statistic S(W) can also be expressed in $N.R^2$ form where R^2 is the uncentred coefficient of determination of the regression of a unit vector on $\tilde{D} = \tilde{G}' \tilde{\bar{A}}_3^{-1} \tilde{H}$. This is easily seen by writing S(W) as

$$S(W) = \underline{1}' \tilde{G}' \tilde{\bar{A}}_3^{-1} \tilde{H}[\tilde{H}' \tilde{\bar{A}}_3^{-1} \tilde{G} \tilde{G}' \tilde{\bar{A}}_3^{-1} \tilde{H}]^{-1} \tilde{H}' \tilde{\bar{A}}_3^{-1} \tilde{G}\underline{1}$$

$$= N \cdot [\underline{1}' \tilde{D}(\tilde{D}' \tilde{D})^{-1} \tilde{D}'\underline{1}/\underline{1}'\underline{1}]$$

where $\underline{1}$ is an N x 1 vector of ones. Therefore, S(W) can be calculated using the $N.R^2$ form, but the necessity to calculate $\tilde{\bar{A}}_3^{-1}$ indicates its computational complexity compared to S(WOE).

Harvey (1981, p. 173) has proposed a modified LM statistic

$$S(HM) = \frac{S(WE)}{N - S(WE)} \cdot \frac{N - p}{r}$$

which is approximately distributed as F(r,N-p) under the null hypothesis. The justification for this modification lies in the relationship between the S statistic and the usual F statistic for testing linear restrictions in the linear regression model with normal disturbances. Kiviet (1981) reports some simulation results in the context of testing serial independence, suggesting that this modification can lead to a closer correspondence between the nominal and actual significance levels.

Davidson and MacKinnon (1983) have examined the small sample performances of S(WOE) and S(DM) for testing linear and log-linear models within a Box-Cox transformation framework. Their main findings are that (i) under the null hypothesis S(DM) is closer to χ^2 distribution in small samples than is S(WOE) and (ii) there is not much difference in terms of power between these two statistics. We took a different model for our simulation study since all the above S statistics cannot be calculated for testing linear and log-linear models. Also we wanted to have a sufficiently flexible model that permitted the study of the behavior of all the statistics under different circumstances.

Our study examines the small sample performance of the S statistics listed above for testing jointly the homoscedasticity (H) and serial independence (I) of normal regression disturbances. For the alternative hypothesis the data were gathered by violating H, I and normality (N) as well as their various combinations. This provides us with an opportunity to compare the performances of the statistics under different situations such as "undertesting" (when the statistic fails to test all departures that occur), "exact testing" (the situation for which the statistic is designed) and "overtesting" (when we test for more departures than actually occur) [for a detailed discussion of these concepts, see Bera and Jarque (1982)].

We considered the following linear regression model

$$y_i = \sum_{j=1}^{4} x_{ij}\beta_j + u_i$$

where we set $x_{i1} = 1$ and generated x_{i2} from N(10,25), x_{i3} from the uniform (U) distribution in the interval [7.5, 12.5] and x_{i4} from χ^2_{10}. These values of the regression matrix, X, were kept fixed from one replication to another. Serially correlated $(\overline{\underline{I}})$ disturbances were generated from the first order autoregressive (AR) process $u_i = \rho u_{i-1} + \varepsilon_i, |\rho| < 1$. "Weak" and "strong" autocorrelations were represented by setting $\rho = \rho 1 = 0.3$ and $\rho = \rho 2 = 0.7$, respectively. Heteroscedastic $(\overline{\underline{H}})$ disturbances were generated by assuming $V(\varepsilon_i) = \sigma_i^2 = 25 + \sigma z_i$, $\sqrt{z_i} \sim N(10,25)$ and α was set to $\alpha = \alpha 1 = 0.25$ and $\alpha = \alpha 2 = 0.85$ for "weak" and "strong" heteroscedasticity. Non-normal $(\overline{\underline{N}})$ ε_i's were obtained from the Student's distribution with five degrees of freedom and the log-normal distribution (say t and log). Combining these departures from H_o: $u \sim \underline{NHI}$ (or $\alpha = \rho = 0$), we have 26 possible alternatives: 6 one-directional, 12 two-directional and 8 three-directional departures. Observations of u_i under H_o were generated by taking $u_i \sim N(0,25)$.

In every experiment, for a given sample size N, we generated the data under the null and 26 alternatives, and calculated the above five S statistics. Algebraic derivations of the statistics are given in the Appendix. The experiments were performed for N = 20, 35, 50, 100 and 200, and for each N we carried out 500 replications. [Therefore, the maximum standard errors of the estimates of type 1 errors and powers in the following tables would be $\sqrt{.5(1-.5)/500} \simeq .022$]. The calculated statistics under H_o were used to study the closeness of their null distributions to χ^2_2 (or the appropriate F distribution) through the Kolmogorov-Smirnov (K-S) test, and to estimate the type I errors. These statistics values were also used to obtain the simulated 10 percent significance points which are the appropriate order statistics (450-th) of the 500 values. The powers of a test were estimated by counting the numbers of times the text statistic exceeded the corresponding empirical and asymptotic significance points and dividing those by 500. We report the values of the K-S statistic (based on 100 replications because tabulated critical values are only available up to 100) in Table I and estimates of the type I errors for all the statistics are presented in Table II.

TABLE I

Kolmogorov-Smirnov statistics for testing departures from χ^2 and F distributions

Sample Size	S(WE)	S(WOE)		S(DM)	S(W)	S(HM)
	χ^2	χ^2	F	χ^2	χ^2	F
20	.108	.161*	.118	.110	.075	.218*
35	.097	.139*	.115	.064	.081	.157*
50	.077	.081	.064	.055	.083	.099
100	.083	.112	.106	.072	.091	.064
200	.047	.053	.049	.051	.040	.056

[a] Statistic values with "*" are significant at 5 percent level.

TABLE II

Estimated type I errors for different statistics
(nominal level .10)

Sample Size	S(WE)	S(WOE)		S(DM)	S(W)	S(HM)
	χ^2	χ^2	F	χ^2	χ^2	F
20	.092	.180	.134	.082	.116	.062
35	.078	.176	.144	.080	.146	.054
50	.084	.172	.162	.100	.142	.072
100	.088	.154	.148	.082	.150	.082
200	.098	.136	.136	.100	.128	.096

From Tables I and II we observe that S(WE) and S(DM) perform very well. Even for small sample sizes, both follow χ^2 distributions and actual type I errors are close to the nominal level of .10. The statistic S(WOE) does not follow the χ^2 distribution for N = 20 and 35. Although its distribution is closer to χ^2 for N > 50, its performance is not good in terms of type I error. However, the hypothesis of an F distribution is accepted for S(WOE), thereby justifying the conjecture made in Bera (1982a). When the appropriate critical values from the F distribution are used, the modification reduces the type I errors of S(WOE) for N = 20 and 35 only as seen from Table II. The statistic S(W) follows χ^2 for all sample sizes. However, the behavior of type I errors corresponding to S(W) is very erratic and are much higher than .10. This shows that although the overall distribution is close to χ^2, the tail part which determines the type I error is not. Lastly, Harvey's correction does not work in our case. This is in contrast to the results obtained by Kiviet (1981).

The estimated powers of the tests, calculated using simulated critical points are presented in Table III for N = 50. Since the behavior of the tests differs under the null hypothesis, as illustrated in Tables I and II, any meaningful power comparison requires the use of estimated (rather than asymptotic) critical values. For S(HM), estimated powers, using simulated critical points, are the same as those for S(WE) since the former is a monotonic transformation of the latter. The left hand column in Table III indicates the characteristics of the regression disturbances. Recall that $\rho 1$ and $\rho 2$ denote "weak" and "strong" first order serial correlation, $\alpha 1$ and $\alpha 2$ denote "weak" and "strong" heteroscedasticity and, "t" and "log" denote that the errors ε_i were obtained from student t and log-normal distributions respectively. So, for example, $u \sim \underline{N\overline{HI}}(\alpha 1, \rho 2)$ indicates the errors are normally distributed, "weakly" heteroscedastic and "strongly" serially correlated.

All the test statistics being investigated are designed to test the null hypothesis of white noise, homoscedastic and normal disturbances ($u \sim \underline{NHI}$) against the alternative of first order serially correlated and heteroscedastic but normal errors ($u \sim \underline{N\overline{HI}}$). For such alternatives, on the basis of the relative numerical magnitudes of the powers, the preference ordering would be S(WE), S(DM), S(W), and S(WOE). Although it should be said that the power differences are not substantial (maximum power difference between the most powerful and least powerful test is only 0.06). In the case of overtesting, where

TABLE III

Estimated powers of the tests
(N=50, with empirical 10 percent significance level)

	S(WE)	S(WOE)	S(DM)	S(W)
One-directional				
$\overline{\text{NHI}}$(t)	.188	.156	.176	.190
(log)	.530	.528	.554	.242
$\overline{\text{NHI}}$(α1)	.642	.530	.580	.572
(α2)	.858	.778	.832	.800
$\overline{\text{NHI}}$(ρ1)	.470	.418	.452	.452
(ρ2)	.986	.974	.984	.978
Two-directional				
$\overline{\text{NHI}}$(t,α1)	.636	.502	.580	.486
(log,α1)	.524	.498	.528	.226
(t,α2)	.816	.702	.786	.640
(log,α2)	.578	.538	.584	.220
$\overline{\text{NHI}}$(α1,ρ1)	.780	.720	.764	.752
(α2,ρ1)	.908	.870	.902	.878
(α1,ρ2)	.994	.988	.994	.990
(α2,ρ2)	.996	.994	.998	.998
$\overline{\text{NHI}}$(t,ρ1)	.530	.460	.522	.466
(log,ρ1)	.708	.582	.726	.158
(t,ρ2)	.990	.964	.990	.964
(log,ρ2)	.990	.928	.994	.790
Three-directional				
$\overline{\text{NHI}}$(t,α1,ρ1)	.786	.670	.772	.672
(log,α1,ρ1)	.704	.582	.704	.182
(t,α2,ρ1)	1.000	.992	.998	.990
(log,α2,ρ1)	.990	.912	.992	.800
(t,α1,ρ2)	.998	.986	.998	.980
(log,α1,ρ2)	.992	.918	.994	.798
(t,α2,ρ2)	1.000	.992	.998	.990
(log,α2,ρ2)	.990	.912	.992	.802

we test for more departures from the null hypothesis than actually occur, i.e., u ~ $\underline{N\overline{H}I}$ or u ~ $\underline{NH\overline{I}}$, the preference ordering remains the same. When the underlying assumption of normality is violated, i.e., u ~ $\underline{\overline{N}HI}$, u ~ $\underline{\overline{NH}I}$, u ~ $\underline{\overline{N}H\overline{I}}$ or u ~ $\underline{\overline{NHI}}$, some interesting results arise. For example, when there is undertesting, u ~ $\underline{\overline{NHI}}$, the powers of S(WE), S(WOE) and S(DM) are not adversely affected. However, quite surprisingly, that is not the case for S(W). It loses substantial power when the alternative is contaminated by the log-normal distribution. For example, when u ~ $\underline{N\overline{HI}}(\alpha 1,\rho 1)$ the power is 0.752 but this falls to 0.182 when u ~ $\underline{\overline{NHI}}(\log,\alpha 1,\rho 1)$. The power loss is smaller but still significant for the other log-normal cases. The S(W) statistic suffers a similar loss of power when the alternative of serially correlated errors or heteroscedastic errors is contaminated by the log-normal distribution. The other tests also tend to have lower power in these cases. For the alternatives u ~ $\underline{\overline{NH}I}$ and u ~ $\underline{\overline{N}H\overline{I}}$, S(WE) is best followed by S(DM), S(WOE) and S(W). When the model is completely misspecified, such as when the alternative is $\underline{\overline{N}HI}$, one would expect S(W) to perform "better" with a power around 0.10 since it is designed for the case when the probability model is not correctly specified. For $\underline{\overline{N}HI}(t)$, the performance of LM(W) and all the other statistics is not good. However, when the alternative is $\underline{\overline{N}HI}(\log)$, S(W)'s performance while still not good is far superior to the other tests.

Similar observations can be made on the power of the tests for N = 20, 35, 100 and 200.† In fact, for these values of N, S(DM) performs

---

†Results corresponding to these values of N are available from the authors on request.

marginally better than S(WE). On the basis of the above limited simulation study our recommendation would be to use S(DM) or S(WE), and when there is a possibility of misspecification S(W) can be used with some caution. In certain cases, S(WE) cannot be obtained whereas S(DM) can be obtained easily, for example, while testing linear and log-linear models within Box-Cox framework. There are also circumstances where S(DM) is not applicable such as for specification tests in the limited dependent variable models [see Bera et al. (1983)]. When both are available either of them can be used since in terms of type I error and power there is not much difference among these two versions of the S test.

3. INVARIANCE OF THE S STATISTIC

A striking feature of the S test is its invariance to different alternatives. There are many interesting examples of this, but here we will mention only a few. The S statistic for testing normality suggested in Bera and Jarque (1981) with the Pearson family of distributions as the alternative, remains unchanged under Gram-Charlier (type A) alternatives [see Bera (1982b, p. 98)]. Statistics for testing homoscedasticity are invariant to different forms of alternatives such as multiplicative and additive heteroscedasticity, as has been noted by Breusch and Pagan (1979) and Godfrey and Wickens (1981b). Testing serial independence against q-th order autoregressive [AR(q)] or q-th order moving average [MA(q)] processes lead to the same test statistic; see for instance, Breusch (1978) and Godfrey (1978). Pesaran (1979) found that the S test is "incapable of differentiating between polynomial lags and rational distributed lags." These examples raise the question whether the S test will be inferior to other asymptotically equivalent tests with respect to power since it does not use precise information of the alternative.

Godfrey (1981) took up this question and compared the powers of the S and LR tests for local alternatives in the context of testing serial independence when the alternatives are AR(1) and MA(1) processes. Using his analysis it can be shown that AR(1) and MA(1) processes are "almost the same" for local alternatives. To illustrate this point, consider the following AR(1) and MA(1) processes

$$u_i = \rho u_{i-1} + \varepsilon_i \tag{3}$$

and

$$u_i = \rho \varepsilon_{i-1} + \varepsilon_i \tag{4}$$

where $\rho = \delta(1/\sqrt{N})$, $\delta < \infty$. An equivalent expression for (4) is

$$u_i = \varepsilon_i + \rho u_{i-1} - \rho^2 u_{i-2} + \ldots$$

For $\rho = \delta(1/\sqrt{N})$, each term beyond ρu_{i-1} will be at least of order $O(1/N)$. Therefore, it follows that both of these alternatives are basically the same.

To make a power comparison it might be desirable to consider _fixed_ alternatives where substantial differences in power can be expected. To achieve this, we consider testing homoscedasticity against additive and multiplicative heteroscedasticity, and through a simulation experiment we compare the powers of the S and LR statistics. The model is the same as in Section 2, with the only difference being the alternatives are now additive heteroscedasticity (<u>AH</u>)

$$\sigma_i^2 = \sigma^2 + \alpha z_i \tag{5}$$

and multiplicative heteroscedasticity (<u>MH</u>)

$$\sigma_i^2 = \exp[\sigma^2 + \alpha z_i] \tag{6}$$

with normal and serially independent disturbances. For the same values of σ^2 and α, the degree of heteroscedasticity can be expected to be higher in <u>MH</u> than in <u>AH</u>. In our simulation study we set $\sigma^2 = 1.0$ and generated z_i from a U[0,1]. We selected two values of α, 3.0 and 6.0, to represent "weak" and "strong" heteroscedasticity and to obtain powers within a convenient illustrative range.

The S statistic [S(WE) version] for testing H_o: $\alpha = 0$ against <u>MH</u> or <u>AH</u> is [Breusch and Pagan (1979, p. 1290)]

$$S = \left[\frac{\tilde{u}'Q\tilde{u}}{\tilde{u}'\tilde{u}} \right]^2$$

where $Q = \text{diag}[N(z_i - \bar{z})/\{2\Sigma_i(z_i - \bar{z})^2\}^{1/2}]$, $\tilde{u} = y - X\tilde{\beta}$ and $\bar{z} = \Sigma_i z_i/N$. The LR statistic when the alternative is <u>MH</u>, is given by [Harvey (1976, p. 464)]

$$LR_M = N \ln\tilde{\sigma}^2 - N\hat{\sigma}^2 - \hat{\alpha}_M\Sigma_i z_i$$

where $\hat{\sigma}_M^2$ and $\hat{\alpha}_M$ are the unrestricted MLEs of σ^2 and α respectively. To obtain the estimates we have used Harvey's (1976, pp. 463-464) algorithm. The expression for the LR statistic when we take <u>AH</u> as the alternative, is

$$LR_A = N(\ln\tilde{\sigma}^2 + 1) - \Sigma_i \ln(\hat{\sigma}_A^2 + \hat{\alpha}_A z_i)$$

$$- \Sigma_i \frac{(y_i - x_i'\hat{\beta}_A)^2}{\hat{\sigma}_A^2 + \hat{\alpha}_A z_i}$$

where $\hat{\sigma}_A^2$, $\hat{\alpha}_A$ and $\hat{\beta}_A$ are the unrestricted MLEs of σ^2, α and β respectively. These estimates were obtained by using the method of scoring. In all cases, the iteration process was terminated when the difference between two successive values of the likelihood function was less than $\epsilon = .01$ or after the 40-th iteration whichever came first. In some cases we set the value of ϵ to .001, but this did not have any significant effects on the results. A by-product of our study is the observation that the inequality LR ⩾ S established for testing linear or nonlinear restrictions in the linear regression model, does not hold for our nonlinear model. A similar observation was made by Mizon (1977, p. 1237) in relation to the inequality W ⩾ LR.

The results for type I errors are given in Table IV, and on power in Table V, for sample sizes N = 20, 40 and 80. All these results are based on 500 replications. In Table V, alternatives 1 and 2 are $\overline{AH}$ with $\alpha = 3.0$ and 6.0, respectively, and alternatives 3 and 4 are $\overline{MH}$ with $\alpha = 3.0$ and 6.0, respectively.

From Table IV we observe that for the S statistic type I errors are less than .10 for all sample sizes, whereas for LR_M and LR_A they are much higher than .10 for N = 20 and 40. However, for N = 80 both LR_M and LR_A have significance levels of around .10. Since type I errors of the three statistics differs substantially, especially when

TABLE IV

Estimated type I errors for the three statistics
(nominal level .10)

Sample size	S	LR_M	LR_A
20	.068	.204	.332
40	.070	.140	.188
80	.072	.082	.102

TABLE V

Estimated powers of the three statistics[a]
(with empirical 10 percent significance level)

Alternatives		S	LR_M	LR_A
N=20	1	.266	.288	.258
	2	.366	.390	.356
	3	.706	.776	.606
	4	.960	.990	.960
N=40	1	.404	.388	.382
	2	.556	.568	.570
	3	.924	.936	.934
	4	1.000	1.000	1.000
N=80	1	.746	.750	.752
	2	.906	.922	.936
	3	1.000	1.000	1.000
	4	1.000	1.000	1.000

[a]Alternatives 1 and 2 are $\underline{\overline{AH}}$ with α = 3.0 and 6.0 respectively, and alternatives 3 and 4 are $\underline{\overline{MH}}$ with α = 3.0 and 6.0 respectively.

the sample size is small, we have used simulated critical points for calculating power. It is observed from Table V that powers for alternatives 3 and 4 are higher than those for alternatives 1 and 2, as expected. For alternatives 3 and 4, LR_M does better than LR_A as one would expect. However, for alternatives 1 and 2, LR_A does not always have higher power than LR_M. Comparing the powers for S with those for LR_M and LR_A, it is seen that for N = 20, the power of S lies between the powers of LR_M and LR_A, and for N = 40 and 80, S has slightly less power than the other two statistics. Therefore, the S test is effective in detecting both kinds of heteroscedasticity. Godfrey (1981) reached to a similar conclusion when he compared the S and LR procedures in the context of testing serial independence against AR(1) and MA(1) alternatives. These observations show that although the algebraic form of the S statistic does not take account of the specific alternatives, its numerical values do take account of the data. Therefore, the invariance of the S test to a class of alternatives is not necessarily a drawback of the procedure. Given the computational complexities of the LR procedure, the S test may be preferred. Moreover, even though the LR statistic uses specific information concerning the alternative, it does not provide any guidance about the exact nature of the alternative. For instance, from Table V it is seen that LR_M (LR_A) has very high power against $\underline{A\bar{H}}$($\underline{M\bar{H}}$). Therefore, if the LR statistic is found to be significant, we cannot infer much about the alternative - the same complaint levelled against the S test.

4. CONCLUSION

In this paper we have studied some of the properties of the S test procedure. It has been observed that various asymptotically equivalent forms of the S test can have different finite sample properties. On the basis of the results of Section 2, we recommend the use of S(WE) and S(DM) versions. We found that the S test has good power to detect heteroscedasticity per se, in spite of its inability to discriminate between various forms of the alternative hypothesis.

As an end-note on this paper, we should add that since our conclusions are based on the simulation results they are open to many criticisms. However, in the absence of analytical results, we do feel, our study sheds further lights on the behavior of the S statistics in addition to the findings of Davidson and MacKinnon (1983) and Godfrey (1981). And since our primary interest is the relative performances of various statistics which have been calculated from the same samples (and hence are positively correlated), the standard errors of the differences of powers can be expected to be less than those of the estimates of powers [see Godfrey (1981, p. 1451)].

APPENDIX

Formulae for the S statistics used in section 2.

For our model $\theta = (\beta', \sigma^2, \alpha, \rho)'$, H_o: $\alpha = \rho = 0$ and the log-likelihood function $\ell(\theta)$ is given by

$$\ell(\theta) = \sum_{i=1}^{N} \ell_i(\theta) = -(N/2)\ln 2\pi - \frac{1}{2}\sum_{i=1}^{N} \ln\sigma_i^2 - \frac{1}{2}\sum_{i=1}^{N} \frac{\varepsilon_i^2}{\sigma_i^2} \tag{A.1}$$

where $\sigma_i^2 = \sigma^2 + \alpha z_i$ and $\varepsilon_i = u_i - \rho u_{i-1}$ with $u_i = y_i - x_i'\beta$. From the above equation following first order derivatives are easily obtained

$$\frac{\partial \ell_i(\theta)}{\partial \beta} = \frac{1}{\sigma_i^2}(x_i - \rho x_{i-1})\varepsilon_i \tag{A.2}$$

$$\frac{\partial \ell_i(\theta)}{\partial \sigma^2} = -\frac{1}{2\sigma_i^2} + \frac{\varepsilon_i^2}{2\sigma_i^4} \tag{A.3}$$

$$\frac{\partial \ell_i(\theta)}{\partial \alpha} = -\frac{z_i}{2\sigma_i^2} + \frac{\varepsilon_i^2 z_i}{2\sigma_i^4} \tag{A.4}$$

and

$$\frac{\partial \ell_i(\theta)}{\partial \rho} = \frac{1}{\sigma_i^2}(y_{i-1} - x'_{i-1}\beta)\varepsilon_i. \tag{A.5}$$

Using these derivatives we get

$$\tilde{d}' = \left[\frac{\partial \ell(\tilde{\theta})}{\partial \theta}\right]' = \left[0,\ 0,\ -\frac{\Sigma_i z_i}{2\tilde{\sigma}^2} + \frac{\Sigma_i \tilde{u}_i^2 z_i}{2\tilde{\sigma}^4},\ \frac{\Sigma_i \tilde{u}_i \tilde{u}_{i-1}}{\tilde{\sigma}^2}\right],$$

and

$$\tilde{\underline{I}} = \begin{bmatrix} \frac{X'X}{\tilde\sigma^2} & 0 & 0 & 0 \\ 0 & \frac{N}{2\tilde\sigma^4} & \frac{\Sigma_i z_i}{2\tilde\sigma^4} & 0 \\ 0 & \frac{\Sigma_i z_i}{2\tilde\sigma^2} & \frac{\Sigma_i z_i^2}{2\tilde\sigma^2} & 0 \\ 0 & 0 & 0 & N \end{bmatrix}$$

where $\tilde u_i = y_i \; x_i'\tilde\beta$ and $\tilde\sigma^2 = \Sigma_i \tilde u_i^2/N$. Using (1) we can then obtain S(WE). To obtain S(WOE) and S(DM), block diagonality of $\underline{I}$ between β and (σ^2, α, ρ) is employed. For S(WOE) we need the matrix $\tilde G$ whose element are obtained from equations (A.3) to (A.5), evaluated at the restricted estimates. To obtain S(DM) we can define the function f(.) as

$$f_i = \frac{u_i - \rho u_{i-1}}{(\sigma^2 + \alpha z_i)^{1/2}} = \frac{\varepsilon_i}{\sigma_i} .$$

This statistic requires construction of two matrices F and J evaluated under H_o. The following derivatives provide the different elements of F:

$$\frac{\partial f_i}{\partial \beta} = -\frac{1}{\sigma_i}(x_i - \rho x_{i-1}),\ \frac{\partial f_i}{\partial \sigma^2} = -\frac{\varepsilon_i}{2\sigma_i^3},\ \frac{\partial f_i}{\partial \alpha} = -\frac{z_i \varepsilon_i}{2\sigma_i^3}$$

and $$\frac{\partial f_i}{\partial \rho} = -\frac{u_{i-1}}{\sigma_i} .$$

Let $\overline{f}_i = \ln|\partial f_i/\partial y_i| = -\ln\sigma_i$, then the elements of J are obtained from

$$\frac{\partial \overline{f}_i}{\partial \beta} = 0, \quad \frac{\partial \overline{f}_i}{\partial \sigma^2} = -\frac{1}{2\sigma_i^2}, \quad \frac{\partial \overline{f}_i}{\partial \alpha} = -\frac{z_i}{2\sigma_i^2}$$

and $$\frac{\partial \overline{f}_i}{\partial \rho} = 0.$$

Derivation of the other forms of the S statistic is straightforward.

ACKNOWLEDGEMENTS

We would like to thank Ray Byron, Michael McAleer, Grayham Mizon, Adrian Pagan, Jean Marie Rolin, Deane Terrell and especially Jan Kiviet for helpful comments, Russell Davidson and James MacKinnon for permission to refer to their unpublished paper and helpful suggestions. Errors, of course, remain our sole responsibility. The second author gratefully acknowledges the financial support provided by the Japanese Government's Department of Education and the research support provided by the Economic Planning Agency. The views and opinions expressed in this paper are not necessarily those of the Economic Planning Agency.

REFERENCES

Bera, A. K. (1982a), "A new test for normality," Economics Letters, 9, 263-268.

Bera, A. K. (1982b), Aspects of Econometric Modelling, unpublished Ph.D. dissertation, The Australian National University.

Bera, A. K. and C. M. Jarque (1981), "An efficient large-sample test for normality of observations and regression residuals," Working Papers in Economics and Econometrics, N° 040, The Australian National University.

Bera, A. K. and C. M. Jarque (1982), "Model specification tests: A simultaneous approach," Journal of Econometrics, 20, 59-82.

Bera, A. K., C. M. Jarque and L. F. Lee (1983), "Testing for the normality assumption in limited dependent variable models," International Economic Review, 25, (forthcoming).

Breusch, T. S. (1978), "Testing for autocorrelation in dynamic linear models," Australian Economic Papers, 17, 334-355.

Breusch, T. S. and A. R. Pagan (1979), "A simple test for heteroscedasticity and random coefficient variation," Econometrica, 47, 1287-1294.

Davidson, R. and J. G. MacKinnon (1981), "Model specification tests based on artificial linear regressions," Discussion Paper N° 426, Department of Economics, Queen's University.

Davidson, R. and J. G. MacKinnon (1983), "Small sample properties of alternative forms of the Lagrange multiplier test," Economics Letters, 12, 269-275.

Efron, B. and D. V. Hinkley (1978), "Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information," Biometrika, 65, 457-487.

Engle, R. F. (1982), "A general approach to Lagrange multiplier model diagnostics," Journal of Econometrics, 20, 83-104.

Godfrey, L. G. (1978), "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables," Econometrica, 46, 1293-1302.

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. and M. R. Wickens (1981a), "Testing linear and log-linear regressions for functional form," Review of Economic Studies, 48, 487-496.

Godfrey, L. G. and M. R. Wickens (1981b), "Tests of misspecification using locally equivalent alternative models," Discussion Paper 66, Department of Economics and Related Studies, University of York.

Harvey, A. C. (1976), "Estimating regression models with multiplicative heteroscedasticity," Econometrica, 44, 461-465.

Harvey, A. C. (1981), The Econometric Analysis of Time Series, (Oxford: Philip Allan).

Kiviet, J. F. (1981), "On the rigour of some specification tests for modelling dynamic relationships," Report AE 11/81, Faculty of Actuarial Science and Econometrics, University of Amsterdam.

Mizon, G. E. (1977), "Inferential procedures in nonlinear models: An application in a UK industrial cross section study of factor substitution and return to scale," Econometrica, 45, 1221-1242.

Pesaran, M. H. (1979), "Diagnostic testing and exact maximum likelihood estimation of dynamic models," in Proceedings of the Econometric Society European Meeting, 1979, ed. E. G. Charatsis, (Amsterdam: North Holland), pp. 63-87.

Rao, C. R. (1948), "Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation," Proceedings of the Cambridge Philosophical Society, 44, 50-57.

White, H. (1982), "Maximum likelihood estimation of misspecified models," Econometrica, 50, 1-25.

TABLE IIIA

Estimated powers of the tests
(N = 20, with 10 percent empirical significance level)

	S(WE)	S(WOE)	S(DM)	S(W)
One-directional				
$\overline{\text{NHI}}$(t)	.126	.124	.122	.118
(log)	.198	.340	.282	.154
$\overline{\text{NHI}}$(α1)	.266	.220	.234	.144
(α2)	.376	.316	.346	.168
$\overline{\text{NHI}}$(ρ1)	.126	.146	.152	.152
(ρ2)	.540	.538	.614	.610
Two-directional				
$\overline{\text{NHI}}$(t,α1)	.284	.234	.272	.152
(log,α1)	.232	.374	.270	.150
(t,α2)	.372	.286	.340	.168
(log,α2)	.280	.372	.296	.166
$\overline{\text{NHI}}$(α1,ρ1)	.284	.244	.298	.226
(α2,ρ1)	.378	.370	.380	.250
(α1,ρ2)	.626	.594	.696	.650
(α2,ρ2)	.650	.644	.744	.650
$\overline{\text{NHI}}$(t,ρ1)	.184	.188	.204	.174
(log,ρ1)	.200	.354	.304	.088
(t,ρ2)	.584	.542	.620	.558
(log,ρ2)	.606	.492	.710	.344
Three-directional				
$\overline{\text{NHI}}$(t,α1,ρ1)	.322	.292	.336	.244
(log,α1,ρ1)	.238	.356	.314	.102
(t,α2,ρ1)	.662	.634	.718	.614
(log,α2,ρ1)	.644	.494	.730	.406
(t,α1,ρ2)	.642	.588	.680	.590
(log,α1,ρ2)	.626	.496	.716	.382
(t,α2,ρ2)	.662	.634	.718	.614
(log,α2,ρ2)	.638	.506	.734	.404

TABLE IIIB

Estimated powers of the tests
(N = 35, with 10 percent empirical significance level)

	S(WE)	S(WOE)	S(DM)	S(W)
One-directional				
$\overline{\text{NHI}}$(t)	.160	.132	.172	.110
(log)	.488	.470	.496	.144
$\overline{\text{NHI}}$(α1)	.362	.234	.310	.238
(α2)	.552	.450	.524	.396
$\overline{\text{NHI}}$(ρ1)	.318	.252	.326	.274
(ρ2)	.924	.898	.920	.902
Two-directional				
$\overline{\text{NHI}}$(t,α1)	.342	.254	.320	.192
(log,α1)	.498	.484	.500	.132
(t,α2)	.508	.370	.472	.262
(log,α2)	.482	.472	.486	.122
$\overline{\text{NHI}}$(α1,ρ1)	.520	.414	.504	.426
(α2,ρ1)	.678	.560	.672	.532
(α1,ρ2)	.938	.920	.942	.916
(α2,ρ2)	.952	.924	.962	.926
$\overline{\text{NHI}}$(t,ρ1)	.388	.322	.380	.296
(log,ρ1)	.586	.480	.598	.112
(t,ρ2)	.906	.842	.912	.854
(log,ρ2)	.974	.800	.976	.570
Three-directional				
$\overline{\text{NHI}}$(t,α1,ρ1)	.504	.402	.482	.356
(log,α1,ρ1)	.588	.506	.594	.120
(t,α2,ρ1)	.946	.894	.952	.874
(log,α2,ρ1)	.982	.792	.980	.622
(t,α1,ρ2)	.942	.880	.936	.858
(log,α1,ρ2)	.978	.804	.978	.604
(t,α2,ρ2)	.946	.894	.952	.874
(log,α2,ρ2)	.982	.790	.980	.620

TABLE IIIC

Estimated powers of the tests
(N = 100, with 10 percent empirical significance level)

	S(WE)	S(WOE)	S(DM)	S(W)
One-directional				
$\overline{\text{NHI}}$(t)	.208	.160	.222	.178
(log)	.590	.566	.652	.232
$\overline{\text{NHI}}$(α1)	.876	.794	.874	.816
(α2)	.986	.974	.986	.980
$\overline{\text{NHI}}$(ρ1)	.776	.752	.796	.758
(ρ2)	1.000	1.000	1.000	1.000
Two-directional				
$\overline{\text{NHI}}$(t,α1)	.740	.590	.746	.576
(log,α1)	.592	.518	.630	.222
(t,α2)	.938	.856	.944	.816
(log,α2)	.614	.562	.642	.242
$\overline{\text{NHI}}$(α1,ρ1)	.970	.960	.978	.966
(α2,ρ1)	.994	.992	.998	.998
(α1,ρ2)	1.000	1.000	1.000	1.000
(α2,ρ2)	1.000	1.000	1.000	1.000
$\overline{\text{NHI}}$(t,ρ1)	.818	.752	.840	.738
(log,ρ1)	.958	.710	.962	.284
(t,ρ2)	1.000	1.000	1.000	1.000
(log,ρ2)	1.000	.970	1.000	.910
Three-directional				
$\overline{\text{NHI}}$(t,α1,ρ1)	.954	.902	.960	.874
(log,α1,ρ1)	.946	.716	.956	.276
(t,α2,ρ1)	1.000	1.000	1.000	1.000
(log,α2,ρ1)	1.000	.972	1.000	.886
(t,α1,ρ2)	1.000	1.000	1.000	1.000
(log,α1,ρ2)	1.000	.976	1.000	.894
(t,α2,ρ2)	1.000	1.000	1.000	1.000
(log,α2,ρ2)	1.000	.974	1.000	.886

TABLE IIID

Estimated powers of the tests
(N = 200, with 10 percent empirical significance level)

	S(WE)	S(WOE)	S(DM)	S(W)
One-directional				
$\underline{\overline{\text{NHI}}}(t)$	.204	.132	.202	.134
(log)	.696	.580	.738	.302
$\underline{\overline{\text{NHI}}}(\alpha 1)$	.970	.946	.970	.952
$(\alpha 2)$	1.000	.998	1.000	1.000
$\underline{\overline{\text{NHI}}}(\rho 1)$	.980	.976	.980	.976
$(\rho 2)$	1.000	1.000	1.000	1.000
Two-directional				
$\underline{\overline{\text{NHI}}}(t,\alpha 1)$	.888	.744	.894	.734
$(\log,\alpha 1)$	.702	.554	.738	.296
$(t,\alpha 2)$	.994	.966	.994	.930
$(\log,\alpha 2)$	.724	.580	.748	.252
$\underline{\overline{\text{NHI}}}(\alpha 1,\rho 1)$	1.000	1.000	1.000	1.000
$(\alpha 2,\rho 1)$	1.000	1.000	1.000	1.000
$(\alpha 1,\rho 2)$	1.000	1.000	1.000	1.000
$(\alpha 2,\rho 2)$	1.000	1.000	1.000	1.000
$\underline{\overline{\text{NHI}}}(t,\rho 1)$	.990	.986	.994	.978
$(\log,\rho 1)$	1.000	.836	1.000	.370
$(t,\rho 2)$	1.000	1.000	1.000	1.000
$(\log,\rho 2)$	1.000	.992	1.000	.952
Three-directional				
$\underline{\overline{\text{NHI}}}(t,\alpha 1,\rho 1)$	.996	.992	.998	.986
$(\log,\alpha 1,\rho 1)$	1.000	.830	1.000	.364
$(t,\alpha 2,\rho 1)$	1.000	1.000	1.000	1.000
$(\log,\alpha 2,\rho 1)$	1.000	.990	1.000	.942
$(t,\alpha 1,\rho 2)$	1.000	1.000	1.000	1.000
$(\log,\alpha 1,\rho 2)$	1.000	.990	1.000	.944
$(t,\alpha 2,\rho 2)$	1.000	1.000	1.000	1.000
$(\log,\alpha 2,\rho 2)$	1.000	.990	1.000	.942