BU-135

INDEPENDENT STEPWISE RESIDUALS FOR TESTING HOMOSCEDASTICITY[1]

A. Hedayat and D. S. Robson[2]

Cornell University

ABSTRACT

Regression models which specify independent, homoscedastic and normally distributed errors may be analyzed in a stepwise manner to produce calculated residuals having this same property. If the n'th residual is calculated as the deviation of the n'th observation from its predicted value based on a least squares fit to only the first n observations then the resulting sequence of residuals, appropriately normalized, are not only mutually independent and homoscedastic but also are independent of all of the calculated regression functions. If error variance is a monotonic function of the mean then, under certain regularity conditions, the calculated stepwise residuals are likewise monotonically heteroscedastic. Simple linear regression with equally spaced values of the independent variable constitutes one such regular case, and a Monte Carlo study of the "peak-test" of homoscedasticity in this instance shows that for small samples the stepwise residuals are substantially more sensitive to monotonic heteroscedasticity than conventional, untransformed residuals.

[1] This paper was originally presented under the title "Homoscedasticity in Linear Regression Analysis with Equally Spaced X's", on April 4, 1967 at the one hundred fourteenth meeting of the Institute of Mathematical Statistics, at the Georgia Institute of Technology, Atlanta, Georgia.

[2] Research supported in part by Grant Number GB-4502 from the National Science Foundation.

J. Am. Stat Assoc. 65: 1970 (~~Sept~~)

BU-341-M

INDEPENDENT STEPWISE RESIDUALS FOR TESTING HOMOSCEDASTICITY[1]

A. Hedayat and D. S. Robson[2]

Cornell University

1. INTRODUCTION AND SUMMARY

Consider the fixed effects general linear model

$$Y = X\beta + \epsilon \qquad (1)$$

where Y is an N-vector of responses, X is an NXP matrix with rank $r \leq p$ having either fixed known coefficients or coefficients that are stochastically independent of the error term, β is a p-vector of unknown parameters, ϵ is an N-vector of unknown stochastic components with mean zero and is usually called the error (residual or disturbance) vector.

Linear models dealt with in practice usually include in their basic structure the assumption that the covariance matrix of ϵ is $\sigma^2 I_N$ where σ^2 is a scalar and I_N denotes the identity matrix of order N. Specifically it is often assumed that $\epsilon \sim N(0, \sigma^2 I_N)$.

A diagnosis of the validity of these conditions imposed on the linear model residuals is impeded by the fact that under the usual hypothesis of independent and identically distributed errors, deviations from the least squares fit are neither independent nor, in general, identically distributed. Calculated residuals $e = Y - X\hat{\beta}$ are linear functions of the true errors $\epsilon = Y - Y\beta$ at the $\underline{N}$ points of the experimental design and are subject to linear constraints equal

[1] This paper was originally presented under the title "Homoscedasticity in Linear Regression Analysis with Equally Spaced X's", on April 4, 1967 at The one hundred fourteenth meeting of the Institute of Mathematical Statistics, At the Georgia Institute of Technology, Atlanta, Georgia.

[2] Research supported in part by Grant Number GB-4502 from the National Science Foundation.

Paper No. BU-135 in the Biometrics Unit and No. 524 in the Department of Plant Breeding and Biometry, Cornell University, Ithaca, N.Y.

in number to the rank $\underline{r}$ of the design matrix $\underline{X}$. If the design is balanced then residuals are marginally identically distributed, but their joint distribution is singular of rank $\underline{N} - \underline{r}$.

A transformation of these $\underline{N}$ estimated residuals into $\underline{N} - \underline{r}$ orthogonal linear functions of the true errors eliminates this complication in the normal case but, in effect, creates a new set of residuals which generally lack an easy intuitive interpretation conducive to the heuristic approach to analysis. Theil [7,8] and Koerts [5] impose some restrictions on the transformation of least squares residuals in order to obtain some optimality properties for the set of transformed residuals which, however, do not seem to stand in a one-to-one correspondence with points of the design. The result is, for example, that an apparently anomalous transformed residual cannot be associated with some particular design point. Such disadvantages may, in turn, be largely avoided by selecting a linear transformation which retains this salient feature of the original least squares residuals--namely that each transformed residual, while representing a linear function of all $\underline{N}$ true errors, is clearly identified with a particular design point.

The choice of a particular transformation having this property will be influenced by the statistician's objective in examining residuals. Motivation for the analysis of residuals is commonly a suspicion directed toward some specific type of alternative to the homoscedastic hypothesis of independent, normally and/or identically distributed errors. An example is the suspicion that variance is a monotonic function of the mean, implying in this context that error variance $\sigma^2_{y\cdot x}$ is a monotonic function of $x\beta$. Independent transformed residuals would facilitate such heuristic approaches as the half-normal plot of residuals, the rank correlation between absolute or squared residuals and $x\hat{\beta}$ or between squared residuals and their expected values with respect to any specified heteroscedastic model, or the "peak-test" of heteroscedasticity as developed by Goldfeld and Quandt [1,2]. For further discussion of the use of transformed residuals see [5], [6], [7], and [8].

As shown in Section 2, an uncorrelated set of residuals retaining essentially the same intuitive appeal as the original residuals may be obtained by a stepwise fitting of the linear model to successively more observations. Thus, if $x_n\hat{\beta}_{(n)}$ is the predicted value of Y_n calculated by fitting only the first n observations $Y_1, \ldots, Y_n$ to the linear model $Y - X\beta + \epsilon$ then, excluding all n for which $x_n\hat{\beta}_{(n)} \equiv Y_n$, the residuals in the sequence

$$f_n = Y_n - x_n\hat{\beta}_{(n)}$$

are linearly uncorrelated if the components of ϵ are uncorrelated and homoscedastic. The degenerate case $Y_n \equiv x_n\hat{\beta}_{(n)}$ arises when inclusion of the n'th observation increases the rank of the design matrix (by unity)*. Normalizing scalars $c_n = \sigma_\epsilon/\sigma_{f_n}$ are known constants and the residuals $d_n = c_n f_n$ are then linearly uncorrelated with common variance $\sigma^2_{y\cdot x} = \sigma^2_\epsilon$, and if the ϵ-distribution is normal then so is the d-distribution.

*If the r degeneracies occur at $Y_1, \ldots, Y_r$ then f_n could be defined as $Y_n - x_n\hat{\beta}_{(n-1)}$ for $n > r$ in order to give more weight to ϵ_n; in any other case, however, the f_n so defined would depend on β as well as ϵ.

The set of numbers $\{d_n\}$ obtained in this manner depends upon the ordering imposed on the set of N observations; for a given set of N observations there are n!/r! possible sets $\{d_n\}$. The choice of a particular set, again, will depend upon the statistician's objective in analyzing residuals. Fortunately, this choice may be made to depend upon calculated values of the regression functions, $X\hat{\beta}_{(n)}$, for any n, without affecting the probability distribution of $\{d_n\}$ under the homoscedastic normal hypothesis. Since residuals are statistically independent of estimated regression functions then in constructing a set of residuals $d_{r+1},\dots,d_N$ to test for monotonic heteroscedasticity, for example, d_N may be chosen as the normalized residual associated with the largest of the N predicted values $X\hat{\beta}$. Similarly, Y_{N-1} may be defined as the observed Y at the design point corresponding to the second largest of $X\hat{\beta}$, and so on.

In the simplest and, in the present context, degenerate case where $Y_1,\dots,Y_N$ are assumed to be identically distributed, say $Y_i = \alpha + \epsilon_i$, the N predicted values $X\hat{\beta}$ are identically $\hat{\alpha}$. For any given ordering of the observations, specified by some external consideration, the sequence $d_2,\dots,d_N$ becomes the Helmert statistics as employed by Hogg, for example, in his heuristic method of iterated tests for equality of means [4]. We note that his iterative scheme may in general be applied to the sequence of test statistics

$$F_{1,i-r-1} = \frac{(i-r-1)d_i^2}{d_{r+1}^2 + \dots + d_{i-1}^2}$$

to test the sequence of nested hypotheses

$$H_{r+2}\colon \sigma^2_{\epsilon_1} = \dots = \sigma^2_{\epsilon_{r+2}};\ H_{r+3}\colon \sigma^2_{\epsilon_1} = \dots = \sigma^2_{\epsilon_{r+3}};\ H_N\colon \sigma^2_{\epsilon_1} = \dots = \sigma^2_{\epsilon_N}.$$

If H_k is true (and the ϵ's are normally distributed) then $F_{1,1},\dots,F_{1,k-r-1}$ are mutually independent and distributed as Snedecor's F with the indicated degrees of freedom.

In the case of simple linear regression, $Y_i = \alpha + \beta x_i + \epsilon_i$, the procedure outlined above reduces to ordering the Y's according to the rank order of the x's. Thus, if the alternative hypothesis is that $\sigma^2_{y \cdot x}$ is an increasing function of x then when $\hat{\beta}_{(N)}$ is positive,

$$f_n = Y_n - \bar{Y}_{(n)} - \hat{\beta}_{(n)}(x_n - \bar{x}_{(n)}) \ , \ n = 3,4,\dots,N$$

where $x_1 \leq x_2 \leq \dots \leq x_n$ and $\bar{Y}_{(n)}$, $\bar{x}_{(n)}$, $\hat{\beta}_{(n)}$, are the sample means and simple linear regression coefficient calculated from $(x_1,Y_1),\dots,(x_n,Y_n)$. The normalizing scalars in this case become

$$c_n = \left[1 - \frac{1}{n} - \frac{(x_n - \bar{x}_{(n)})^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x}_{(n)})^2} \right]^{-\frac{1}{2}} .$$

If $\hat{\beta}_{(N)}$ were negative then the ordering would be reversed.

From the heuristic point of view the above transformation of residuals would be especially satisfactory for testing against the alternative hypothesis that $\sigma^2_{y \cdot x_i}$ increases with $\alpha + \beta x_i$ if it were true that under this alternative model the transformed residuals $\{d_n\}$ had the corresponding monotonic property, $\sigma^2_{d_{n+1}} \leq \sigma^2_{d_{r+2}} \leq \dots \leq \sigma^2_{d_N}$. Evidently this property cannot be guaranteed in general when the ordering of the observations is determined by the rank order in $X\hat{\beta}$, which by chance may differ from the rank order in $X\beta$; and, unfortunately, even with a correct ordering of the observations there are design configurations for which monotonicity of $\sigma^2_{y \cdot x}$ is not sufficient to guarantee monotonicity of the sequence $\{\sigma^2_{d_n}\}$. Counterexamples violating this property are easily constructed with simple linear regression models; in the important special case of simple linear regression with equally spaced values of x, however, the monotonicity is preserved. This fact is demonstrated in Section 3.

The utility of independent residuals is illustrated in Section 4 where the "peak-test" developed by Goldfeld and Quandt [1] is applied to simulated residuals from a simple linear regression with equally spaced values of x. This test was devised to detect monotonic trends in a sequence of random variables, and the distribution of the peak-test statistic was tabulated for the case of independent and identically distributed (continuous) random variables. Monte Carlo computations are given here, comparing the properties of the peak-test applied to $|d_3|, \ldots, |d_N|$ and applied (as in Goldfeld and Quandt) to the original residuals $|e_1|, \ldots, |e_N|$.

2. ZERO CORRELATION BETWEEN OLD AND NEW RESIDUALS WHEN ADDITIONAL OBSERVATIONS ARE INCORPORATED INTO A LINEAR (MULTIPLE) REGRESSION ANALYSIS.

Let us rewrite the model (1) in the following form

$$\begin{bmatrix} Y_{(1)} \\ Y_{(2)} \end{bmatrix} = \begin{bmatrix} X_{(1)} \\ X_{(2)} \end{bmatrix} \beta + \begin{bmatrix} \epsilon_{(1)} \\ \epsilon_{(2)} \end{bmatrix} \qquad (2)$$

where $Y_{(1)}$ and $Y_{(2)}$ contain n and N-n observations respectively. Now suppose that we ignore the $Y_{(2)}$ observations and fit only the n observations $Y_{(1)}$; viz.,

$$Y_{(1)} = X_{(1)}\beta + \epsilon_{(1)} \quad . \qquad (3)$$

If H is a generalized inverse of $X'_{(1)}X_{(1)}$, i.e., $X'_{(1)}X_{(1)}HX'_{(1)}X_{(1)} = X'_{(1)}X_{(1)}$ where $X'_{(1)}$ denotes the transpose of $X_{(1)}$, the least squares estimate $f_{(1)}$ of $\epsilon_{(1)}$ from model (3) will be

$$f_{(1)} = Y_{(1)} - X_{(1)}HX'_{(1)}Y_{(1)} \quad . \tag{4}$$

If we now fit the entire N observations to the model, and if G is a generalized inverse of X'X, then the least squares estimate $e_{(2)}$ of $\epsilon_{(2)}$ will be

$$e_{(2)} = Y_{(2)} - X_{(2)}GX'_{(1)}Y_{(1)} - X_{(2)}GX'_{(2)}Y_{(2)} \quad . \tag{5}$$

Note that while $f_{(1)}$ is a function of $Y_{(1)}$ only, $e_{(2)}$ is a function of both $Y_{(1)}$ and $Y_{(2)}$.

We now prove the following theorem

THEOREM 2.1. $f_{(1)}$ *and* $e_{(2)}$ *are linearly uncorrelated (independent) if the components of* ϵ *are independent and identically (normally) distributed.*

To prove the theorem we need the following well-known lemma

LEMMA 2.1. Let W *be a* p x q *matrix. Then if* K *is a generalized inverse of* W'W, *then* WKW'W = W .

Proof of Theorem. $f_{(1)}$ and $e_{(2)}$ can be expressed as follows:

$$f_{(1)} = Y_{(1)} - X_{(1)}HX'_{(1)}Y_{(1)} = (I_n - X_{(1)}HX'_{(1)})\epsilon_{(1)}$$

$$\begin{aligned} e_{(2)} &= Y_{(2)} - X_{(2)}GX'_{(1)}Y_{(1)} - X_{(2)}GX'_{(2)}Y_{(2)} \\ &= (I_{N-n} - X_{(2)}GX'_{(2)})\epsilon_{(2)} - X_{(2)}GX'_{(1)}\epsilon_{(1)} \quad . \end{aligned}$$

Now if the components of ϵ are independent and identically distributed, then the covariance between $f_{(1)}$ and $e_{(2)}$ is

$$\Omega = E(f_{(1)}e'_{(2)}) = [I_n - X_{(1)}HX'_{(1)}]E(\epsilon_{(1)}\epsilon'_{(2)})[I_{N-n} - X_{(2)}G'X'_{(2)}]$$
$$- [I_n - X_{(1)}HX'_{(1)}]E(\epsilon_{(1)}\epsilon'_{(1)})X_{(1)}G'X'_{(2)}$$
$$= -[X_{(1)}G'X'_{(2)} - X_{(1)}HX'_{(1)}X_{(1)}G'X'_{(2)}]\sigma^2_\epsilon \,.$$

Since by lemma 2.1 $X_{(1)}HX'_{(1)}X_{(1)} = X_{(1)}$, $\Omega \equiv 0$. The independence of $f_{(1)}$ and $e_{(2)}$ is obvious under normality assumptions.

The following corollary is an immediate consequence of theorem 2.1.

COROLLARY 2.1. *If $x_n\hat{\beta}$ is the predicted value of Y_n calculated by fitting only the first n observations $Y_1, Y_2, \ldots, Y_n$ to the linear model $Y = X\beta + \epsilon$, then, excluding all n for which $x_n\hat{\beta}_{(n)} \equiv Y_n$, the residuals in the sequence*

$$f_n = Y_n - x_n\hat{\beta}_{(n)} \qquad , \qquad n = r+1, \ldots, N$$

are linearly uncorrelated (independent) if the components of ϵ are independent and identically (normally) distributed.

The proof follows by a successive application of theorem 2.1.

Remark. The residual f_n is a linear combination of $\epsilon_1, \ldots, \epsilon_n$, say $f_n = \sum_{i=1}^{n} \omega_{ni}\epsilon_i$. The weights ω_{ni} were calculated for several common designs, and usually ω_{nn} turned out to be larger (and sometimes much larger) than the other weights ω_{ni} (i-1,...,n-1). Exceptions to this rule exist, but it seems that in many cases $\omega_{nn}\epsilon_n$ is indeed the dominant term in f_n, and thus f_n yields considerable information about the true residual ϵ_n.

3. THE MONOTONICITY PROPERTY OF THE VARIANCE OF d_n IN HETEROSCEDASTIC SIMPLE REGRESSION MODELS

If the errors ϵ_i in the simple linear regression model $Y_i = \alpha + \beta x_i + \epsilon_i$ are normally and independently distributed with mean 0 and variance $\sigma^2_{y \cdot x_i} = \sigma^2_i$ then the transformed residuals $f_3, \ldots, f_N$ are likewise normal with mean 0 and variances

$$\sigma^2_{f_n} = \sum_1^n \left[\frac{1}{n} + \frac{(x_i - \bar{x}_{(n)})(x_n - \bar{x}_{(n)})}{\sum_1^n (x_i - \bar{x}_{(n)})^2} \right]^2 \sigma^2_i + \left[1 - \frac{2}{n} - 2 \frac{(x_n - \bar{x}_{(n)})^2}{\sum_1^n (x_i - \bar{x}_{(n)})^2} \right] \sigma^2_n \quad .$$

When the error variance $\sigma^2_{y \cdot x}$ is an increasing function of x the condition $x_1 < \ldots < x_n$, which implies $\sigma^2_1 \leq \ldots \leq \sigma^2_n$, is not sufficient to ensure that the normalized residuals

$$d_n = f_n \Bigg/ \sqrt{1 - \frac{1}{n} - \frac{(x_n - \bar{x}_{(n)})^2}{\sum_1^n (x_i - \bar{x}_{(n)})^2}}$$

will have increasing variances; however, if $x_i = d + bi$, $b > 0$, then the following theorem obtains:

<u>THEOREM 3.1.</u> *Let $\{\sigma^2_i\}$ be an increasing sequence. Then $\left\{\sigma^2_{d_n}\right\}$ is an increasing sequence if x's are equally spaced.*

<u>Proof.</u> The variance $\sigma^2_{d_n}$ in this case becomes

$$\sigma^2_{d_n} = \sum_1^{n-1} \frac{(6i - 2 - 2n)^2}{n(n^2 - 1)(n - 2)} \sigma^2_i + \frac{(n - 2)(n - 1)}{n(n + 1)} \sigma^2_n$$

and

$$\sigma^2_{d_{n+1}} - \sigma^2_{d_n} = \sum_{1}^{n-1} \frac{-144i^2 + i(72n + 144) - 4(n + 2)(2n + 5)}{n(n^2 - 1)(n^2 - 4)}\, \sigma^2_i$$

$$+ \frac{(n - 1)(20 - n^2)}{n(n + 1)(n + 2)}\, \sigma^2_n + \frac{n(n - 1)}{(n + 1)(n + 2)}\, \sigma^2_{n+1}$$

$$= \sum_{i=1}^{n+1} \delta_{ni}\sigma^2_i \quad \text{(say)} \quad .$$

Note that $\sum_{i=1}^{n+1} \delta_{ni} = 0$, because of the normalization, and that for $n \geq 5$

$$\delta_{ni} > 0 \text{ for } \begin{cases} i = n + 1 \\ \left[\dfrac{3n + 6 - \sqrt{n^2 - 4}}{12}\right] < i \leq \left[\dfrac{3n + 6 + \sqrt{n^2 - 4}}{12}\right] = \left[\dfrac{n + 1}{3}\right] \end{cases}$$

and $\delta_{ni} < 0$, otherwise (where $[\underline{a}]$ denotes the integer part of $\underline{a}$). This information concerning the signs of the δ_{ni} implies that

$$\sum_{i=1}^{k} \delta_{ni} \leq \sum_{i=1}^{\left[\frac{n+1}{3}\right]} \delta_{ni} \quad \text{for } \left[\frac{n+1}{3}\right] < k \leq n$$

where

$$\sum_{i=1}^{\left[\frac{n+1}{3}\right]} \delta_{ni} = \left\{ \begin{array}{ll} -\dfrac{4(n^2-n-2)}{9n(n-1)(n^2-4)} & \text{if } \left[\dfrac{n+1}{3}\right] = \dfrac{n+1}{3} \\ -\dfrac{4(n^2-6)}{9(n^2-1)(n^2-4)} & \text{if } \left[\dfrac{n+1}{3}\right] = \dfrac{n}{3} \\ -\dfrac{4(n^2+n-2)}{9n(n+1)(n^2-4)} & \text{if } \left[\dfrac{n+1}{3}\right] = \dfrac{n-1}{3} \end{array} \right\} < 0 \text{ if } n \geq 5 \, .$$

The monotonicity property $\sigma^2_{d_{n+1}} - \sigma^2_{d_n} = \sum_{i=1}^{n+1} \delta_{ni}\sigma_i^2 \geq 0$ then follows from

LEMMA 3.1. For any non-null vector $\delta' = (\delta_1,\dots,\delta_{n+1})$ *such that* $\sum_1^{n+1} \delta_i = 0$, *a necessary and sufficient condition for* $\delta'\sigma^2$ *to be non-negative (positive) for every vector* $\sigma^{2\prime} = (\sigma_1^2,\dots,\sigma_{n+1}^2)_i$ *with* $\sigma_i^2 \leq \sigma_{i+1}^2 (\sigma_i^2 < \sigma_{i+1}^2)$ *for* $i = 1,\dots,n$, *is that the partial sums* $D_i = \sum_{j=1} \delta_j$ *be non-positive for* $i = 1,\dots,n$.

Proof of sufficiency: Since $D_i \leq 0$ and not all $D_i = 0$ for $i = 1,\dots,n$, then

$$\delta'\sigma^2 \equiv \sum_1^n (\sigma_i^2 - \sigma_{i+1}^2)D_i + D_{n+1}\sigma^2_{n+1}$$

$$= \sum_1^n (\sigma_i^2 - \sigma_{i+1}^2)D_i \geq 0$$

with strict inequality if $0 < \sigma_1^2 < \dots < \sigma^2_{n+1}$.

Proof of necessity: Suppose $D_{i*} > 0$ for some $i*$, $1 \leq i* \leq n$, then a vector σ^2 satisfying

$$\sigma^2_{i+1} - \sigma_i^2 = \begin{cases} 1 & \text{if } \ i \neq i* \\ \dfrac{1}{D_{i*}}\left(1 + \sum_{i \neq i*} |D_i|\right) & \text{if } \ i = i* \end{cases}$$

also satisfies the conditions of the lemma, but in this case

$$\delta'\sigma^2 = \sum_{1}^{n} (\sigma_i^2 - \sigma_{i+1}^2)D_i \leq -1$$

in contradiction to the assumption $\delta'\sigma^2 \geq 0$.

Calculation of the numerical values of δ_{ni} for $n = 3$ and $n = 4$ reveals that the conditions of the lemma are also satisfied in these cases. Thus, the monotonicity preserving property holds for all n when $x_i = a + bi$, $b > 0$, and the correct direction of monotonicity is preserved provided only that $\text{sgn}(\hat{\beta}) = \text{sgn}(\beta)$.

4. SIMULATION OF THE "PEAK-TEST" OF HOMOSCEDASTICITY IN SIMPLE LINEAR REGRESSION

Goldfeld and Quandt [1] discuss the problem of testing homoscedasticity against a monotone heteroscedastic alternative hypothesis, and present tabulated critical values for a so-called "peak-test" of the residuals. A "peak" residual is said to occur at $\hat{Y}_j = \hat{\alpha} + \hat{\beta}x_j$, $\hat{Y}_1 < \dots < \hat{Y}_N$, if and only if $|Y_i - \hat{Y}_i| < |Y_j - \hat{Y}_j|$ for all $i < j$, and the peak-test statistic is then defined as the number of peaks occurring among $|Y_2 - \hat{Y}_2|, \dots, |Y_N - \hat{Y}_N|$. Critical values are obtained from the tabulated distribution of the number of peaks occurring in a random sample of size N from an absolutely continuous distribution.

In their original application of the peak-test to simple linear regression residuals, Goldfeld and Quandt [1] failed to take into account both the dependence which exists between residuals and the fact that the distribution of $Y_i - \hat{Y}_i$ is a function of x_i; under the homoscedastic hypothesis the stochastically largest absolute residual occurs with the x_i nearest to $\bar{x}$. If sample size is large then these shortcomings of their procedure are minor,as the authors later pointed out [2]; however, as sample size increases, the mechanics of performing the peak-test become unduly time consuming and a computationally simpler procedure such as the F-test described by Goldfeld and Quandt becomes more expedient. For small samples, the peak-test applied to the untransformed residuals is clearly invalid with respect to the size of the test, and also has poor power characteristics.

Table 1 illustrates these points for sample size N = 10, and also indicates how they are overcome by applying the peak-test to normalized, stepwise residuals. The cumulative distribution of number of peaks for selected, monotone heteroscedastic alternative hypotheses was estimated by generating 1000 samples of size N = 10 from the standard normal distribution and transforming to heteroscedastic errors by appropriate scale changes. After scale changes the least squares residuals and stepwise least squares residuals were then constructed as appropriate linear functions of the errors; each sample of size N = 10 was thus used in all eight columns of observed values in Table 1. The columns labeled "H_o Nominal c.d.f." and "H_o Exact c.d.f." were calculated from recursion formulae presented by Goldfeld and Quandt for the exact probability distribution of number of peaks in random samples of size 10 and 8, respectively. Note that the homoscedastic exact and observed distribution of peaks in normalized stepwise residuals stand in close agreement, as expected, providing a crude guide as to the amount of precision inherent in the other columns of observed probabilities.

Table 1. Monte Carlo results comparing the properties of the Goldfeld-Quandt "peak-test" applied to least squares residuals and to independent, normalized stepwise residuals in simple linear regression with N = 10 and $x_i = i$.

	Distribution of the number of peaks among the last 9 least squares residuals					Distribution of the number of peaks among the last 7 normalized stepwise residuals				
Number of peaks	H_o Nominal c.d.f.	Observed cumulative distribution				H_o Exact c.d.f.	Observed cumulative distribution			
		$\sigma^2_{y \cdot x} = 1$	$\sigma^2_{y \cdot x} = 2x$	$\sigma^2_{y \cdot x} = x^2$	$\sigma^2_{y \cdot x} = \ln x$		$\sigma^2_{y \cdot x} = 1$	$\sigma^2_{y \cdot x} = 2x$	$\sigma^2_{y \cdot x} = x^2$	$\sigma^2_{y \cdot x} = \ln x$
0	.100	.305	.361	.380	.352	.1250	.117	.002	.030	.034
1	.3829	.671	.696	.758	.716	.4456	.428	.070	.202	.244
2	.7061	.888	.864	.887	.876	.7707	.771	.341	.551	.604
3	.9055	.972	.957	.947	.959	.9385	.944	.709	.837	.882
4	.9797	.993	.992	.986	.991	.9871	.994	.922	.968	.976
5	.9971	.998	.998	.998	.997	.9950	.999	.989	.995	.996
6	.9997	1.000	1.000	1.000	.998	.9956	1.000	1.000	1.000	1.000
7	.9999	1.000	1.000	1.000	1.000	1	1	1	1	1
8	1.0000	1.000	1.000	1.000	1.000	-	-	-	-	-
9	1	1	1	1	1	-	-	-	-	-

The "peak-test" which treats the least squares residuals as if they were independent and identically distributed errs substantially in the size of the test. Thus, taking 4 or more peaks among the 10 residuals as the critical region gives a nominal significance level $\alpha_4 = 1 - .9055 = .0945$ while the actual size of the test is approximately $1 - .9710 \approx .03$; and if any of the three heteroscedastic models obtained then the probability of rejecting homoscedasticity would be at best approximately .05 (less than the nominal size of the test). Applied to normalized stepwise residuals the critical region of 4 or more peaks has size $1 - .9385 = .0615$, and the probability of detecting the alternative $\sigma^2_{y \cdot x} = 2x$ is approximately $1 - .709 \approx .29$.

Since only the errors ϵ were simulated in this Monte Carlo operation the estimated distributions under the heteroscedastic models in Table 1 must be regarded as estimates of conditional probabilities, the condition being that $\mathrm{sgn}(\hat{\beta}) = \mathrm{sgn}(\beta)$. With independent, normally distributed heteroscedastic errors

$$P\left(\mathrm{sgn}(\hat{\beta}) = \mathrm{sgn}(\beta)\right) = \Phi\left(\frac{|\beta| \sum_{1}^{N} (x_i - \bar{x})^2}{\sqrt{\sum_{1}^{N} (x_i - \bar{x})^2 \sigma^2_{y \cdot x_i}}} \right)$$

where $\Phi(\cdot)$ denotes the standard cumulative normal distribution.

ACKNOWLEDGMENT

The authors are grateful to the referees, associate editor, and editor for several helpful suggestions.

References

[1] Goldfeld, S. M. and Quandt, R. E. "Some tests for homoscedasticity," Journal of the American Statistical Association, 60(1965)539-47.

[2] Goldfeld, S. M. and Quandt, R. E. Corrigenda, Journal of the American Statistical Association, 62(1967)1518.

[3] Hogg, R. V. "On the resolution of statistical hypotheses," Journal of the American Statistical Association, 56(1961)978-89.

[4] Hogg, R. V. "Iterated tests of the equality of several distributions," Journal of the American Statistical Association, 57(1962)579-85.

[5] Koerts, J. "Some further notes on disturbance estimates in regression analysis," Journal of the American Statistical Association, 62(1967) 169-83.

[6] Putter, J. "Orthonormal bases of error spaces and their use for investigating the normality and variances of residuals," Journal of the American Statistical Association, 62(1967)1022-36.

[7] Theil, H. "The analysis of disturbances in regression analysis," Journal of the American Statistical Association, 60(1965)1067-79.

[8] Theil, H. "A simplification of the BLUS procedure for analyzing regression disturbances," Journal of the American Statistical Association, 63(1968) 242-51.