ANALYZING UNIFORM RESIDUALS FOR HETEROSCEDASTICITY

R. A. Hester, Jr. and C. P. Quesenberry*

Abstract

We discuss the detection of monotonic heteroscedasticity in normal regression data, using graphical methods and two new, exact tests; and conclude from the results of a systematic power comparison of these tests and those of Goldfeld and Quandt, Theil, Harvey and Phillips, and Harrison and McCabe, that the new test based on NU residuals is somewhat superior, overall. We also use the methods posed here to analyze a data set found in the literature. Both proposed test statistics have exact (null hypothesis) χ^2 distributions, so that exact p-values of tests are readily available for all sample sizes.

KEY WORDS: Monotonic heteroscedasticity; Normal linear regression models; NU and uniform residuals; Exact χ^2 distributions; Invariant transformations; Power estimates.

---

*R. A. Hester, Jr. is a Graduate Student, Department of Statistics, North Carolina State University and C. P. Quesenberry is Professor, Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203.

1. INTRODUCTION AND SUMMARY

In a recent paper, Quesenberry (1983), Q83, proposed analyzing uniform residuals from a general normal linear regression model to detect specification errors in the model for regression data ordered in time. A number of properties of these residuals were discussed, and analysis methods designed for a number of purposes were proposed. In particular, a plot of these residuals against an ordering variable (time) was suggested as a general method of analysis, a Neyman smooth omnibus test of fit was recommended, and methods for detecting change points in time, as well as formulas to evaluate p-values for detecting outliers were given.

In the present paper, we propose some methods for using these residuals to detect a tendency of the variance to increase (decrease) in time - or in any variable that can be used to order the regression data points. We suggest here plotting these residuals against time and watching for certain data patterns that will appear with heteroscedasticity. We also propose two new test statistics for heteroscedasticity. These new tests are based on statistics which have exact chi-squared distributions under the null hypothesis of constant variance, and thus the p-values of the tests can be readily evaluated for all sample sizes.

The powers of these new tests have been compared with those of a number of competing tests in a simulation study. These new tests are found to compare favorably with competing tests from the literature, and one of them appears to be somewhat better overall.

2. THE MODEL AND UNIFORM RESIDUALS

We consider the general linear regression (null hypothesis) model setup:

$\underset{\sim}{y}_n = (y_1, \ldots, y_n)'$, $X_n = \{x_{jk}\}$; $j = 1, \ldots, n$; $k = 1, \ldots, p$

$\underset{\sim}{\varepsilon}_n \sim MVN(0, \sigma^2 I)$, Rank $X_n = p$, $\underset{\sim}{\beta} = (\beta_1, \cdots, \beta_p)'$

$$\underset{\sim}{y}_n = X_n \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}_n \qquad (2.1)$$

The quantities that we call "uniform residuals" were essentially given originally in O'Reilly and Quesenberry (1973), O-Q, and are given explicitly in Q83. We give the formulas again here, since it will be convenient to have them in view in the following development. Let

$\underset{\sim}{x}_j' = (x_{j1}, \ldots, x_{jp})$ denote the jth row of X_n;

X_j is the first j rows of X_n, assume Rank $(X_{p+2}) = p$;

$T_j = (\underset{\sim}{t}_j', S_j^2)$, for $\underset{\sim}{t}_j = X_j' \underset{\sim}{y}_j$, $S_j^2 = \underset{\sim}{y}_j'[I - X_j(X_j'X_j)^{-1}X_j']\underset{\sim}{y}_j$;

$$A_j = \frac{(j-p-1)^{\frac{1}{2}}[y_j - \underset{\sim}{x}_j'(X_j'X_j)^{-1}\underset{\sim}{t}_j]}{\{[1 - \underset{\sim}{x}_j'(X_j'X_j)^{-1}\underset{\sim}{x}_j]S_j^2 - [y_j - \underset{\sim}{x}_j'(X_j'X_j)^{-1}\underset{\sim}{t}_j]^2\}^{\frac{1}{2}}}$$

$G_\nu(\cdot)$ is the Student-t d.f. with ν degrees of freedom;

$u_{j-p-1} = G_{j-p-1}(A_j)$; $j = p+2, \ldots, n$;

$\underset{\sim}{u} = (u_1, \ldots, u_N)'$, $N = n - p - 1$, is the vector of <u>uniform residuals</u>.

In Q83 four properties of the vector $\underset{\sim}{u}$ of uniform residuals under the model assumptions (2.1) were discussed: (1) the components of $\underset{\sim}{u}$ are i.i.d. uniform r.v.'s on (0,1), i.e., are i.i.d. U(0,1) r.v.'s; (2) $\underset{\sim}{u}$ is independent of the complete sufficient statistic $\underset{\sim}{T}_j$; (3) $\underset{\sim}{u}$ is a maximal invariant for a natural class of transformations; and (4) the vector $\underset{\sim}{u}$ contains all information for testing misspecifications in the model. The NU residuals are defined in Q83 as $z_j = \Phi^{-1}(u_j)$; i = 1, ..., N; for Φ a N(0,1) d.f. In the following sections we shall consider analyzing the uniform and NU residuals for heteroscedasticity.

3. ANALYSIS OF UNIFORM RESIDUALS FOR MONOTONE PATTERNS OF HETEROSCEDASTICITY

3.1 INCREASING VARIANCES

We consider first the effect on the vector $\underset{\sim}{u}$ if σ is an increasing function of t (time); since the y_j are ordered in time, this means that $Var(y_1) \leq Var(y_2) \leq \ldots \leq Var(y_n)$. Note first that the ith residual, u_i, is associated with the ith data point $(y_j, \underset{\sim}{x}_j')$ for $j = i + p + 1$, since $u_i = G_i(A_j)$. Consider the formula for A_j, and observe that the denominator of A_j is essentially a mean squared estimator of σ based on the first j y values, that is, on y_j and the other y's observed before it. Therefore, under the alternative hypothesis that σ increases with t, this mean squared estimator would tend to underestimate σ. Thus, A_j will tend to be too large in absolute value, and u_i will tend to be too near either zero or one, with the tendency symmetric on the (0,1) interval.

This will tend to give a splatter pattern of points of the type shown in Figure 1 and the corresponding "union"-shaped density. Thus, if heteroscedasticity is present, the uniform residuals u_i; i = 1, ..., N; should tend to cluster more near zero and

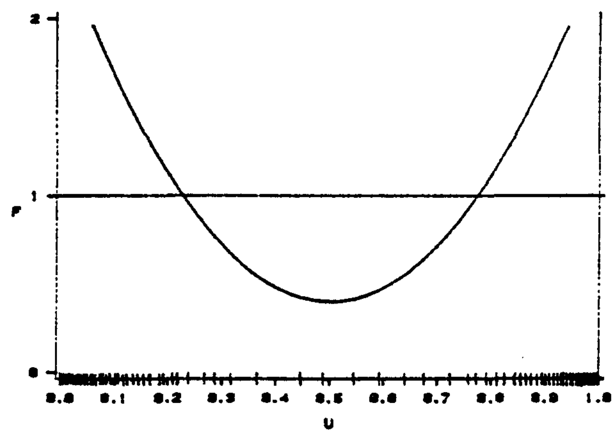


Figure 1: Alternative Hypothesis Splatter Pattern for u_i's and Conceptualized "Density" Shape for Increasing σ.

one than would i.i.d. U(0,1) r.v.'s. This is, of course, only a heuristic argument; however, we shall report below simulation results which do support is validity.

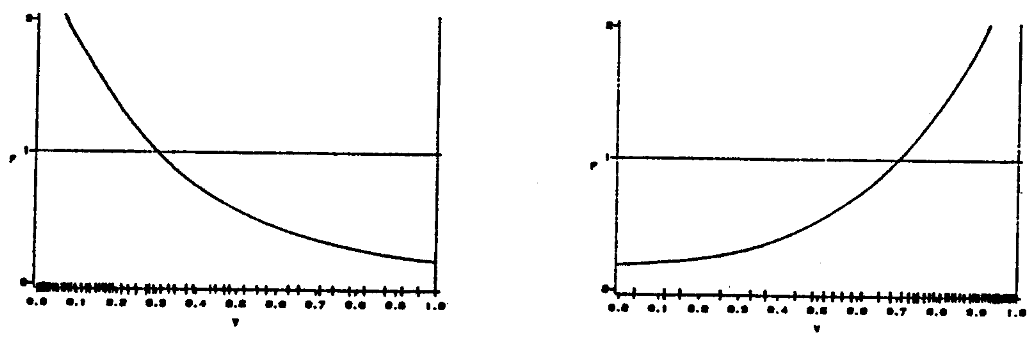
3.1.1. PLOTTING METHODS AND TEST CONSTRUCTION

In view of the above discussion, one obvious method of analyzing for increasing variances is to plot the uniform residuals vs time as recommended in Q83, and watch for the pattern described above to develope. A plot of the NU residuals vs time will also reflect a corresponding pattern, with more observations in the tails of the distribution than would be expected for a N(0,1) distribution.

In order to construct a test with good power against data patterns of the type shown in Figure 1, we proceed as follows. First, we fold the union-shaped distribution of Figure 1 to obtain an inverse J-shaped distribution by the following transformation. Put

$$v_i = 2u_i I_{[0,\frac{1}{2})}(u_i) + 2(1 - u_i) I_{[\frac{1}{2},1]}(u_i);\ i = 1,\ldots,N; \qquad (3.1)$$

for $I_{(a,b)}(u)$ the indicator function of the interval (a,b). Observe that, if the u_i's have a splatter pattern as shown in Figure 1, then the v_i's will have a splatter pattern as in Figure 2a.



(a) σ Increasing (b) σ Decreasing

Figure 2: Alternative Hypothesis Splatter Pattern for v_i's and Conceptualized "Density" Shape for Monotone σ.

Now the statistic $Q_1 = v_1 v_2 \ldots v_n$ has good power for testing against alternative densities of the shape of the density in Figures 2a and 2b - see Miller and Quesenberry (1979), Section 2.3; and under the null hypothesis of constant variance, the random variables $v_1, \ldots, v_N$ are i.i.d. U(0,1), so that $Q = -2 \ln Q_1$

$= -2 \sum_{i=1}^{N} \ln v_i$ is a $\chi^2(2N)$ r.v. We reject the null hypothesis of constant σ for the alternative of increasing σ when small values of Q_1 or large values of Q are observed. The p-value for this test is then $P\{\chi^2(2N) > Q\}$. The statistic Q was proposed by Fisher (1932) for his combined test of significance.

We consider another new test formed as follows. According to the heuristic argument given above, the NU residuals $z_1, \ldots, z_n$ should, under the heteroscedasticity alternative, tend to give too many values in the tails of a $N(0,1)$ distribution. Thus, while the statistic $H^* = \Sigma z_i^2$ is a $\chi^2(N)$ r.v. under the null hypothesis, it will tend to be larger under the heteroscedasticity alternative. Actually, we shall use the statistic $H = \sum_{i=1}^{N} (z_i - \bar{z})^2$, which is a $\chi^2(N-1)$ r.v. under the null hypothesis, and has better power properties than Σz_i^2. For a similar situation in other testing problems, see the remarks in Q83, Quesenberry and Starbuck (1976), and Stephens (1974). The null hypothesis of constant variance is rejected here when $H > \chi^2_\alpha(N-1)$; the corresponding p-value is $P\{\chi^2(N-1) > H\}$.

3.2 DECREASING VARIANCES

We now consider an alternative hypothesis with the variances decreasing in time, that is, Var $(y_1) \geq$ Var $(y_2) \geq$ Var (y_n). By an argument similar to that given above for the case of increasing variances, it is seen that the pattern in the uniform residuals which should appear is one that concentrates many points symmetrically near ½, with few points near zero and one. Similarly, the NU residuals should be more concentrated in the center near zero, with fewer points in the tails than there would be for true i.i.d. N(0,1) observations. Also, the v_i's should have a splatter pattern with points concentrated near one, as shown in Figure 2b.

A test based on the v_i's is thus made by rejecting when Q_1 is large or, equivalently, when Q is small, i.e., for $Q < \chi^2_{1-\alpha}(2N)$; the p-value for this test is $P\{\chi^2(2N) < Q\}$. A test based on NU residuals is made by rejecting when H is small, i.e. when $H < \chi^2_{1-\alpha}(N-1)$; and the p-value for this test is $P\{\chi^2(N-1) < H\}$.

3.3 MONOTONE VARIANCES

Suppose next that we wish to test the homogeneity null hypothesis against an alternative that the variances are either increasing throughout the range of t, or, decreasing throughout the range of t. Then we shall make two-tailed tests for this problem by using the statistics defined above as follows. Reject homogeneity when

$$Q > \chi^2_{\alpha/2}\ (2N),\text{ or } Q < \chi^2_{1-\alpha/2}\ (2N).$$

Put $$p^* = P\{\chi^2\ (2N) > Q\},$$

then the p-value of this test for monotone variances is given by

$$p = 2p^*I_{(0,\frac{1}{2}]}\ (p^*) + 2\ (1 - p^*)I_{(\frac{1}{2},\ 1]}\ (p^*). \qquad (3.2)$$

A test based on NU residuals is made by rejecting when

$$H > \chi^2_{\alpha/2}\ (N-1),\text{ or } H < \chi^2_{1-\alpha/2}\ (N-1).$$

For this test put $p^* = P\{\chi^2(N-1) > H\}$. The p-value of this test is again given by formula (3.2) above.

4. SIMULATION RESULTS

In this section we report the results of a simulation study designed to determine to what degree the data patterns anticipated from the heuristic argument in Section 3 actually do occur under heteroscedastic alternative models, and to compare the tests based on the Q and H statistics with four comparable tests from the literature. We consider tests to be comparable with Q and H if they are based on essentially the same model assumptions; and if they have either exact distribution theory under the null hypothesis, or approximate distribution theory that is sufficiently accurate to allow precise determinations of significance points for small samples under the null hypothesis of homoscedasticity. It is important that the nominal test size be accurately attained in comparative power studies, since even small inaccuracies in test sizes are often greatly magnified in power results, which vitiates the study. Also, tests have been chosen for comparison that make the same assumptions under both the null and alternative hypotheses.

The four tests from the literature which we have studied in power comparisons include those of Goldfeld and Quandt (1965), Theil (1971, with further references), Harvey and Phillips (1974), and Harrison and McCabe (1979). These are relatively well-known tests and provide a good base for comparison for these new tests.

Since each of these four tests can be modified to yield better power in many cases, we have considered two versions of each test, which we shall call the "unmodified" and the "modified" versions. Consider first the tests of Theil, Goldfeld and Quandt, and Harvey and Phillips. Each of these tests can be modified by omitting some of the central observations from a particular part of the analysis. Following the recommendation of Harvey and Phillips, we omit enough central observations for each test so that the numerator and denominator degrees of freedom for the resulting F-statistic are both approximately n/3. The resulting tests are referred to here as the "modified" versions of these tests. The "unmodified" versions of the tests are the versions in which no central observations are omitted. In both versions of these tests, we make the numerator and denominator degrees of freedom as nearly equal as possible.

Similarly, if we write Harrison and McCabe's test statistic as

$$\sum_{j=1}^{m} \hat{e}_j^2 \Big/ \sum_{j=1}^{n} \hat{e}_j^2 ,$$

where $m = [\gamma n + .5]$, $[x]$ = the integer part of x, and $\hat{e}_j$ = the j-th ordinary least squares (o.l.s.) predicted residual, then the labels "unmodified" and "modified", used in reference to Harrison and McCabe's test, are meant here to refer to the versions of the test with γ-values of .5 and .4, respectively. Both values of γ produced good power results in the studies conducted by

Harrison and McCabe.

4.1 ALTERNATIVE HYPOTHESIS BEHAVIOR OF THE UNIFORM RESIDUALS

4.1.1 U-SHAPED HISTOGRAMS

The tests given above in Section 3 are based on the assumption that the uniform residuals will have heavy-tailed, symmetric, union-shaped distributions on the unit interval in the presence of heteroscedasticity. To investigate the validity of this assumption for a particular pattern of increasing variance, histograms have been compiled, one for each of the n-p-1 uniform residuals, using repeated samples of generated regression data. The samples were generated for a particular variance pattern and for a fixed n x p regression matrix X of rank p, by suitably modifying the normal residuals obtained from the subroutine GGNML of The International Mathematical and Statistical Libraries, Inc. (IMSL). For each of the n-p-1 histograms, the unit interval was divided into 20 bins of equal length for recording relative frequencies for the generated samples.

The histograms for the variance pattern, $\sigma_j^2 = j$, are graphed together in Figure 3. They are based on 2500 generated samples of size 20 with p = 2. For this variance pattern, the sample index j is the variance <u>and</u> the index for the 17 histograms. The index j runs from 1 to 20, but the first histogram is for j = 4. The symmetry and U-shaped pattern evident in these histograms were observed for the other forms of increasing variance investigated as well. The degree of U-shapedness for the distributions of individual uniform residuals did vary, though, with the

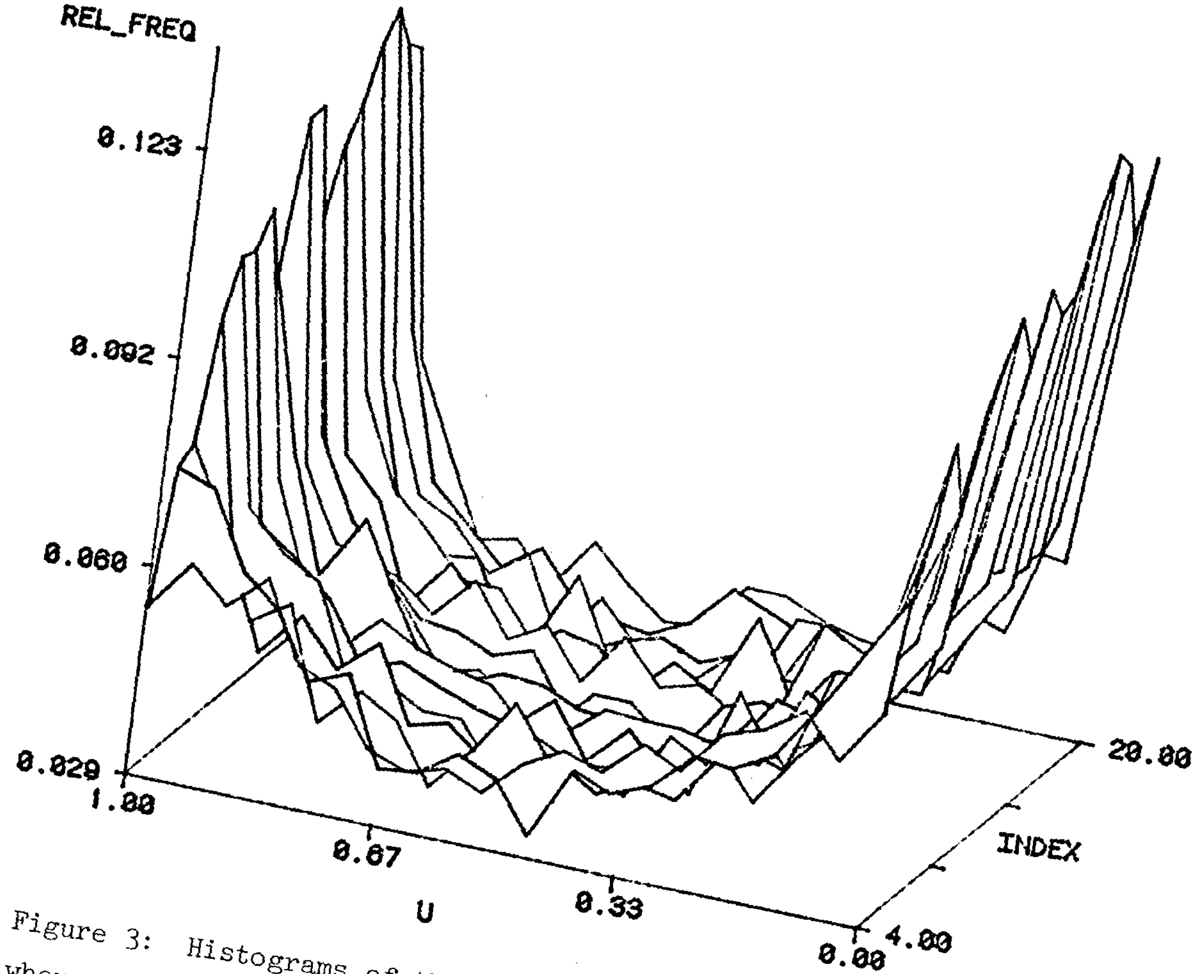
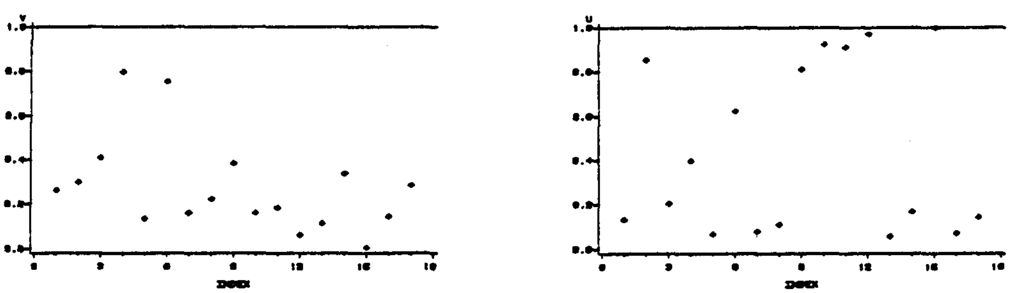


Figure 3: Histograms of the Uniform Residuals for the Case, $\sigma_j^2 = j$, where the Index j Runs from 1 to 20.

particular patterns of increasing variance considered. For example, the histograms for the variance pattern, $\sigma_j = .25$ for $j = 1$ to 5 and $\sigma_j = 1.0$ for $j = 6$ to 20 (corresponding to function f_{22} of Figure 6 below, with $n = 20$): are approximately uniform for $j = 4$ and 5, are extremely heavy tailed for $j = 6$ and 7, and are U-shaped but gradually less heavy-tailed as j goes from 8 to 20. Tables of relative frequencies for these and other forms of increasing variance are given in Hester (1984).

4.1.2 SPLATTER PATTERNS FOR A SINGLE GENERATED SAMPLE

In addition to the histograms in Figure 3, we have also graphed the uniform residuals and the v_i's, given in Section 3, for a single sample with the same case of increasing variance. These are plotted against the sample index $i = j-p-1$ in Figure 4. For this sample, the value of the NU residuals test statistic H is 31.26, and the



(a) (b)

Figure 4: Plots of the v_i's, (a), and u_i's, (b), for a Single Generated Sample with $\sigma_j^2 = j$, $j = 1$ to 20, $i = 1$ to 17. The p-value for the NU Residuals Test H is .013.

p-value is $P(\chi^2_{16} > 31.26) = .013$. Most of the uniform residuals for this single sample are near the endpoints zero and one, and they occur in those two areas in about equal proportions. The corresponding v_i's are mostly in the lower tail near zero.

These plots illustrate the usefulness of ordered plots in checking for the two main features of the uniform residuals in the presence of ordered heteroscedasticity: symmetry about the midpoint ½, and the clustering of values near the endpoints zero and one. Also, by plotting against the ordering index, changes in the behavior of the uniform residuals, occurring as the index changes, can be noted and investigated. In general, plots of the uniform residuals are useful in identifying various potential departures from the uniformity occurring under the null hypothesis, as suggested in Q83. And, in the case of potential heteroscedasticity, they are useful when the null hypothesis is rejected; to check that the residual plot conforms with the hypothesis of heteroscedasticity, or, at least, that rejection of the null hypothesis is not due to some other identifiable inconsistency with the model assumptions.

4.2 PATTERNS OF INCREASING VARIANCE

The patterns of increasing variance used in the power comparisons below are patterns over the unit square, in the positive quadrant of the plane defined by the jth standard deviation σ_j, and by the sample index variable divided by the sample size, j/n. This choice of patterns is based on the invariance of each of the test statistics considered here under transformations of the form $\underset{\sim}{y}^* = a\,\underset{\sim}{y} + X\underset{\sim}{c}$,

($a > 0$, $\underset{\sim}{c}$ a vector of constants), which is listed as property (3) of uniform residuals in Section 2 above. The use of generated variables of the form, $y_j = \underset{\sim}{x}_j'\underset{\sim}{\beta} + \sigma_j\xi_j$, is thus equivalent to using variables of the form, $y_j^+ = \underset{\sim}{x}_j'\underset{\sim}{\beta} + a\sigma_j\xi_j$, (with $a > 0$, $\sigma_j > 0$, $\xi_j \overset{ind}{\sim} N(0,1)$; and $j = 1, \ldots, n$); since y_j^+ can be written as $y_j^+ = ay_j + \underset{\sim}{x}_j'\underset{\sim}{\beta}(1-a)$. And thus the power comparisons under consideration may be conducted more efficiently by considering only the set of y_j^+'s with error terms $a\sigma_j\xi_j$, instead of the many corresponding sets of y_j's with error terms $\sigma_j\xi_j$, through an appropriate choice of the coefficient "a".

4.2.1 TRANSFORMING VARIANCE PATTERNS TO THE UNIT SQUARE

Suppose patterns of increasing variance are visualized as plots of the jth standard deviation σ_j on the vertical axis, against the sample index variable j, along the horizontal axis. The loci in the resulting plane of all possible patterns of increasing variance are then given by the n vertical lines in the positive quadrant, occurring at the n values of the sample index j, as in Figure 5a. Now, because of the property of invariance above, the variance patterns corresponding to $\{j,\sigma_{k,j}\}$ and to $\{j,a_{k,n}\sigma_{k,j}\}$ are equivalent for any finite positive constant $a_{k,n}$ (where k denotes the k-th pattern of increasing variance over the n values of j). By choosing $a_{k,n} = [\max_{1\leq j\leq n} \{\sigma_{k,j}\}]^{-1} = \sigma_{k,n}^{-1}$, the problem of looking at all possible patterns of variance on an unbounded rectangle over the positive quadrant of the plane, defined by the

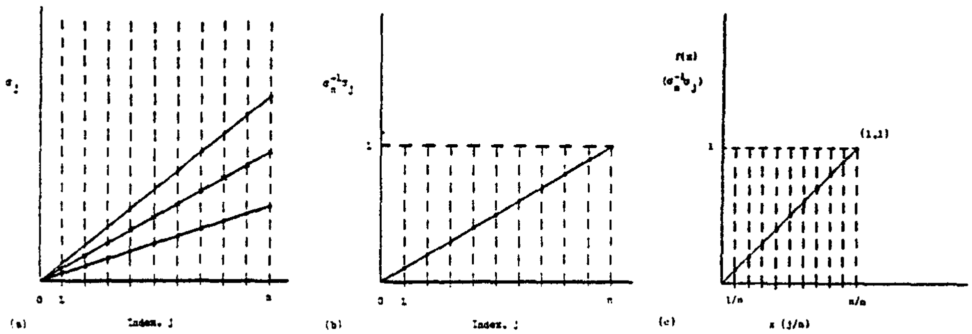


Figure 5: Representations Corresponding to the Increasing Variance Pattern, $\sigma_j = \alpha j$ ($\alpha > 0$; $j = 1, \ldots, n$).

variables $\sigma_{k,j}$ and j, reduces to looking at all possible patterns over a bounded rectangle in the positive quadrant, defined by the variables $\sigma_{k,n}^{-1}\sigma_{k,j}$ and j. Here k is used as an index of patterns over the bounded rectangle rather than of patterns over the unbounded rectangle, since many patterns in the unbounded rectangle correspond to a single pattern in the bounded rectangle. For example, in Figure 5 a few of the many patterns corresponding to $\sigma_j = \alpha j$, $\alpha > 0$, are presented in Figure 5a, along with the corresponding single representations in Figures 5b and 5c.

To allow for different sample sizes, we use the variable j/n, shown plotted in Figure 5c on the horizontal axis, instead of the variable j shown in Figure 5b. This gives every possible pattern of variance for an arbitrary number of n data points an equivalent representation of n points on the unit square in the positive quadrant, defined by the variable pair $(j/n, \sigma_{k,n}^{-1}\sigma_{k,n})$. Note, finally, that looking at patterns in this coordinate system is equivalent, by an argument similar to the one above, to looking at patterns on the unit square (or other bounded rectangle) in the coordinate system defined by the variable pair $(j/n, \sigma_j)$, as claimed above.

4.2.2 FUNCTIONAL REPRESENTATION OF VARIANCE PATTERNS OVER THE UNIT SQUARE

For the representations of variance patterns considered here, we use piecewise continuous functions over the unit square. For convenience, these functions are denoted by $f(x)$, f (or f_k when more than one function (pattern) is considered); and we denote the abscissa by x. The actual values of the standard deviations used correspond to the values of the piecewise continuous functions $f_k(x)$ evaluated at the points $x = j/n$, $j = 1, \ldots, n$, as shown in Figure 5c.

The resulting functional representations of different variance patterns have the following important features in common: they are nondecreasing functions with the common endpoint (1,1); they are nonnegative functions constructed so that $f(x) > 0$ for $0 < x \leq 1$; and when $f(0)$ is defined, it is allowed to take any value in the interval $[0,1]$. Where $f(0)$ is not explicitly defined, its implied value is $\lim f(x)$ as $x \longrightarrow 0^+$. Note that $f_o(x) \equiv 1$, $0 < x \leq 1$, may be considered the unique representation of the null hypothesis of constant variance for all possible sample sizes.

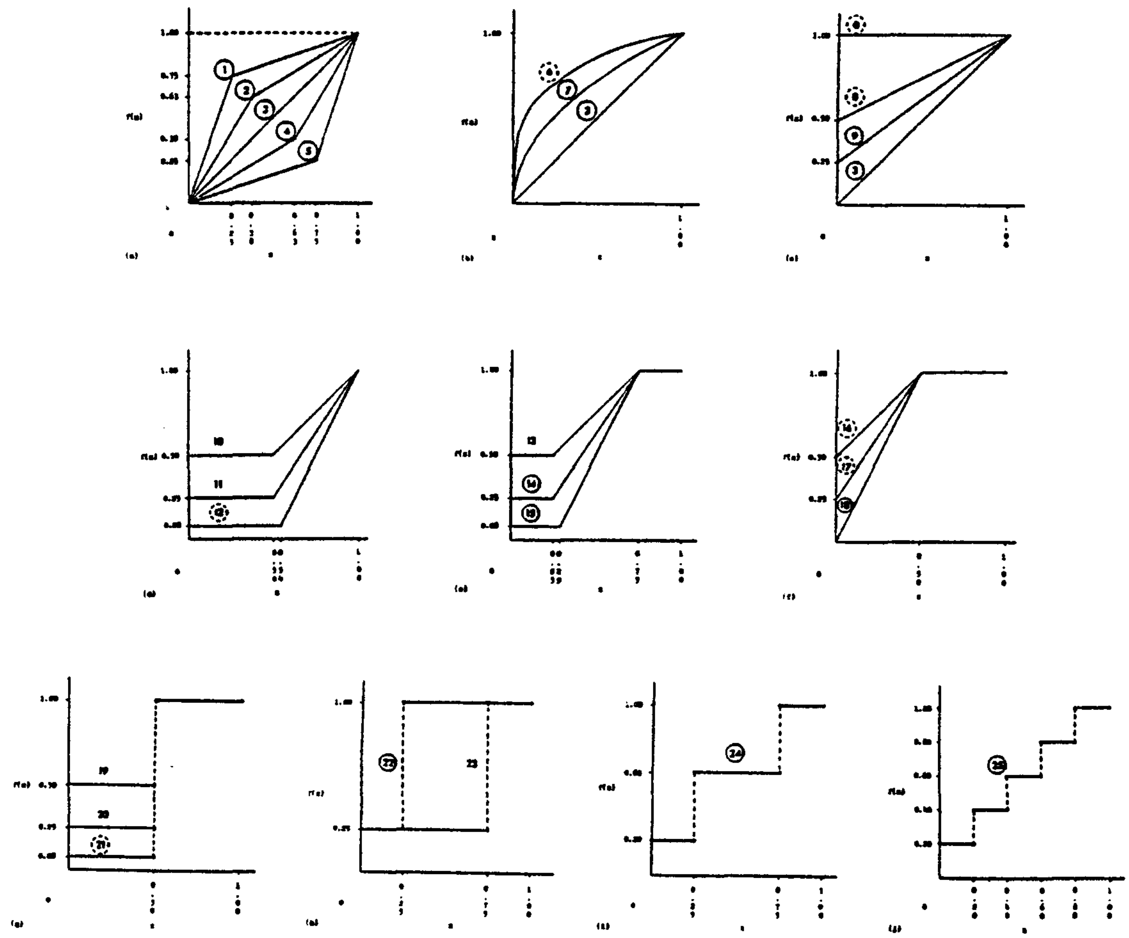


Figure 6: Patterns of Increasing Variance Used in Power Comparisons. The j-th Standard Deviation, for Pattern k and Sample Size n, is Computed from $\sigma_{k,j} = f_k(j/n)$.

Code

Complete Circle: For pattern k, one of the NU residuals tests had the largest estimated power $\hat{P}_{M,k}$, say and none of the other tests' powers (excluding the uniform residuals test) were within two standard deviations of $\hat{P}_{M,k}$.

Broken Circle: For pattern k, the estimate of the power of the NU residuals test H was within $2s_{DM}$ of the largest observed power estimate.

No Circle: For pattern k, the power estimates for the NU residuals tests were not within $2s_{DM}$ of the largest observed power estimate.

Table 1. Formulas for the Patterns of Increasing Variance Considered. Standard Deviations are Obtained from $\sigma_j = f(j/n)$.

k	$f_k(x)$	
0	$I_{(0,1]}(x)$	
1,2,3,4,5	$\frac{(1-a_k)}{a_k} x I_{(0,a_k]}(x) + \{(1-a_k) + \frac{a_k}{(1-a_k)}(x-a_k)\} I_{(a_k,1]}(x)$	$a_k = (k+1)/8,\ k=1,\ldots,5$
6,7	$x^{a_k} I_{(0,1]}(x)$	$a_6 = .25,\ a_7 = .5$
8,9	$(a_k+b_k x) I_{(0,1]}(x)$	$a_8 = b_8 = .5$ $a_9 = .25,\ b_9 = .75$
10,11,12	$\max(c_k, a_k+b_k x) I_{(0,1]}(x)$	$c_{10} = .50,\ a_{10} = 0,\ b_{10} = 1.0$ $c_{11} = .25,\ a_{11} = -.5,\ b_{11} = 1.5$ $c_{12} = .08,\ a_{12} = -1,\ b_{12} = 2.0$
13,14,15	$\min\{1, \max(c_k, a_k+b_k x)\} I_{(0,1]}(x)$	$c_{13} = .50,\ a_{13} = .25,\ b_{13} = 1.0$ $c_{14} = .25,\ a_{14} = -.125,\ b_{14} = 1.5$ $c_{15} = .08,\ a_{15} = -.5,\ b_{15} = 2.0$
16,17,18	$\min(1, a_k+b_k x) I_{(0,1]}(x)$	$a_{16} = .50,\ b_{16} = 1.0$ $a_{17} = .25,\ b_{17} = 1.5$ $a_{18} = 0,\ b_{18} = 2.0$
19,20,21	$a_k I_{(0,.5]}(x) + I_{(.5,1]}(x)$	$a_{19} = .50$ $a_{20} = .25$ $a_{21} = .08$
22,23	$.25 I_{(0,a_k]}(x) + I_{(a_k,1]}(x)$	$a_{22} = .25,\ a_{23} = .75$
24	$.2 I_{(0,.25]}(x) + .6 I_{(.25,.75]}(x) + I_{(.75,1]}(x)$	
25	$\sum_{m=1}^{5} .2m I_{(.2(m-1),.2m]}(x)$	

Note: $I_A(x)$ is 1 if $x \in A$, and is 0 if $x \notin A$.

The graphs of the functions used are given in Figure 6. Their corresponding algebraic expressions are given in Table 1. Note that the step functions used are taken to be continuous from the left at all points in the interval (0,1].

4.3 POWER COMPARISON CONSIDERATIONS

4.3.1 EXTENSION OF THE POWER RESULTS TO THE COLUMN SPACE OF X

Results of the power comparisons for the tests above are given in Table 2. These simulation results were obtained using an X matrix of 20 rows and 2 columns, say X_o, which remained fixed over all samples and variance patterns.

These power results are correct for any other X matrix with the same column space. This follows from the invariance property noted above in Section 4.2, and the fact that the vector $\underset{\sim}{y}^* = (X_o M)\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$, for a p x p nonsingular matrix M, can be rewritten $\underset{\sim}{y}^* = \underset{\sim}{y} + X_o[(M-I)\underset{\sim}{\beta}]$. Note that the invariance of the test statistics would not change for singular M, but then the dimensionality of the model itself would change. The matrix X_o used here consisted of the unit vector, $\underset{\sim}{1} = (1,1, \ldots, 1)'$, and a column of U(0,20) pseudo-random variables, obtained by using the IMSL subroutine GGUBS with seed 647827.DØ.

Results similar to those in Table 2 were found for other column spaces as well. One hundred samples for two other column spaces with n = 20 and p = 2, and 2500 samples for a column space with n = 40 and p = 4 were generated for this parallel study. The latter column space contained the unit vector and three vectors of U(0,20) r.v.'s generated from the same seed used above. The two other column spaces each contained the unit vector as their first generating column and second columns of, respectively, the vector of index numbers, $\underset{\sim}{j} = (1,2,3, \ldots, 20)'$, and a vector of lognormal pseudo-random variables. The variance patterns used in simulations for these last two column spaces were those given by f_3, f_7, f_9, and f_{10}. Twenty of the twenty-five patterns were used for the space with $n = 40$ and $p = 4$.

4.3.2 MEASURES OF RELATIVE EFFICIENCY

Along with the power estimates for each variance pattern, we have included two quantities which measure the relative efficiency of using ordinary least squares (o.l.s.) estimators for the regression parameters, compared to the use of generalized least squares (g.l.s.) estimators, when the regression errors are in fact heteroscedastic. The two measures of relative efficiency used here, τ_1 and τ_2, correspond to computation of the o.l.s. estimators under two different assumptions: τ_1 is computed assuming that the errors are homoscedastic, the usual assumption when o.l.s. estimators are used; and τ_2 is computed assuming that the errors are heteroscedastic, the correct assumption here.

Instead of using the ratio of generalized variances, we have taken a ratio of "generalized standard deviations", or, essentially, the square root of the ratio of generalized variances, as a measure of relative efficiency. The term "generalized variance" here refers to the quantity $|\Sigma|^{1/p}$ for a general p x p covariance matrix Σ. The measures of relative efficiency used are thus ratios of quantities with the same units of measure as the regression parameter estimators, instead of ratios of quantities with the squares of those units.

The two measures of relative efficiency are

$$\tau_i = \{|\Sigma_{22}|/|\Sigma_{21}^{(i)}|\}^{1/(2p)},\ i = 1,2, \text{ with}$$

$$\Sigma_{22} = \tilde{\sigma}_2^2(X'V^{-1}X)^{-1},\ \Sigma_{21}^{(2)} = \tilde{\sigma}_2^2[(X'X)^{-1}X'VX(X'X)^{-1}],$$

$$\Sigma_{21}^{(1)} = \tilde{\sigma}_1^2(X'X)^{-1}, \text{ and with } \operatorname{Var}(\underset{\sim}{y}) = \sigma^2 V,$$

the matrix V changing with the different patterns of variance considered. The variance matrices Σ_{22} and $\Sigma_{21}^{(2)}$ of the g.l.s. and o.l.s. estimators, respectively, are computed based on the assumption that Var $(\underset{\sim}{y}) = \sigma^2 V$, with $V = \text{diag}\ (v_1, \ldots, v_n)$, known. The "estimator" of σ^2 used to correspond to this assumption is

$$\tilde{\sigma}_2^2 = n^{-1}\sum_{i=1}^{n} \{\operatorname{Var}(y_i)/v_i\} = \sigma^2.$$

In contrast, the covariance matrix $\Sigma_{21}^{(1)}$ of the o.l.s. estimators is based on the assumption that $\operatorname{Var}(\underset{\sim}{y}) = \sigma^2 I$, when, in fact, $\operatorname{Var}(\underset{\sim}{y}) = \sigma^2 V$. The "estimator" of σ^2 used for this incorrect assumption is

$$\tilde{\sigma}_1^2 = n^{-1}\sum_{i=1}^{n} \operatorname{Var}(y_i) = n^{-1}\sigma^2\sum_{i=1}^{n} v_i,$$

which, in general, is not equal to σ^2, but would equal σ^2 if the

Var($\underset{\sim}{y}$) were $\sigma^2 I$.

4.3.3 GENERAL TRENDS IN POWER AND RELATIVE EFFICIENCY

Some interesting, general trends in power and relative efficiency, over the range of variance forms used here, may be observed by examining Table 2 in conjunction with Figure 6. For example, the power estimates increase and relative efficiency decreases over the sets of variance patterns $\{f_1,f_2,f_3,f_4,f_5\}$, $\{f_6,f_7,f_3,\}$, $\{f_0,f_8,f_9,f_3\}$, $\{f_{10},f_{11},f_{12}\}$, $\{f_{13},f_{14},f_{15}\}$, $\{f_{16},f_{17},f_{18}\}$, and $\{f_{19},f_{20},f_{21}\}$, as the functions f_k move vertically downward in various fashions away from f_0, the functional representation of constant variance. In constrast, the power estimates (relative efficiency), although not as clearly as before, appear to rise (fall) and then fall (rise) for most of the tests over the sets $\{f_{10},f_{13},f_{16}\}$ and $\{f_{22},f_{20},f_{23}\}$, as the functions f_k shift horizontally to the right. Again, similar results were found for the simulation of 2500 samples with $n = 40$ and $p = 4$.

4.4 POWER COMPARISON RESULTS

4.4.1 CRITERION AND RESULTS

Table 2 contains the power estimates for the two NU residuals tests, the uniform residuals test, and the modified and unmodified versions of the tests of Harrison and McCabe, Theil, Harvey and Phillips, and Goldfeld and Quandt. These results are for 2500 samples from a normal simple linear regression model with heter-

Table 2. Estimated Power of the Tests for Heteroscedasticity for the 25 Variance Patterns of Figure 6. Based on 2500 Generated Samples of Size n=20, with p=2. Test Size α=.05.

Pattern (k)	Power Range	NU Residuals Test	Uniform Residuals Test	Harrison McCabe Test	Theil's Test	Harvey Phillips Test	Goldfeld Quandt Test	$2S_{DM}$	Relative Efficiency τ_1	Relative Efficiency τ_2
0	.0431 - .0511	(.0503) .0487	.0483	(.0439) .0459	(.0455) .0511	(.0431) .0483	(.0435) .0431	-	1.000	1.000
1	.225 - .324	(.303) .324**	.304	(.238) .266	(.242) .273	(.225) .242	(.235) .251	.0195	.619	.615
2	.456 - .700	(.676) .700**	.666	(.484) .533	(.487) .526	(.456) .490	(.485) .526	.0230	.432	.429
3	.748 - .921	(.918)* .921**	.916*	(.751) .787	(.759) .789	(.748) .747	(.778) .812	.0196	.311	.310
4	.898 - .986	(.985)* .986**	.986*	(.902) .898	(.913) .930	(.930) .936	(.936) .939	.0132	.227	.234
5	.937 - .996	(.996)** .994*	.996**	(.951) .937	(.967) .973	(.976) .981	(.976) .977	.0102	.160	.184
6	.182 - .202	(.186)* .198*	.188*	(.190)* .193*	(.186)* .202**	(.182) .182	(.186)* .192*	.0179	.926	.921
7	.386 - .479	(.461)* .479**	.461*	(.411) .422	(.411) .418	(.386) .391	(.409) .435	.0216	.728	.723
8	.216 - .234	(.217)* .230*	.216*	(.227)* .217*	(.225)* .234**	(.222)* .220*	(.224)* .232*	.0182	.935	.929
9	.448 - .517	(.503)* .517**	.494	(.466) .454	(.467) .468	(.448) .448	(.474) .493	.0218	.777	.772
10	.262 - .339	(.301) .300	.290	(.346)* .262	(.331) .318	(.359)** .336	(.347)* .336	.0186	.897	.894
11	.700 - .856	(.774) .771	.750	(.808) .700	(.818) .784	(.856)** .829	(.844)* .820	.0174	.631	.640
12	.965 - 1.000	(.997)* .995*	.996*	(.986) .965	(.997)* .994*	(.9996)** .9996**	(.999)* .998*	.0073	.251	.268
13	.345 - .398	(.355) .358	.345	(.380)* .358	(.372) .368	(.360) .345	(.381)* .398**	.0208	.864	.857
14	.676 - .826	(.826)** .823*	.812*	(.716) .764	(.718) .735	(.700) .676	(.736) .756	.0211	.577	.571
15	.893 - .998	(.998)** .936*	.998**	(.909) .939	(.909) .912	(.916) .893	(.944) .967	.0148	.218	.216
16	.144 - .195	(.168) .178*	.166	(.162) .195**	(.163) .180*	(.144) .149	(.157) .168	.0175	.932	.929
17	.236 - .338	(.334)* .330*	.304	(.264) .338**	(.271) .304	(.236) .240	(.262) .286	.0210	.791	.787
18	.330 - .640	(.598) .640**	.595	(.379) .520	(.387) .438	(.330) .353	(.384) .446	.0237	.439	.436
19	.401 - .582	(.417) .422	.401	(.572)* .432	(.502) .403	(.488) .385	(.582)** .540	.0215	.826	.805
20	.737 - .974	(.915) .903	.889	(.934) .860	(.865) .780	(.843) .737	(.974)** .956*	.0196	.495	.478
21	.827 - 1.000	(.9996)* .9996*	.9996*	(.984) .956	(.934) .887	(.902) .827	1.000** 1.000**	.0166	.168	.162
22	.214 - .572	(.542) .572**	.522	(.247) .347	(.253) .327	(.214) .259	(.235) .268	.0212	.514	.519
23	.719 - .865	(.829) .824	.786	(.805) .719	(.830) .828	(.865)** .864*	(.840) .819	.0161	.538	.587
24	.492 - .746	(.729)* .746**	.711	(.492) .532	(.513) .384	(.495) .543	(.502) .525	.0215	.527	.533
25	.634 - .816	(.807)* .816**	.806*	(.650) .719	(.664) .678	(.634) .643	(.664) .709	.0212	.571	.573

Note: The estimated power for the unmodified versions of the tests of Harrison and McCabe, Theil, Harvey and Phillips, and Goldfeld and Quandt are given in parentheses over the corresponding values for the modified test versions. The estimated power for the NU Residuals test statistic H* is given in parentheses over the estimated power for the test statistic H.

oscedastic errors, using the 20 x 2 matrix X_o, described above in Section 4.3.1, and the patterns of increasing variance found in Figure 6. The relative efficiency measures for X_o and the variance patterns of Figure 6 are also given.

For the row associated with the k-th variance pattern, the largest power estimate $\hat{P}_{M,k}$, say, is denoted in the table by a double asterisk, or double star (**). Power estimates in the same row within a distance $2s_{DM,k}$ of $\hat{P}_{M,k}$, are also marked, using a single asterisk (*). The statistic $s_{DM,k}$, for the k-th variance pattern, is the largest of $\binom{11}{2} = 55$ estimated standard deviations of paired differences between power estimates, for the eleven tests.

The absence of any asterisks beside a power estimate implies a statistically significant difference between the power for that test and the power for the test associated with $\hat{P}_{M,k}$. On the other hand, the presence of a single asterisk next to a power estimate does not necessarily guarantee that statistically significant differences do not exist, since standard deviations for some pairs of tests are substantially smaller than $s_{DM,k}$. The single asterisk, however, does indicate high probability of a difference less than $2s_{DM,k}$($\leq$ 0.0237 here) between the power of the starred test and the test associated with $\hat{P}_{M,k}$. Complete variance tables for paired differences of power estimates and observed correlation matrices of power estimates for the tests considered here may be found in Hester (1984).

As a criterion for comparing the power of these eleven tests, we use the number of variance patterns with at least one asterisk for a given test. This number includes those cases, denoted by two stars (**), for which a test has the largest estimated power, and the cases with one star (*), where, for pratical purposes the test can be considered to be no worse than the test with the largest estimated power. A test's having at least one star can thus be equated, for practical purposes, with its being no worse than or, equivantently, being just as good as the test with the largest estimated power.

(Aside: The criterion used here is a function of the twenty-five variance patterns examined, and may be written in the form

$$\sum_{k=1}^{25} w_k I_{[*\text{ or }**]}(\text{power}_{i,k}),$$

with $w_k \equiv 1$, where the indicator function is 1 or 0 depending on whether or not the power for the i-th test and k-th variance pattern has at least one star. Other criteria may be devised by using different values of w_k, depending on the prior expectations of the scaled pattern of variance.)

The best performing test using the criterion above is the NU residuals test H, which has at least one asterisk for 19 of the 25 variance patterns. Following this test, in order, are: the other NU residuals test H^*, the uniform residuals test Q, Goldfeld and Quandt's unmodified and modified tests, and Harvey and Phillips' and Harrison and McCabe's unmodified tests, with tallies of 14,11,9,6, 5 and 5, respectively. The 19 cases where H was "no worse" than the test with largest estimated power include 13 cases where H, H*, or Q had statistically significantly larger power than all the remaining tests. In all of these 13 cases either H or H* had the largest power estimates, and in the three cases where the estimate for H* was larger, the power estimates for H and H* were within 0.003 of each other.

4.4.2 FURTHER OBSERVATIONS

One interesting observation from the power results is that for the majority of variance patterns where H performed well, the left endpoint f(0) of the corresponding functional representation took relatively small values. Small values for f(0) generally correspond to relatively wide ranges in the values f(1/n), f(2/n), ..., f(n/n), which represent the scaled standard deviations for a sample of size n.

We now focus on the six cases in which the power estimates for H were statistically significantly less than the largest power estimates. In three of these cases the power of H was still relatively high, with values of 0.901, 0.927, and 0.953 for the ratios of the two power estimates. These values correspond to the variance patterns f_{11}, f_{20}, and f_{23}, respectively. The ratios of the power estimates in the remaining cases were 0.836, 0.899, and 0.725, which correspond, respectively, to the variance patterns f_{10}, f_{13}, and f_{19}.

For this latter group of variance patterns the respective values of the relative efficiency measure τ_1 were 0.897, 0.864, and 0.826. Note that both τ_1 and τ_2 take the value one in the case of constant variance. High relative efficiency thus implies little difference in the use of o.l.s. estimators and g.l.s. estimators for the regression parameters. Thus, with high relative efficiency, little is lost by accepting the null hypothesis of constant variance when it is false, and it makes little difference which test is used.

Two of the six variance functions mentioned above, f_{19} and f_{20}, together with f_{21}, correspond to cases in which the sample is divided into two groups of equal size with constant, but unequal, variance within each group. These equally divided "two-sample" cases are the ideal cases for Goldfeld and Quandt's unmodified test. Note that Goldfeld and Quandt's tests do not fare as well in the two-sample cases with functions f_{22} and f_{23}, although for f_{23}, the unmodified test, at least, does not do poorly. The functions f_{22} and f_{23} correspond to the "two-sample" cases in which the sample is divided into two groups of unequal size.

4.4.3 POWER RESULTS FOR n = 40 and p = 4.

Power estimates were also obtained for 2500 samples for which n was 40 and p was 4. The most notable difference between the two studies was a substantial overall increase in the levels of the power estimates, which occurred with the increase in sample size. Because of this power increase, computations were not made for

f_4, f_5, f_{12}, f_{15}, and f_{21}.

For most variance patterns, the relative performances of the tests did not change much when n and p were increased. The performance of H was noticeably improved, however, for the patterns f_6, f_{16}, and f_{17}, for which H had the largest estimated power. The complete results of this study can be found in Hester (1984).

4.5 DISCUSSION

The results of these power studies indicate that the uniform and NU residuals tests have good power for testing heteroscedasticity when compared to the tests of Goldfeld and Quandt, Theil, Harvey and Phillips, and Harrison and McCabe. We feel that these results indicate that the NU residuals test H is somewhat superior, overall, to these competing tests.

5. AN EXAMPLE

Ryan, Joiner, and Ryan (1976, p.278) present data for 31 black cherry trees grown in Allegheny National Forest, Pennsylvania. The variables given are tree volume y, tree diameter d, and tree height h. This set of data was examined by Atkinson (1982) who suggested that a cube root transformation for the dependent variable y was appropriate, yielding $y^{1/3} = \beta_0 + \beta_1 d + \beta_2 h + \varepsilon$ as a reasonable model.

This model was also used by Cook and Weisberg (1983) who tested the data set for heteroscedasticity. Their analysis was based on the score statistic, which has an approximate χ^2 distribution for large samples. A value of 3.24 was obtained for this statistic for the test of the null hypothesis of constant variance against an alternative that variance was of the form $\sigma^2 e^{\lambda h}$, $(\lambda > 0)$. The approximate p-value for this value of the score statistic, based on the asymptotic $\chi^2(1)$ distribution, is .072. The corresponding exact p-values of the uniform and NU residuals tests Q and H for this model are .097 and .114, respectively.

In addition to the model above with intercept term β_0, we have also examined the data using the model, $y^{1/3} = \beta_1 d + \beta_2 h + \varepsilon$, which has no intercept term. This no-intercept model yields the predicted volume that would be expected for a hypothetical tree with zero height and zero diameter; although a nonzero intercept estimate would not be unexpected for this set of data since the tree heights considered are limited to the range between 63 and 87, with none near zero. The no-intercept model is of more than theoretical interest for this data set, however, since the intercept estimate in the first model is not significantly different from zero. The estimate of the intercept is, in fact, less than half its estimated standard error, leading to the conclusion that the no-intercept model is the more appropriate one. The exact p-values for Q and H for this no-intercept model are .040 and .044, respectively. The approximate p-value of the score statistic for this no-intercept model, assuming that the alternative hypothesis variance has the form $\sigma^2 e^{\lambda h}$, is .088.

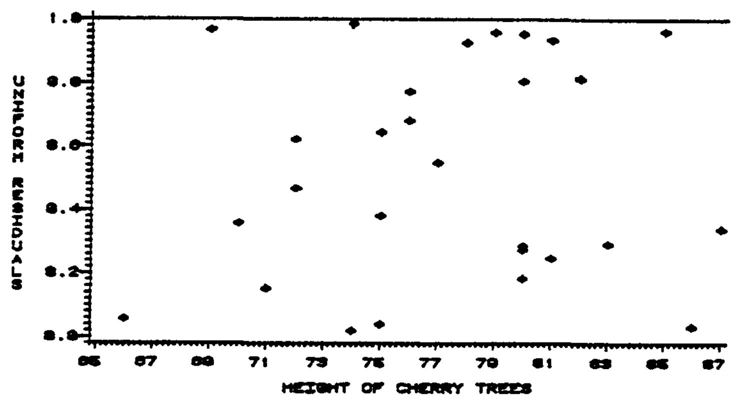


Figure 7: Heights and Corresponding Uniform Residuals of Cherry Trees for the Model, $y^{1/3} = \beta_1 d + \beta_2 h + \varepsilon$.

(Aside: Cook and Weisberg found that for a small simulation study the use of the χ^2 approximation for the score statistic yielded a conservative test. This indicates that the actual p-values for their score statistic are probably less than the values of .072 and .088 given here.)

The graph of the uniform residuals against tree height for this model is given in Figure 7. The residuals are scattered more or less symmetrically about ½ and exhibit some affinity for the endpoints of the unit interval. Thus the graph is consistent with the hypothesis of heteroscedasticity.

The application of the score statistic and the NU and uniform residuals tests to this particular set of data illustrates two points. First, the score test, and many other tests for heteroscedasticity, are based on large sample approximating distributions and thus have only approximate p-values; whereas the NU and uniform residuals test statistics have exact, well-known distributions under the null hypothesis, which yield exact and readily available p-values for all sample sizes. Secondly, the score test is an example of tests for which a functional form of alternative variance must be specified for the test to be carried out. While the NU and uniform residuals tests, as well as the other tests considered above in Sec. 4, do require the dependent variable to be ordered according to increasing variance, they do not require that the functional form of the variance be specified before they can be implemented. Thus, for these tests, questions about the functional form of the variance may be deferred until a decision is made concerning the presence or absence of heteroscedasticity.

For the cherry tree data, a prediction equation for variance was obtained, after concluding that the existence of heteroscedasticity was likely for this data set, by fitting the model, $\ln(\hat{\varepsilon}^2) = \mu + \beta h + \xi$, where $\hat{\varepsilon}$ represents the o.l.s. residual from the model, $y^{1/3} = \beta_1 d + \beta_2 h + \varepsilon$. The prediction equation,

$$\hat{\sigma}^2(h) = \exp\{\hat{\mu} + \hat{\beta} h\} = \exp\{-18.8 + .165h\},$$

has a corresponding predicted "variance" pattern $\hat{f}$, as in Sec. 4.2, whose values for this sample of size 31 range from $\hat{\sigma}(63)/\hat{\sigma}(87) = .138$ for $j/n = 1/31$ to $\hat{\sigma}(87)/\hat{\sigma}(87) = 1$ for $j/n = 31/31 = 1$. The values of

REFERENCES

Atkinson, A.C. (1982), "Regression Diagnostics, Transformations and Constructed Variables," Journal of the Royal Statistical Society, Ser. B, 44, 1-36.

Cook, R. D., and Weisberg, S. (1983), "Diagnostics for Heteroscedasticity in Regression," Biometrika, 70, 1-10.

Fisher, R. A. (1932), Statistical Methods for Research Workers, 4th Ed., Edinburgh: Oliver and Boyd.

Goldfeld, S. M., and Quandt, R. E. (1965), "Some Tests for Homoscedasticity," Journal of the American Statistical Association, 60, 539-547.

Harrison, M. J., and McCabe, B.P.M. (1979), "A Test for Heteroscedasticity Based on Ordinary Least Squares Residuals," Journal of the American Statistical Association, 74, 494-499.

Harvey, A. C., and Phillips, G. D. A. (1974), "A Comparison of the Power of Some Tests for Heteroskedasticity in the General Linear Model," Journal of Econometrics, 2, 307-316.

Hester, R. A., Jr. (1984), "Uniform Residuals and NU Residuals Tests for Heteroscedasticity," unpublished Ph.D. thesis, North Carolina State University, Department of Statistics.

Miller, F. L., Jr. and Quesenberry, C. P. (1979), "Power Studies of Tests for Uniformity," Communications in Statistics-Simulation and Computations, B8(3), 271-290.

O'Reilly, F. J., and Quesenberry, C. P. (1973), "The Conditional

Probability Integral Transformation and Applications to Obtain Composite Chi-Square Goodness-of-Fit Tests," Annals of Statistics, 1, 74-83.

Quesenberry, C. P. (1983), "Analyzing Uniform Residuals for Time Changes in Regression Models," unpublished manuscript.

Quesenberry, C. P., and Starbuck, R. R. (1976), "On Optimal Tests for Separate Hypotheses and Conditional Probability Integral Transformations," Communications in Statistics, A5(6), 507-524..

Ryan, T. A., Joiner, B.L., and Ryan, B. F. (1976), MINITAB Student Handbook, North Scituate, Massachusetts: Duxbury Press.

Stephens, M. A. (1974), "EDF Statistics for Goodness of Fit and Some Comparisons," Journal of the American Statistical Association, 69, 730-737.

Theil, H. (1971), Principles of Econometrics, New York: John Wiley and Sons.