The Economic and Social Review, Vol. 12, No. 4, July 1981, pp. 235-252

A Comparison of the Bounds, Beta-approximate, and Exact Variants of Two Tests for Heteroscedasticity based on Ordinary Least Squares Residuals

M. J. HARRISON*
Trinity College, Dublin

Précis: This paper compares the small sample empirical size, power and incidence of inconclusiveness of the bounds tests for heteroscedasticity proposed by Szroeter (1978) and Harrison and McCabe (1979). It also examines the performance of the exact and beta-approximate variants of the tests. Probabilities are computed numerically using both simulated and actual data and various heteroscedasticity specifications. No consistent power superiority of either test is found, although for the types of heteroscedasticity most commonly postulated in applied economics, Szroeter's test is the more powerful. On the other hand, Szroeter's test suffers from the higher incidence of inconclusiveness in all of the cases examined. Two-moment beta-approximations perform well compared with the exact tests. An example of the use of both bounds tests is given.

I INTRODUCTION

Szroeter (1978) has recently proposed, amongst other things, a parametric bounds test for heteroscedasticity in linear regression models with nonstochastic regressors. A similar test has been developed by Harrison and McCabe (1979). Both tests are small sample tests based on the direct use of the ordinary least squares residuals from a single regression on the complete sample of observations. Their computational simplicity, which compares with that of the Durbin-Watson bounds test for autocorrelation (1950, 1951), exceeds that of other recent tests for heteroscedasticity, such as the likelihood ratio test of Harvey (1976) and the Lagrange multi-

*The research for this paper was undertaken during the author's tenure of a Visiting Research Fellowship in the Department of Econometrics and Social Statistics, University of Manchester. The author is most grateful to two anonymous referees for their valuable comments on an earlier draft of the paper, and to B. J. Whelan and G. Keogh, the Economic and Social Research Institute, Dublin, for making available the data on which the example in Section II is based.

plier test of Godfrey (1978),[1] and makes them most attractive practical procedures. However, little is known about their relative efficacy in the kinds of situation likely to be encountered in practice. Harrison and McCabe briefly considered the problem of inconclusiveness of their test and its small sample power compared with that of the tests of Goldfeld and Quandt (1965), Theil (1971, pp. 214-218), and Harvey and Phillips (1974). Szroeter derived an asymptotic power function for his class of tests; Harrison (1980) has since examined the problem of inconclusiveness of the Szroeter bounds test in small samples. Yet no direct study of the comparative performance of the two bounds procedures has been undertaken. Similarly, little is known about the relative performance of associated procedures for use in the event of the bounds tests being inconclusive.

The main purpose of this paper is to compare the empirical size, the power against several specific forms of heteroscedasticity, and the incidence of inconclusiveness of the Szroeter (S) and Harrison-McCabe (HM) bounds tests in a variety of small sample circumstances. The small sample performances of the exact and beta-approximate variants of the two tests are also examined. It is hoped that the results may be of value to applied economists who may wish to use a simple bounds procedure to effect a test for heteroscedasticity.

The organisation of the paper is as follows. Section II briefly describes the two bounds procedures and illustrates their application using data from a recent Irish demographic study. The relative performance of the bounds tests is examined in Section III. The exact and beta-approximate variants of the tests are assessed in Sections IV and V, respectively. Section VI contains a number of conclusions.

II THE S AND HM TESTS

Consider the problem of testing the null hypothesis H_O of homoscedasticity against an alternative hypothesis H_A of heteroscedasticity in the familiar context of the general normal linear regression model (see Theil 1971, Sec. 5.4). Let the n sample observations be ordered such that, under H_A, the disturbance variances σ_i^2, i = 1, 2, . . ., n, satisfy $\sigma_{i-1}^2 \leqslant \sigma_i^2$, i = 2, 3, . . ., n. Then the form of the S test statistic used in this study is

$$h = 2 \sum_{i=1}^{n} h_i e_i^2 \Big/ \sum_{i=1}^{n} e_i^2 , \qquad (1)$$

1. While these likelihood ratio and Lagrange multiplier tests are also based on the ordinary least squares residuals, they are strictly large sample procedures. However, on the evidence presented by Godfrey (1978) and Breusch and Pagan (1979), the Lagrange multiplier test appears to perform fairly well even with quite small samples.

where $h_i = 1 - \cos[i\pi/(n+1)]$, $i = 1, 2, \ldots, n$, and the e_i, $i = 1, 2, \ldots, n$, are the ordinary least squares regression residuals. One other form for the S test statistic is available based on a somewhat less straightforward definition of the h_i. The simpler specification (1) was chosen with the practitioner in mind; it also accords with S's own choice of statistic in his worked example (see Szroeter 1978, p. 1317). The distribution of h, which is dependent on the observations on the regressors, is bounded below and above by the distributions of the variables $h_L = 4 - d_U(n+1, k+1)$ and $h_U = 4 - d_L(n+1, k+1)$, respectively, where d_L and d_U are the well-known lower and upper bounding Durbin-Watson variables (see Szroeter 1978, Sec. 4).[2] Thus the one-sided S bounds test criterion is to reject H_O if $h > h_U^\alpha$, and accept if $h < h_L^\alpha$, where h_L^α and h_U^α denote the bounding 100α per cent critical values; otherwise the test is inconclusive.

The form of the HM test statistic used in the study is

$$b = \sum_{i=1}^{m} e_i^2 \Big/ \sum_{i=1}^{n} e_i^2, \tag{2}$$

where $m = [n/2]$, i.e., the integer part of n/2. The choice of m, like the choice of the number of omitted observations in the Goldfeld-Quandt test, is an important practical consideration having implications for the power of the HM test. The present choice accords with HM's suggestion for situations in which, as is common in practice, H_A does not postulate the precise form of the heteroscedasticity. As HM (1979, sec. 2.1) show, b is bounded below and above by the beta-distributed variables $b_L(\frac{m-k}{2}, \frac{n-m}{2})$ and $b_U(\frac{m}{2}, \frac{n-m-k}{2})$, respectively. Thus, for the 100α per cent significance level, the one sided HM bounds test criterion is to reject H_O if $b < b_L^\alpha$, and accept if $b > b_U^\alpha$, where b_L^α and b_U^α denote the bounding beta critical values; otherwise the test is inconclusive.

The ease of application of both bounds tests is illustrated in the following example.

Example

The example relates to a recent study by Whelan and Keogh (1980) of the relationship between the population and the number of registered electors in Irish counties in census years. Conscious of the potential problem of heteroscedasticity due to the great variation in population across counties,

2. Readers familiar with standard Durbin-Watson theory may be puzzled by the use of i and n + 1 instead of i − 1 and n, in the definition of h_i; and by the use of n + 1 and k + 1, instead of n and k, in the choice of d_L and d_U. The explanation for this is to be found in Szroeter (1978, Proposition 3.1, p. 1314).

Whelan and Keogh adopted an equation specification in which the ratio of county population and Electoral Register is the dependent variable. On the assumption that this ratio is constant in any given county or county borough, but may vary across counties, in the census years covered by the data, the right-hand side of the equation contained an intercept and 30 county dummy variables. The equation was estimated by stepwise ordinary least squares regression, using the SPSS package and observations on 31 counties and county boroughs from 3 censuses. The overall result was statistically highly significant, with a coefficient of determination of 0.95. However, 8 of the county coefficients were not significantly different from zero at the 5 per cent level,[3] and the dummy variables associated with these were excluded from the version of the equation actually used by Whelan and Keogh in estimating county populations in intercensal years.

Despite their awareness of the possibility of heteroscedasticity, Whelan and Keogh did not test for the problem, either before choosing the ratio specification or after estimating their preferred equation. A test for heteroscedasticity in their chosen model is undertaken here, using the ordinary least squares residuals from their final regression; both the S and HM bounds procedures are employed. From Whelan and Keogh's regression output, $n = 93$, $k = 23$ and the sum of squares of the least squares residuals $\sum_{i=1}^{93} e_i^2 = 0.03989$. The latter number is required for the denominator when calculating the sample value of both the S and HM test statistics. Before the numerators of the statistics are computed, however, the residuals require to be ordered. Since the use of the ratio of county population and Electoral Register for the dependent variable may be viewed as a form of "correction" of the data for heteroscedasticity based on the assumption that the disturbance variance is a function of the size of the Electoral Register, the residuals were ordered according to the magnitude of the Electoral Register for the present tests.

Using Equation (1), the numerator of the S statistic, h, was then computed as $2 \sum_{i=1}^{93} h_i e_i^2 = 2 \times 0.04247 = 0.08494$. Thus the sample value of h is $0.08494/0.03989 = 2.1294$. Unfortunately, the values for $d_L(n+1, k+1) = d_L(94, 24)$ and $d_U(n + 1, k + 1) = d_U(94, 24)$, which are required to determine the critical values h_U^α and h_L^α, respectively, are not tabulated, even in the extended Durbin-Watson tables of Savin and White (1977). However, extrapolating from Savin and White's tables for $\alpha = 0.05$, which cater for values of k up to 20, $\hat{d}_L^{0.05}(94, 24) = 1.123$ and $\hat{d}_U^{0.05}(94, 24) = 2.274$ are thought to be

3. After scrutiny of their regression output, the present author is of the opinion that Whelan and Keogh's (1980, p. 305) statement that nine of the county coefficients are insignificant, is in error.

reasonable estimates.[4] Using these values, $\hat{h}_L^{0.05} = 4 - \hat{d}_U^{0.05}(94,24) = 4 - 2.274 = 1.726$, and $\hat{h}_U^{0.05} = 4 - \hat{d}_L^{0.05}(94, 24) = 4 - 1.123 = 2.877$. Since the sample value of h lies between these critical values, the S bounds test is inconclusive at the 5 per cent level of significance.

Similarly, using Equation (2), the numerator of the HM statistic, b, was computed as $\sum_{i=1}^{m} e_i^2 = 0.01711$, where $m = [n/2] = 46$. Therefore the sample value of b is 0.01711/0.03989 = 0.4289. The critical values $b_L^{0.05}$ and $b_U^{0.05}$ were obtained from standard F tables, which are more readily available than beta tables, with the aid of the transformation given in HM (1979, Sec. 2). Specifically,

$$b_L^{0.05} = \left[1 + \frac{(n-m)F^{0.05}(n-m, m-k)}{m-k}\right]^{-1} = \left[1 + \frac{47 \times F^{0.05}(47, 23)}{23}\right]^{-1} = \left[1 + \frac{47 \times 1.901}{23}\right]^{-1} = 0.205; \text{and}$$

$$b_U^{0.05} = \left[1 + \frac{(n-m-k)F^{0.05}(n-m-k, m)}{m}\right]^{-1} = \left[1 + \frac{24 \times F^{0.05}(24, 46)}{46}\right]^{-1}$$

$$= \left[1 + \frac{24 \times 1.750}{46}\right]^{-1} = 0.523.$$

As in the case of the S test, the computed value of b lies between the appropriate critical values and, therefore, the HM bounds test is also inconclusive at the 5 per cent significance level.

The fact that both bounds tests are inconclusive, and therefore provide no basis for accepting or rejecting the null hypothesis of homoscedasticity, is not altogether surprising given the large value of k. Even with large n, such large values of k inevitably give rise to large inconclusive regions. However, lest it be thought that these inconclusive test results, though illustrative, cast doubt on the efficiency of Whelan and Keogh's ordinary least squares estimates, it should be pointed out that on further testing, the null hypothesis is accepted at the 5 per cent significance level. For example, using the simple beta approximation to the true distribution of Szroeter's h statistic proposed by Harrison (1980, Sec. V), the 5 per cent critical value of h is found to be 2.334. Since the sample value of h = 2.1294 is less than 2.334, the null hypothesis of homoscedasticity is accepted.

4. In practice, values of k as large as that used by Whelan and Keogh are rare. If required, exact Durbin-Watson values could be calculated numerically, of course, but this complicated exercise was not felt to be warranted for the purposes of this illustration, particularly as the HM test can be applied using published tables.

As well as demonstrating their simplicity in application, this example has shown clearly the possibility of inconclusiveness of both bounds procedures. This matter, and others, is examined more systematically in the following section.

III COMPARISONS OF THE S AND HM BOUNDS TESTS

Model and Methodology

To facilitate comparisons of the two bounds tests a simple regression model was postulated involving an intercept and single explanatory variable, i.e.,

$$y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, 2, \ldots, n, \tag{3}$$

where y_i, X_i and u_i denote the i^{th} observation on the dependent variable, independent variable and unknown disturbance, respectively, and β_0 and β_1 are unknown parameters. Thus, including the unit dummy variable associated with the intercept, $k = 2$. The u_i are assumed to be $N(0,\sigma_i^2)$ variables.

Both cross-section and time-series observations on the explanatory variable were used, some of the data in each category being artificially generated and some being actual economic data. In the cross-section category, observations were randomly generated on a uniformly distributed variable with range (0, 20) by the method of Downham and Roberts (1967); on a normally distributed variable with mean 10 and standard deviation 10 by the method of Marsaglia and Bray (1964); and on a lognormally distributed variable with coefficient of variation 1 by exponentiation of random normal numbers. In each of these cases, values for n of 10, 20, 30 and 40 were used. In addition, two sets of actual cross-section data were used, namely, the observations on output and employment in 28 UK industrial groupings in 1968 from Stone's *A Programme for Growth* (1974, Table 26, p. 129 and Table 35, p. 135).

In the time-series data category, the artificial observations used were those on the pure trend variable $i = 1, 2, \ldots, n$, the values used for n again being 10, 20, 30, and 40. Three actual economic time-series were used, namely, the final 16 annual observations on real income *per capita* in The Netherlands from Theil's "textile" example (1971, Table 3.1, p. 102), the final 20 annual observations on real income *per capita* in the UK from Durbin and Watson's "spirits" data (1951, Table 1, p. 160), and 40 quarterly observations (second quarter of 1966 to first quarter of 1977, inclusive) on the index of industrial production in the Republic of Ireland *(Irish Statistical Bulletin)*. As Dubbelman *et al.*, (1978) have pointed out, when time series data are being used, the powers of statistical procedures, such as the S and HM tests, may be influenced considerably by the characteristics of the series; and often, simulated time-series data may not possess the typical characteristics of economic time series. Hence the emphasis on actual economic time series here. However, it still seemed worthwhile to calculate the values of the indicators of the typicality of time series proposed by Dubbelman *et al.* Thus, for each of the time series used, the quantities $\tau(s) = \mathrm{tr}[X'A^sX(X'X)^{-1}]$, where X is the $n \times k$ matrix of observations on the regressors and A is the familiar first-differencing matrix of order $n \times n$, were evaluated for s = 1, 2, 3, and 4 (see Dubbelman *et al.*, 1978, p. 300). Of these $\tau(s)$ values, which are available on request from the author, most were small; only 4 exceeded unity, and of those, only 1 exceeded 3.0, namely, $\tau(4) = 11.463$, for the Irish industrial production series. All were of the low order of magnitude expected of representative economic time series.

The following disturbance variance structures were used

$$\left.\begin{aligned} H_O &: \sigma_i^2 = \sigma^2, \\ H_{A1} &: \sigma_i^2 = \sigma^2 X_i, \\ H_{A2} &: \sigma_i^2 = \sigma^2 X_i^2, \\ H_{A3} &: \sigma_i^2 = \sigma^2 i, \\ H_{A4} &: \sigma_i^2 = \sigma^2 i^2, \end{aligned}\right\} \quad i = 1, 2, \ldots, n,$$

and

$$H_{A5} : \sigma_i^2 = \begin{cases} \sigma^2 & i = 1, 2, \ldots, m \\ \sigma_*^2 & i = m+1, m+2, \ldots, n, \end{cases}$$

where σ^2 and $\sigma_*^2 (\sigma^2 < \sigma_*^2)$ are constants. Since h and b are each independent of the scale of the u_i, the value for σ^2 was, without loss of generality, taken to be unity; σ_*^2 was given the values 2, 4, and 8. The pairs of alternatives H_{A1} and H_{A3}, and H_{A2} and H_{A4}, are equivalent in the case of the trend variable, of course.

Thus 160 different combinations of explanatory variable, sample size, and disturbance variance specification were used. For each of these, and given numerical values for β_0 and β_1 and appropriately generated disturbances, Equation (3) defines the corresponding sets of values for the dependent variable. However, it was not necessary to derive y_i values in this way, as it would have been had the study to rely on Monte Carlo methods. Simulation techniques were unnecessary. Rather, the probabilities of rejecting H_O using the S and the HM tests were calculated accurately, for both the lower and upper bounds of the tests, using the numerical integration technique due to Imhof (1961), which requires only the observations on the explanatory variables and the numerical form of the disturbance variance-covariance matrix for its application. Throughout the study, one-sided tests at the $100\alpha = 5$ per cent level of significance were used, critical values being obtained from Savin and White's (1977) Durbin-Watson tables in the case of the S test, and Pearson's (1968) beta tables in the case of the HM test. The probability calculations were performed in the manner described in HM (1979, Sec. 3) using the method of Imhof (1961). The main computer subroutine used was that of Koerts and Abrahamse (1969, pp. 155-160) for which the truncation and integration errors were set at 1.0×10^{-4}. The eigenvalue calculations required for the implementation of the Koerts and Abrahamse algorithm were performed using the NAG subroutine F02AAF (1977). The computer used was the CDC 7600 at the University of Manchester Regional Computer Centre.

Results

A representative selection of results is given in Table 1; the results not reported are available on request from the author.

Before commenting on the results in Table 1, it may be useful to clarify their meaning. Columns 1 to 4 in the body of the table give the theoretically calculated probabilities of certain events under H_O; Columns 5 to 8, 9 to 12, and 13 to 16 give the calculated probabilities of events under H_{A1}, H_{A2}, and H_{A5}, respectively. For example, each number in Column 1 gives the probability, $Pr(h > h_L^\alpha | H_O)$, that the sample value of Szroeter's h will exceed the $\alpha = 0.05$ critical value of h_L when H_O is true. Now, under the S procedure, H_O is accepted if $h < h_L^\alpha$, where $Pr(h_L < h_L^\alpha | H_O) = 1 - \alpha$. Since $h > h_L$, $Pr(h < h_L^\alpha | H_O) < 1 - \alpha$; therefore $Pr(h > h_L^\alpha | H_O) > \alpha$. Note also that $Pr(h > h_L^\alpha | H_O) = Pr(h_L^\alpha < h \leqslant h_U^\alpha | H_O) + Pr(h > h_U^\alpha | H_O)$. Each number in Column 2 gives the probability, $Pr(h > h_U^\alpha | H_O)$ that h will exceed the α critical value of h_U when H_O is true. The S procedure rejects H_O if $h > h_U^\alpha$, where $Pr(h_U > h_U^\alpha | H_O) = \alpha$. Therefore, since $h < h_U$, $Pr(h > h_U^\alpha | H_O) < \alpha$. A similar argument applies to the entries in Columns 3 and 4, which give $Pr(b < b_L^\alpha | H_O) < \alpha$ and $Pr(b < b_U^\alpha | H_O) = Pr(b_U^\alpha > b \geqslant b_L^\alpha | H_O) + Pr(b < b_L^\alpha | H_O) > \alpha$, respectively. In the discussion that follows these pairs of probabilities are referred to as the actual or bounding sizes of the S and HM bounds tests. For a given set of sample circumstances, the difference between the bounding sizes of each test yields the probability that the test will be inconclusive under H_O.

Further, each number in Column 5 gives the probability, $Pr(h > h_L^\alpha | H_{A1}) = Pr(h_L^\alpha < h \leqslant h_U^\alpha | H_{A1}) + Pr(h > h_U^\alpha | H_{A1})$, that the S test statistic h will exceed h_L^α when H_{A1} is true, while each number in Column 6 gives the probability, $Pr(h > h_U^\alpha | H_{A1})$, that h will exceed h_U^α when H_{A1} is true. Likewise, Columns 7 and 8 give the calculated values of $Pr(b < b_L^\alpha | H_{A1})$ and $Pr(b < b_U^\alpha | H_{A1}) = Pr(b_U^\alpha > b \geqslant b_L^\alpha | H_{A1}) + Pr(b < b_L^\alpha | H_{A1})$, respectively, for the HM test. On the basis of reasoning similar to that just outlined for the H_O case, these pairs of values are referred to as the bounding powers of the tests against H_{A1}; the differences between associated bounding powers are the

Table 1: *Powers of the S and HM bounds tests at the 0.05 significance level*

Variable	n	H_O S h_L	H_O S h_U	H_O HM b_L	H_O HM b_U	H_{A1} S h_L	H_{A1} S h_U	H_{A1} HM b_L	H_{A1} HM b_U	H_{A2} S h_L	H_{A2} S h_U	H_{A2} HM b_L	H_{A2} HM b_U	H_{A5} (σ_*^2=4) S h_L	H_{A5} (σ_*^2=4) S h_U	H_{A5} (σ_*^2=4) HM b_L	H_{A5} (σ_*^2=4) HM b_U
a. Cross-section data																	
Lognormal	10	.187	.005	.004	.096	.501	.059	.025	.296	.776	.250	.125	.596	.554	.039	.045	.449
	20	.133	.015	.019	.092	.453	.113	.109	.323	.776	.367	.333	.626	.716	.255	.427	.732
	30	.102	.018	.023	.076	.496	.209	.174	.357	.860	.626	.529	.732	.805	.471	.678	.853
	40	.092	.022	.027	.074	.749	.493	.442	.632	.991	.956	.929	.972	.888	.670	.840	.931
Normal	10	.211	.006	.005	.110	.672	.105	.061	.460	.883	.338	.241	.747	.592	.046	.051	.489
	20	.149	.018	.022	.099	.560	.162	.160	.405	.846	.444	.415	.689	.740	.281	.446	.743
	30	.114	.021	.026	.083	.518	.216	.181	.371	.839	.572	.482	.690	.823	.497	.692	.861
	40	.100	.025	.029	.078	.871	.653	.584	.754	.997	.977	.951	.981	.897	.689	.847	.934
Uniform	10	.251	.008	.008	.134	.860	.151	.214	.705	.964	.335	.525	.883	.671	.055	.085	.552
	20	.134	.016	.019	.090	.733	.325	.297	.589	.975	.819	.795	.934	.717	.272	.430	.734
	30	.108	.020	.025	.081	.862	.593	.596	.789	.998	.974	.980	.994	.824	.503	.701	.865
	40	.103	.026	.030	.080	.907	.714	.633	.787	.998	.980	.950	.980	.900	.695	.849	.935
Stone's	28	.090	.013	.019	.072	.939	.751	.838	.877	1.000	.993	.993	.999	.760	.376	.609	.821
b. Time-series data																	
Trend	10	.247	.008	.007	.131	.585	.051	.040	.382	.813	.159	.138	.638	.652	.054	.073	.538
	20	.141	.017	.021	.095	.700	.275	.266	.549	.948	.689	.670	.872	.730	.276	.443	.743
	30	.113	.022	.026	.082	.816	.679	.467	.679	.990	.923	.888	.957	.828	.508	.700	.865
	40	.100	.025	.029	.078	.894	.695	.623	.781	.998	.985	.964	.967	.897	.688	.847	.934
Theil's	16	.164	.015	.018	.104	.319	.045	.048	.220	.512	.111	.117	.392	.699	.193	.319	.681
Durbin-Watson's	20	.136	.016	.019	.092	.196	.029	.030	.129	.271	.049	.045	.176	.712	.257	.422	.729
Irish Industrial Production	40	.103	.025	.030	.080	.264	.091	.090	.196	.494	.235	.210	.375	.899	.692	.847	.935

Note: The results for Stone's variable in this and the following tables relate to the output data; the results for the employment data are very similar.

probabilities of inconclusiveness of the tests under H_{A1} in the various sample situations. The remaining columns contain similar information on the bounding powers of the two tests against H_{A2} and H_{A5}.

In accordance with theoretical expectations, the actual sizes of the bounds of both tests, given by the results for H_O, bound the nominal test size of $\alpha = 0.05$ in all cases. The sizes of h_U and b_L are similar, but, without exception, the sizes of h_L are greater than the corresponding sizes of b_U, and in several cases exceed 4 times the nominal size of the test. The sizes associated with b_U rarely exceed 2½ times the nominal value. Consequently, the differences between the bounding sizes are greater in the case of the S test than in the case of the HM test. In the case of both tests, however, these differences decrease as sample size increases, with both bounding sizes approaching 0.05. More specifically, as sample size increases from 10 to 40, the differences decrease from about 0.20 to about 0.07 for the S test, and from about 0.12 to about 0.05 for the HM test.

The power of both tests, for given explanatory variable and sample size, varies directly with the degree of heteroscedasticity, being considerably higher for H_{A2} and H_{A4} than for H_{A1} and H_{A3}, and, in the case of H_{A5}, for $\alpha_*^2 = 8$ than for $\alpha_*^2 = 2$. Similarly, for a given variable and a given H_A, the power of both tests varies directly with sample size. However, there are no systematic variations in the power of the tests with type of variable, *ceteris paribus*. None the less, it seems noteworthy that for H_{A1} and H_{A2}, which are commonly used alternatives in applied econometrics, the power of both tests is markedly higher with the actual than with the simulated cross-section data, but lower with the actual time-series data than with the pure trend variable.

Concerning relative performance, the S test appears to be more powerful than the HM test against alternatives H_{A1} to H_{A4}, but, with the exception of cases in which $n = 10$, less powerful than the HM test against H_{A5}. This result appears to provide some confirmation, for small samples, of Szroeter's asymptotic result that power increases with the correlation between the h_i and σ_i^2 (Szroeter 1978, Sec. 6). However, for any given set of circumstances in Table 1, it is difficult to draw accurate conclusions about the relative power of the two bounds tests because of the differences in their actual sizes. To circumvent this problem, the powers of the exact variants of the S and HM tests are examined in Section IV.

As has been stated, an indication of the incidence of inconclusiveness of the bounds tests is provided by the differences between their bounding powers. As in the H_O case, these differences are consistently larger for the S test than for the HM test in all of the H_A cases considered, suggesting a higher likelihood of inconclusiveness for the S than for the HM test. Moreover, while inconclusiveness declines with increasing sample size for both tests, the results suggest that its incidence may remain quite large even for n as large as 40. For n = 10, inconclusiveness varies from about 0.3 to over 0.6 in the case of the S test, and from about 0.2 to 0.5 in the case of the HM test; for n = 40, it may still be as high as 0.3 in the case of the S test and 0.2 in the case of the HM test. Only when the degree of heteroscedasticity is high, as in the case of H_{A2}, H_{A4} and $H_{A5}(\sigma_*^2 = 8)$, is the incidence of inconclusiveness of both tests small for the larger sample sizes. This finding suggests that the availability of supplementary procedures for use when the bounds tests are inconclusive is a matter of considerable practical importance. Besides exact tests, certain approximations are feasible. The performances of two-moment beta-approximations to the distributions of h and b are examined in Section V.

IV COMPARISON OF THE EXACT S AND HM TESTS

The scope of the comparisons for the exact variants of the tests was extended in two ways. First, in view of Szroeter's asymptotic power function result, a third test, whose construction is dependent on the H_A in question, was included. We may refer to this as the "generalised" Szroeter (GS) test and define it in a similar fashion to h, but with $h_i = \sigma_i^2$, $i = 1, 2, \ldots, n$. For the purposes of this study, the h_i in the GS test for alternatives H_{A1} to H_{A4} were "normalised" by appropriate scaling to yield a test statistic with range (0, 1).[5] Second, an additional alternative hypothesis was considered, namely,

$$H_{A6} \;:\; \sigma_i^2 = \begin{cases} \sigma^2, & i = 1, 2, \ldots, r, \\ \sigma_*^2, & i = r+1, r+2, \ldots, n, \end{cases}$$

where r ($\neq$ m) was set at [n/3], and σ_*^2 was given the values 2 and 4.

Exact critical values for each test and data set were obtained by using a search procedure based on the repeated application of Imhof's method under H_O. Values were calculated which yielded a test size of 0.050 correct to 3 decimal places,convergence being achieved in between 5 and 13 iterations per case. Given these values, exact powers were computed using the methodology described in Section III. A selection of results is given in Table 2.

The contents of Table 2 are more straightforward than those of Table 1. For each alternative hypothesis, the column of numbers relating to each test gives the exact powers of that test in the various sample circumstances;

5. Szroeter (1978, p. 1315) has proposed an exact test based on the use of a set of BLUS residuals. However, as this study is concerned only with testing for heteroscedasticity using the OLS residuals, this rather complex BLUS-based procedure was not examined.

Table 2: *Powers of the S, HM, and GS exact tests at the 0.05 significance level*

Variable	n	H_{A1}			H_{A2}			H_{A5} $(\sigma_*^2=4)$			H_{A6} $(\sigma_*^2=4)$		
		S	HM	GS	S	HM	GS	S	HM	GS	S	HM	GS
a. Cross-section data													
Lognormal	10	.246	.188	.235	.555	.467	.460	.235	.303	.303	.131	.138	.194
	20	.251	.217	.248	.580	.504	.497	.483	.612	.612	.237	.229	.449
	30	.355	.279	.344	.768	.660	.667	.670	.796	.796	.401	.361	.707
	40	.638	.555	.587	.981	.958	.910	.805	.900	.900	.502	.425	.839
Normal	10	.349	.301	.315	.660	.610	.525	.233	.307	.307	.127	.136	.184
	20	.310	.274	.332	.638	.562	.588	.481	.609	.609	.234	.227	.451
	30	.345	.279	.361	.712	.599	.646	.667	.794	.794	.395	.358	.705
	40	.771	.678	.776	.991	.970	.961	.805	.899	.899	.499	.422	.842
Uniform	10	.453	.508	.485	.694	.771	.573	.244	.318	.318	.131	.141	.182
	20	.527	.467	.507	.923	.891	.858	.485	.619	.619	.238	.230	.450
	30	.745	.712	.713	.992	.990	.951	.680	.803	.803	.407	.367	.712
	40	.816	.715	.857	.992	.968	.979	.804	.899	.899	.498	.421	.843
Stone's	28	.895	.838	.686	.999	.998	.814	.641	.767	.767	.365	.329	.662
b. Time-series data													
Trend	10	.209	.195	.206	.438	.426	.382	.236	.310	.310	.125	.135	.179
	20	.464	.416	.458	.843	.796	.741	.483	.617	.617	.234	.229	.450
	30	.668	.586	.664	.967	.932	.908	.675	.800	.800	.402	.364	.715
	40	.804	.711	.804	.994	.978	.971	.804	.900	.900	.498	.423	.841
Theil's	16	.124	.121	.125	.253	.247	.251	.401	.522	.522	.208	.214	.358
Durbin-Watson's	20	.081	.074	.092	.125	.105	.127	.474	.609	.609	.232	.225	.447
Irish Industrial Production	40	.153	.136	.158	.342	.287	.345	.802	.898	.898	.495	.420	.843

i.e., the probability that that test will lead to rejection of H_O when H_O is false and the particular alternative is true. Thus, for example, each number in Column 1 in the body of the table gives the probability $\Pr(h > h_E^{\alpha} | H_{A1})$, where h_E denotes the exact 100α per cent critical value of h; and similarly for the numbers in the other columns.

The exact powers confirm the impression given by the results in Table 1 that the S test is more powerful than the HM test against alternatives H_{A1} to H_{A4}. In the case of H_{A5}, the HM exact test, unlike the HM bounds test, is more powerful than the corresponding S test for all sample sizes. The exact GS test is not in general superior to the exact S test for H_{A1} to H_{A4}, but it is consistently more powerful than it in the case of H_{A5}.[6] The exact HM test under H_{A5} is equivalent to the exact GS test, of course, and both yield identical powers. In the case of H_{A6}, however, both the S and HM exact tests are inferior to the GS exact test, especially for the larger sample sizes. Moreover, while the HM test is more powerful than the S test for n = 10 under H_{A6}, it is less powerful than the S test for $n > 20$.

V COMPARISON OF THE BETA-APPROXIMATE S AND HM TESTS

For each data set, beta-approximate critical values were computed for the HM test using the method described in HM (1979, Sec. 2.2.). A similar procedure, using the correct range of h, was used to obtain beta-approximate critical values for the S test.[7] With these values, probability calculations were carried out for the tests under H_O. Table 3 contains both the critical values and the test sizes for the two beta approximations. For comparative purposes, the exact 0.05 critical values are also included in the table. The h_A and b_A columns contain the beta approximations to h_E and b_E, respectively. The S size gives the probability, $\Pr(h > h_A | H_O)$, that the sample value of h will exceed h_A when H_O is true, while the HM size gives the probability, $\Pr(b < b_A | H_O)$, that the sample value of b will be less than b_A when H_O is true.

6. The inferior performance of the exact GS test *vis-à-vis* the exact S test for H_{A1} to H_{A4}, inclusive, is puzzling because it conflicts with expectations based on Szroeter's asymptotic power function. Yet it conforms to an alternative, but as yet unpublished, theoretical result on the asymptotically most powerful form of the S statistic obtained in correspondence from G. Bornholt, Department of Economic Statistics, University of Sydney. Together with Bornholt's result, the numerical results presented here might be viewed as casting some doubt on Szroeter's asymptotics, though to date the present author has not found any error.

7. While such a procedure was not suggested by Szroeter, it would seem entirely appropriate given that h is a Durbin-Watson variable. Following Durbin and Watson's (1951) suggestion, beta distributions have been extensively used to approximate the distribution of the Durbin-Watson statistic. See, for example, Theil and Nagar (1961), Henshaw (1966) and Durbin and Watson (1971). Incidentally, the correct range of h is given by the largest and smallest non-zero eigenvalues of the matrix appearing in the numerator of h when the statistic, defined in (1), is written in matrix terms as a ratio of quadratic forms in the true disturbances. For further details see Henshaw (1966, Sec. 3).

Table 3: *Critical values and sizes of the Beta-Approximate S and HM tests for the 0.05 significance level*

Variable	n	S			HM		
		Critical value		*Size*	*Critical value*		*Size*
		h_E	h_A	h_A	b_E	b_A	b_A
		a. Cross-section data					
Lognormal	10	2.813	2.785	.096	.178	.172	.046
	20	2.676	2.613	.069	.246	.246	.050
	30	2.557	2.556	.050	.294	.293	.049
	40	2.492	2.498	.048	.320	.320	.050
Normal	10	2.856	2.749	.075	.166	.160	.046
	20	2.703	2.610	.080	.240	.240	.050
	30	2.578	2.538	.064	.289	.289	.050
	40	2.506	2.488	.056	.316	.317	.050
Uniform	10	2.912	2.693	.114	.143	.141	.049
	20	2.684	2.657	.057	.247	.247	.050
	30	2.570	2.565	.052	.290	.290	.050
	40	2.512	2.512	.059	.315	.315	.050
Stone's	28	2.547	2.599	.036	.294	.293	.049
		b. Time series data					
Trend	10	2.906	2.688	.112	.147	.144	.048
	20	2.691	2.635	.067	.243	.243	.050
	30	2.578	2.552	.058	.289	.291	.052
	40	2.506	2.491	.055	.317	.317	.050
Theil's	16	2.764	2.698	.068	.213	.212	.050
Durbin-Watson's	20	2.682	2.635	.063	.246	.246	.050
Irish Industrial Production	40	2.512	2.484	.060	.315	.316	.050

Note: The subscripts E and A denote the exact and beta-approximate variants, respectively.

In the case of both tests, the beta-approximate critical values are, without exception, closer to the exact critical values than to either of the bounding critical values. This is reflected by the fact that the empirical sizes of the approximations are closer to 0.05 than to the corresponding bounding sizes given in Table 1. Clearly, the powers of the approximations against the various H_A's would likewise be closer to the exact than to the bounding powers of the tests.

Comparing the two tests, the HM beta-approximate critical values and sizes are, in general, closer to the exact critical values and the actual size of the test, respectively, than those of the beta-approximate S test. This generally superior accuracy is no doubt due to the fact that, in the case of the HM test, only k of the eigenvalues necessary for the approximation are not known to be zero or one, whereas in the case of the S test all of the eigenvalues used for the approximation are unknown. For it is the fact that the corresponding eigenvalues are all ones or zeros in the case of the HM bounds, that results in the bounds being exactly beta distributed.

VI CONCLUDING REMARKS

Three main findings have emerged from this study. First, for the range of circumstances examined, there is no consistent power superiority of either of the tests over the other; the relative power performance depends on the nature of the heteroscedasticity postulated in the alternative hypothesis. However, for the alternatives commonly used in applied econometrics (H_{A1} and H_{A2}), the S test is more powerful than the HM test. The HM test is the more powerful procedure against the H_{A5} type of alternative which is more common in other areas of applied statistics. Second, both bounds tests are likely to be characterised by a high incidence of inconclusiveness in the kinds of small sample situation typically encountered in practice, although the HM bounds test is somewhat less prone to inconclusiveness than the S bounds test, *ceteris paribus*. Therefore in practice, the choice between the two bounds tests will generally involve a trade-off between considerations of power and inconclusiveness. Third, when a two-moment beta approximation is used, greater accuracy is generally obtained in the case of the HM than in the case of the S test. This is not altogether surprising given that the HM bounding statistics are themselves beta variables and all but k of the eigenvalues used in the approximation to the true distribution of b are known to be ones and zeros. In the case of both tests, however, the performance of the beta approximation accords more with that of the exact test than with that of either of the bounding statistics. On the other hand, the two-moment beta-approximations used in this study are relatively complex; they offer little saving in computational cost over the exact variants of the tests, and therefore are not attractive for practical purposes. In the event of inconclusiveness, other, considerably simplified, beta-approximations have been suggested (see, e.g., Harrison 1980).

The results on which these conclusions are based relate to specific sample situations, of course, and no claim to generality is made for them. In particular they relate to a regression model with only two explanatory variables, including the dummy variable unity to account for the intercept. When the number of explanatory variables is larger, one would expect that the bounds tests would have larger regions of inconclusiveness, and the approximate tests may be poorer. For example, comparing the case in which $n = 20$ and $k = 2$ with that in which $n = 20$ and $k = 5$, the inconclusive interval for the S test at the 5 per cent level increases from (2.462, 2.875) to (2.036, 3.171), and that of the HM test increases from (0.193, 0.289) to (0.095, 0.375). The speculation that, for given n, beta approximations may be poorer the larger is k, is based on the previously mentioned fact that, in the HM case, k of the eigenvalues used in the approximation are unknown. Therefore, an increase in k would seem to be tantamount to a loss of information. However, given the range of circumstances investigated, it seems likely that the broader findings would carry over to other situations.

Finally, it may be noted that the findings of this study lead to no change in the view, expressed in Section I, that the S and HM bounds tests are attractive practical procedures. Despite the possibility that both tests may prove to be inconclusive, they are reliable in the sense that the probabilities $\Pr(h > h_U^\alpha)$ and $\Pr(b < b_L^\alpha)$ are less than the significance level α under H_O, and the probabilities $\Pr(h < h_L^\alpha)$ and $\Pr(b > b_U^\alpha)$ are less than $\Pr(h_L < h_L^\alpha)$ and $\Pr(b_U > b_U^\alpha)$, respectively, under H_{A1}, whatever the value of k. Moreover, just as the Durbin-Watson bounds test is a conservative test for autocorrelation, so the S and HM bounds tests are conservative tests for heteroscedasticity in linear regression models. This, together with the fact that they are computationally so simple, suggests that they would at least be useful as first tests for heteroscedasticity. The HM test, which does not require the evaluation of cosine expressions to obtain the sample value of its test statistic, as the S test does, is particularly simple to apply, and therefore would seem eminently suitable for use by researchers.

REFERENCES

BREUSCH, T.S. and A. R. PAGAN, 1979. "A Simple Test for Heteroscedasticity and Random Coefficient Variation", *Econometrica*, Vol. 47, pp. 1287-1293.

DOWNHAM, D. Y. and F. D. K. ROBERTS, 1967. "Multiplicative Congruential Pseudo-Random Number Generators", *Computer Journal*, Vol. 10, pp. 74-77.

DUBBELMAN, S., A. S. LOUTER, and A. P. J. ABRAHAMSE, 1978. "On Typical Characteristics of Economic Time Series and the Relative Qualities of Five Autocorrelation Tests", *Journal of Econometrics*, Vol. 8. pp. 295-306.

DURBIN, J. and G. S. WATSON, 1950. "Testing for Serial Correlation in Least Squares Regression. I", *Biometrika*, Vol. 37, pp. 409-428.

DURBIN, J. and G. S. WATSON, 1951. 'Testing for Serial Correlation in Least Squares Regression. II", *Biometrika*, Vol. 38, pp. 159-178.

DURBIN, J. and G. S. WATSON, 1971. "Testing for Serial Correlation in Least Squares Regression. III", *Biometrika*, Vol. 58, pp. 1-19.

GODFREY, L. G., 1978. "Testing for Multiplicative Heteroscedasticity", *Journal of Econometrics*, Vol. 8, pp. 227-236.

GOLDFELD, S. M. and R. E. QUANDT, 1965. "Some Tests for Homoscedasticity", *Journal of the American Statistical Association*, Vol. 60, pp. 539-547.

HARRISON, M. J., 1980. "The Small Sample Performance of the Szroeter Bounds Test for Heteroscedasticity and a Simple Test for Use when Szroeter's Test is Inconclusive", *Oxford Bulletin of Economics and Statistics*, Vol. 42, pp. 235-250.

HARRISON, M. J. and B. P. M. McCABE, 1979. "A Test for Heteroscedasticity Based on Ordinary Least Squares Residuals", *Journal of the American Statistical Association*, Vol. 74, pp. 494-499.

HARVEY, A. C., 1976. "Estimating Regression Models with Multiplicative Heteroscedasticity", *Econometrica*, Vol. 44, pp. 461-465.

HARVEY, A. C. and G. D. A. PHILLIPS, 1974. "A Comparison of the Power of Some Tests for Heteroscedasticity in the General Linear Model", *Journal of Econometrics*, Vol. 2, pp. 307-316.

HENSHAW, R. C., Jr., 1966. "Testing Single Equation Least Squares Regression Models for Autocorrelated Disturbances", *Econometrica*, Vol. 34, pp. 646-660.

IMHOF, J. P., 1961. "Computing the Distribution of Quadratic Forms in Normal Variables", *Biometrika*, Vol. 48, pp. 419-426.

Irish Statistical Bulletin. Dublin: Central Statistics Office (various issues).

KOERTS, J. and A. P. J. ABRAHAMSE, 1969. *On the Theory and Application of the General Linear Model*. Rotterdam: Rotterdam University Press.

MARSAGLIA, G. and T. A. BRAY, 1964. "A Convenient Method of Generating Normal Variables", *SIAM Review*, Vol. 6, pp. 260-264.

NAG FORTRAN Library Manual, 1977. Mark 6, Vol. 3. Oxford: Numerical Algorithms Group.

PEARSON, K. (ed.), 1968. *Tables of the Incomplete Beta Function*, 2nd Edition. Cambridge: Cambridge University Press.

SAVIN, N. E. and K. J. WHITE, 1977. "The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or many Regressors", *Econometrica*, Vol. 45, pp. 1989-1996.

STONE, R. (ed.), 1974. "Structural Change in the British Economy 1948-1968", Vol. 12 of *A Programme for Growth*, London: Chapman and Hall.

SZROETER, J., 1978. "A Class of Parametric Tests for Heteroscedasticity in Linear

Econometric Models", *Econometrica*, Vol. 46, pp. 1311-1327.

THEIL, H., 1971. *Principles of Econometrics*, Amsterdam: North-Holland Publishing Company.

THEIL, H. and A. L. NAGAR, 1961. "Testing the Independence of Regression Disturbances", *Journal of the American Statistical Association*, Vol. 56, pp. 793-802.

WHELAN, B. J. and G. KEOGH, 1980. "The Use of the Irish Electoral Register for Population Estimation", *The Economic and Social Review*, Vol. 11, pp. 301-318.