EXAMINATION AND ANALYSIS OF RESIDUALS:
DIAGNOSTIC CHECKING OF RESIDUALS FOR DETECTING
A SPECIAL TYPE OF HETEROSCEDASTICITY IN LINEAR REGRESSION
THROUGH THE ORIGIN

A. Hedayat
Cornell University, Ithaca, New York

Abstract

The main purpose of this paper is to use the
Theil residuals for detecting a monotonic relation
between mean and variance by means of the peak test
applied to linear regression through the origin. A
numerical illustration is provided.
EXAMINATION AND ANALYSIS OF RESIDUALS:
DIAGNOSTIC CHECKING OF RESIDUALS FOR DETECTING
A SPECIAL TYPE OF HETEROSEDASTICITY IN LINEAR REGRESSION
THROUGH THE ORIGIN

A. Hedayat
Cornell University, Ithaca, New York

SUMMARY

The main purpose of this paper is to use the Theil residuals for
detecting a monotonic relation between mean and variance by means of the
peak test applied to linear regression through the origin. A numerical
illustration is provided.

1. INTRODUCTION

Consider the simple linear model $Y = X\beta + \epsilon$ where $Y$ represents an
n-dimensional random vector, $X$ is an n-dimensional column vector with known
coefficients consisting of nonstochastic elements, or elements which are
distributed independently of the error terms, $\beta$ is an unknown scalar, and $\epsilon$
is an n-dimensional random vector having multivariate normal distributions
with

$$EE' = C, \quad E\epsilon' = \sigma^2 I_n$$

where $\sigma^2 > 0$ is an unknown parameter and $I_n$ is used to denote the $n \times n$
identity matrix.

Research was supported by NIH GM-05900
The least square (LS) estimates $\hat{\beta}$ of $\beta$ and $\hat{\epsilon}$ of $\epsilon$ are

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} \quad \text{and} \quad \hat{\epsilon} = Y - X\hat{\beta} = Y - X \left( \frac{\sum x_i y_i}{\sum x_i^2} \right).$$

where

$$P = I_n - \left( \frac{1}{\sum x_i^2} \right) X' X.$$

Under the above assumptions

$$E\hat{\epsilon} = PX\beta = 0 \beta = 0$$

$$E\hat{\epsilon}\epsilon' = PEYY'P = PE\epsilon\epsilon'P = \sigma^2P.$$  

We see that even if assumption (1) is true, the LS estimates of residuals are neither independent nor do they have constant variance, since $P \neq I_n$.

Theil [1965] has presented an estimator of $\epsilon$ which has all the ordinary properties of $\hat{\epsilon}$ except that the covariance of the Theil estimator $\sigma^2 I_{n-1}$ under assumption (1). The dimension of the Theil estimator of $\epsilon$ is $n-1$ due to the fact that the residual has $n-1$ degrees of freedom.

While Theil [1965] has given the general procedure for deriving uncorrelated residuals with constant variance under the homoscedasticity assumption in multiple linear regression, Koerts [1967] has derived the explicit form of the Theil estimator for the simple linear model through the origin. The
latter will be used here in our diagnostic checking of residuals to detect a monotonic relation between mean and variance by means of the peak test introduced by Goldfeld and Quandt [1965].

2. THEIL ESTIMATOR OF RESIDUALS

We denote this estimator by $\hat{\epsilon}$ and for a simple linear model through the origin it can be represented simply as

$$\hat{\epsilon}_i = y_i - b^* x_i \quad i=1,2,\ldots,k-1, k+1,\ldots,n$$

where

$$b^* = (1-a) \left( \frac{\hat{\beta}_{n-1} + a y_k}{x_k} \right)$$

and

$$a = \frac{|x_k|}{\left( \sum_{i=1}^{n} x_i^2 \right) \frac{1}{3}} \quad \text{and} \quad \hat{\beta}_{n-1} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

$k$ can take any value from 1 to $n$ and the choice is largely a matter of power with respect to a specific alternative hypothesis. Properties of Theil residuals are:

1. $\hat{\epsilon}_i$ is a linear function of $y$,

2. $\mathbb{E}\hat{\epsilon}_i = 0, \quad i=1,2,\ldots,k-1, k+1,\ldots,n$,

3. $\text{Cov}(\hat{\epsilon}_i, \hat{\epsilon}_j) = 0$ if $i \neq j \quad i,j = 1,2,\ldots,k-1, k+1,\ldots,n$,

   $= \sigma^2$ if $i = j$
(4) have the minimum expected sum of squares in the class of estimators with properties 1, 2, and 3,

\[ \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 . \]

Properties (3) and (5) make Theil residuals very interesting indeed. Theil residuals have been derived based on the first four properties. Koerts [1967] has shown that Theil residuals also have the fifth property.

3. APPLICATION OF THEIL RESIDUALS

Consider the case where x's have been ordered such that \( x_i < x_j \) if \( i < j \). And suppose we are interested in testing the following hypothesis:

\[ H_0 : E\varepsilon_i^2 = \sigma^2 \quad \text{for all } i \]

versus

\[ H_1 : E\varepsilon_i^2 = \sigma_i^2 < E\varepsilon_j^2 = \sigma_j^2 \quad \text{for } i < j . \]

Note that alternative hypothesis says that as \( x \) increases the variance of \( \varepsilon \) or \( y \) increases also. We are considering the case where we have only a single observation for each level of \( x \), as is the case in most experiments.

Two alternative tests for testing \( H_0 \) against \( H_1 \) are suggested by Goldfeld and Quandt [1965].
(1) F test

The obvious choice for k is then the middle observation, so that one can compute the ratio of the sum of squares of the first \((n-1)/2\) estimated residuals to that of the last \((n-1)/2\), which is F distributed. When \(n-1\) is not even, one can use either \((n-2)/2\) first and \(n/2\) last observations or \(n/2\) first and the \((n-2)/2\) last observations, and for this choice see Theil [1965].

(2) Peak Test

While the F test is a general test, the peak test has been constructed specifically for testing \(H_0\) against the particular \(H_1\) under consideration and therefore one expects to obtain greater sensitivity from this test than the F test. Especially when the number of observations is small the greater sensitivity of the peak test becomes important.

The idea of this test was originally given by Goldfeld and Quandt [1965]. For residuals ordered by the ordering of \(x_i, x_i < x_{i+1}\), they define a peak at \(x_i\) to be an instance where \(|\hat{\epsilon}_i^*| > |\hat{\epsilon}_j^*|\) for \(j = 1, 2, ..., i-1\).

Hedayat and Robson [1966] have demonstrated the failure of the Goldfeld and Quandt peak test applied to the LS residuals \(\hat{\epsilon}\), but application to \(\hat{\epsilon}^*\) is valid and appropriate because

(a) under \(H_0\), \(\hat{\epsilon}_i^*\)'s are uncorrelated and therefore under the normality assumption will be independent,

(b) under \(H_1\), \(\text{var} \hat{\epsilon}_i^* (= \text{var} \hat{\epsilon}_i) < \text{var} \hat{\epsilon}_{i+1}^* (= \text{var} \hat{\epsilon}_{i+1})\) and hence one expects \(\hat{\epsilon}_i^* < \hat{\epsilon}_{i+1}^*\) or equivalently \(|\hat{\epsilon}_i^*| < |\hat{\epsilon}_{i+1}^*|\).

Proof of part (b) stems from the following theorem, a proof of which is given in the appendix.
Theorem: If $E e_i e_j = 0$, $i \neq j$ and $E e_i^2 = \sigma_i^2 < E e_i e_j = \sigma_j^2$, then $\text{var } e_i < \text{var } e_j$ for $i < j$.

4. A NUMERICAL ILLUSTRATION

We apply the peak test using Theil residuals to the example (see table 1) given on page 180 of Steel and Torrie [1960]. As Steel and Torrie [1960] have pointed out, in this instance regression of $y$ on $x$ should pass through the origin. Therefore,

$$\beta = \frac{\sum_{i=1}^{13} x_i y_i}{\sum_{i=1}^{13} x_i^2} = 3.67$$

and hence the regression line is given by $y = 3.67x$. The reduction in sum of squares attributable to regression is

$$\left( \frac{\sum_{i=1}^{13} x_i y_i}{\sum_{i=1}^{13} x_i^2} \right)^2 = 351,819.$$

And the residual sum of squares is $356,259 - 351,819 = 4,440$. The individual least square residuals are:

$\hat{e}_1 = + 2.088$
$\hat{e}_2 = - 3.013$
$\hat{e}_3 = - 5.437$
$\hat{e}_4 = - 4.952$
$\hat{e}_5 = - 16.888$
$\hat{e}_6 = - 16.066$
$\hat{e}_7 = - 8.167$

$\hat{e}_8 = + 11.887$
$\hat{e}_9 = + 17.527$
$\hat{e}_{10} = + 29.756$
$\hat{e}_{11} = + 19.716$
$\hat{e}_{12} = - 44.688$
$\hat{e}_{13} = + 6.541$
TABLE 1

Induced Reversions to Independence per 10^7 Surviving Cells y per Dose (ergs/Bacterium) 10^{-5} x of Stroptorycin Dependent Escherichia Coli Subjected to Monochromatic Ultraviolet Radiation of 2,967 Angstroms Wavelength.

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.6</td>
<td>32</td>
</tr>
<tr>
<td>13.9</td>
<td>48</td>
</tr>
<tr>
<td>21.1</td>
<td>72</td>
</tr>
<tr>
<td>25.6</td>
<td>89</td>
</tr>
<tr>
<td>26.4</td>
<td>30</td>
</tr>
<tr>
<td>39.8</td>
<td>130</td>
</tr>
<tr>
<td>40.1</td>
<td>139</td>
</tr>
<tr>
<td>43.9</td>
<td>173</td>
</tr>
<tr>
<td>51.9</td>
<td>208</td>
</tr>
<tr>
<td>53.2</td>
<td>225</td>
</tr>
<tr>
<td>65.2</td>
<td>259</td>
</tr>
<tr>
<td>66.4</td>
<td>199</td>
</tr>
<tr>
<td>67.7</td>
<td>255</td>
</tr>
</tbody>
</table>

\[ \sum_{i=1}^{13} x_i = 528.8 \quad \sum_{i=1}^{13} y_i = 1,929 \]
\[ \sum_{i=1}^{13} x_i^2 = 26,062.10 \quad \sum_{i=1}^{13} y_i^2 = 356,259 \]
First of all, examination of these residuals suggests that there is a pattern for the distribution of plus and minus signs among the $\hat{e}_i$'s. Second, one gets the impression that the absolute value of $\hat{e}_i$ increases as $i$ increases. Now suppose we suspect the assumption $\sum (\hat{e}_i)^2 = \sigma^2$ for all $i$ and in particular we suspect that the variance may increase with the mean; i.e., that the variance of $y$ increases as $x$ increases. To test against this alternative hypothesis we first find the Theil residuals

$$\hat{e}_i = y_i - b^* x_i, \quad i=1,2,\ldots,12 \quad (k=13)$$

where

$$b^* = (1-a)\frac{\sum \hat{e}_i x_i}{\sum x_i^2} = 3.7014.$$ 

Thus,

$$\hat{e}_i = y_i - 3.7014 x_i, \quad i=1,2,\ldots,12.$$ 

and

$$\hat{e}_1 = +1.66096 \quad \hat{e}_7 = -9.42614$$
$$\hat{e}_2 = -3.44946 \quad \hat{e}_8 = +10.50854$$
$$\hat{e}_3 = -6.09954 \quad \hat{e}_9 = +15.89734$$
$$\hat{e}_4 = -5.75584 \quad \hat{e}_{10} = +28.08552$$
$$\hat{e}_5 = -17.71696 \quad \hat{e}_{11} = +17.66872$$
$$\hat{e}_6 = -17.31572 \quad \hat{e}_{12} = -46.77296$$

No. of peaks = 5

$$\sum_{i=1}^{12} \hat{e}_i^2 = -4,439.4133$$ where the slight departure from 4,440 is due to rounding error in computing $b^*$. 

\[\frac{12}{\hat{e}_i^2} = -\frac{1}{4,440}\]
The $\epsilon_i$'s are independent and identically distributed under the homoscedasticity and normality assumption of the $\epsilon_i$'s. Now we can compute the probability of obtaining five or more peaks in a sequence of 12 independent and identically distributed random variables using table 2 from Goldfeld and Quandt [1965].

By interpolation from this table we see that the probability is about .036 that a sequence of 12 independent and identically distributed random variables produces five or more peaks. If we can accept a risk of 3.6 percent and if our suspicion about homoscedasticity has biological support, then we should fit a weighted regression rather than the unweighted one for obtaining an efficient estimate of $\beta$ and hence the regression line.

5. DISCUSSION

If we neglect checking the textbook assumptions related to the linear model, this means either that we are willing to accept these assumptions or simply that we are not aware of their importance. In practice, however, there is rarely good reason for supposing that conventional assumptions are satisfied, and methods are therefore needed for detecting and measuring any sort of departure from these ideal conditions.

As Anscombe and Tukey [1960] have put it, "If we are to improve our analysis of data to which the conventional technique can be applied, it is not likely that we shall do this by improving the techniques themselves. Rather we must learn either to go further, beyond the place where the conventional techniques stop, or we must learn to use the technique better. Either path demands the analysis of residuals, where

\[
\text{(residual)} = \text{(observed value)} - \text{(fitted value)}
\]
<table>
<thead>
<tr>
<th>n</th>
<th>x = 0</th>
<th>x = 1</th>
<th>x = 2</th>
<th>x = 3</th>
<th>x = 4</th>
<th>x = 5</th>
<th>x = 6</th>
<th>x = 7</th>
<th>x = 8</th>
<th>x = 9</th>
<th>x = 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>.2000</td>
<td>.6167</td>
<td>.9083</td>
<td>.9917</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>.1000</td>
<td>.3829</td>
<td>.7061</td>
<td>.9055</td>
<td>.9797</td>
<td>.9971</td>
<td>.9997</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>.0667</td>
<td>.2834</td>
<td>.5833</td>
<td>.8211</td>
<td>.9433</td>
<td>.9866</td>
<td>.9976</td>
<td>.9997</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>.0500</td>
<td>.2274</td>
<td>.5022</td>
<td>.7530</td>
<td>.9056</td>
<td>.9720</td>
<td>.9935</td>
<td>.9988</td>
<td>.9996</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>.0400</td>
<td>.1910</td>
<td>.4441</td>
<td>.6979</td>
<td>.8705</td>
<td>.9559</td>
<td>.9879</td>
<td>.9973</td>
<td>.9995</td>
<td>.9999</td>
<td>1.0000</td>
</tr>
<tr>
<td>30</td>
<td>.0333</td>
<td>.1654</td>
<td>.4001</td>
<td>.6525</td>
<td>.8386</td>
<td>.9395</td>
<td>.9815</td>
<td>.9953</td>
<td>.9990</td>
<td>.9998</td>
<td>1.0000</td>
</tr>
<tr>
<td>35</td>
<td>.0286</td>
<td>.1462</td>
<td>.3554</td>
<td>.6144</td>
<td>.8098</td>
<td>.9234</td>
<td>.9745</td>
<td>.9929</td>
<td>.9984</td>
<td>.9997</td>
<td>.9999</td>
</tr>
<tr>
<td>40</td>
<td>.0250</td>
<td>.1313</td>
<td>.3373</td>
<td>.5818</td>
<td>.7837</td>
<td>.9078</td>
<td>.9674</td>
<td>.9903</td>
<td>.9975</td>
<td>.9995</td>
<td>.9999</td>
</tr>
<tr>
<td>45</td>
<td>.0222</td>
<td>.1194</td>
<td>.3138</td>
<td>.5536</td>
<td>.7600</td>
<td>.8930</td>
<td>.9601</td>
<td>.9874</td>
<td>.9966</td>
<td>.9992</td>
<td>.9998</td>
</tr>
<tr>
<td>50</td>
<td>.0200</td>
<td>.1096</td>
<td>.2940</td>
<td>.5288</td>
<td>.7383</td>
<td>.8783</td>
<td>.9530</td>
<td>.9844</td>
<td>.9956</td>
<td>.9989</td>
<td>.9998</td>
</tr>
<tr>
<td>55</td>
<td>.0182</td>
<td>.1014</td>
<td>.2769</td>
<td>.5068</td>
<td>.7184</td>
<td>.8653</td>
<td>.9456</td>
<td>.9813</td>
<td>.9944</td>
<td>.9986</td>
<td>.9997</td>
</tr>
<tr>
<td>60</td>
<td>.0167</td>
<td>.0944</td>
<td>.2620</td>
<td>.4871</td>
<td>.7001</td>
<td>.8524</td>
<td>.9384</td>
<td>.9780</td>
<td>.9932</td>
<td>.9982</td>
<td>.9996</td>
</tr>
</tbody>
</table>
In the first path we analyze residuals to learn what they can tell us of direct interest. In the second path we must analyze the residuals from a first application of conventional methods to learn how a second application might be better made.

Least square residuals, even under ideal conditions, are in general correlated and have different variances. Perhaps for graphical examination of residuals in certain cases we can neglect both the covariance and heterogeneity of variances which exists among the least squares residuals, but certainly for constructing tests or for any rigorous examination of residuals, we prefer to work with a new type of residuals which are free from the above criticism.

As data analysis becomes more sophisticated, computation will surely get more extensive rather than simpler, and if sophisticated data analysts are to gain in depth and power, they must have both the time and stimulation to try out new procedures of analysis. We take advantage here to quote Tukey's [1962] words of wisdom, "The future of data analysis can involve great progress, the overcoming of real difficulties, and the provision of a great service to all fields of science and technology. Will it? That remains to us, to our willingness to take up the rocky road of real problems in preference to the smooth road of unreal assumptions, arbitrary criteria, and abstract results without real attachments. Who is for the challenge?"
ACKNOWLEDGEMENT

The author wishes to thank D. S. Robson for the suggestion that orthogonal residuals might preserve monotonicity of variance.

REFERENCES


We now prove that if \( E\epsilon_i \epsilon_j = 0, \ i \neq j \) and \( E\epsilon_i^2 = \sigma_i^2 \leq E\epsilon_{i+1}^2 = \sigma_{i+1}^2 \)

then \( \text{var } \epsilon_i^* < \text{var } \epsilon_{i+1}^* \). Under these assumptions and by the definition of \( \epsilon_i^* \) we have

\[
\text{Var } \epsilon_i^* = E\epsilon_i^*^2 - (E\epsilon_i^*)^2
\]

\[
= E\epsilon_i^*^2
\]

\[
= E \left[ y_i - x_i \hat{\beta}_{n-1} - ax_i \left( \hat{\beta}_{n-1} - \frac{y_n}{x_n} \right) \right]^2
\]

\[
= x_i^2 \beta^2 + \sigma_i^2 + \beta^2 \left[ x_i^2 (1 - a)^2 + x_i^2 a^2 + 2a (1 - a) x_i^2 \right]
\]

\[
= x_i^2 \beta^2 + \sigma_i^2 + \beta^2 \left[ \sum_{i=1}^{n-1} x_i^2 (1 - a) + \frac{x_i^2 a^2}{n-1} + 2a (1 - a) x_i^2 \right]
\]

\[
= x_i^2 + x_i^2 (1 - a)^2 + x_i^2 a^2 + 2a (1 - a) x_i^2 - 2x_i^2 (1 - a) - 2x_i^2 a
\]

\[
= x_i^2 \left[ \sum_{i=1}^{n-1} x_i^2 (1 - a) + \frac{x_i^2 a^2}{n-1} + 2a (1 - a) x_i^2 \right]
\]

\[
+ \sigma_i^2 + \frac{(1-a)^2 x_i^2}{n-1} \sum_{i=1}^{n-1} x_i^2 \sigma_i^2 + \frac{ax_i^2 a^2}{n-1} \sum_{i=1}^{n-1} x_i^2 - 2x_i^2 (1 - a) \frac{x_i^2 a^2}{n-1} x_i^2 a
\]
\[
\sigma_i^2 + A^2 x_i^2 \sum_{i=1}^{n-1} x_i \sigma_i^2 + Bx_i^2 - 2Ax_i \sigma_i^2
\]

where

\[
A = \frac{1-a}{n-1}, \quad B = \frac{\sigma_n^2(1-a)^2}{n-1}
\]

\[
= \sigma_i^2(Ax_i^2 + 1)^2 + x_i^2 A^2 \sum_{j=1}^{n-1} \Sigma_{j=1}^{i-1} x_j \sigma_j^2 + x_i^2 A^2 x_i x_i + Bx_i^2
\]

\[
= \sigma_i^2(Ax_i^2 + 1)^2 + CA^2 x_i^2 + A^2 x_i x_i + Bx_i^2
\]

where

\[
C = \sum_{j=1}^{n-1} \Sigma_{j=1}^{i-1} x_j \sigma_j^2
\]

Now

\[
\text{Var } \epsilon_{i+1} - \text{Var } \epsilon_i = \sigma_i^2 (Ax_{i+1}^2 + 1)^2 - \sigma_i^2(Ax_i^2 + 1)^2 = CA^2(x_{i+1}^2 - x_i^2)
\]

\[
+ A^2 x_i x_{i+1} (\sigma_i^2 - \sigma_{i+1}^2) + B(x_{i+1}^2 - x_i^2)
\]

For \( x_{i+1} > x_i \) we have
\[ A^2(x_{i+1}^2 - x_i^2) > 0 \quad \text{and} \quad B(x_{i+1}^2 - x_i^2) > 0 . \]

Now in order to show that \( \text{var } \varepsilon_{i+1}^2 > \text{var } \varepsilon_i^2 \), we have to show that

\[
\sigma_{i+1}^2 (A x_{i+1}^2 + 1)^2 - \sigma_i^2 (A x_i^2 + 1) + A^2 x_i^2 x_{i+1}^2 (\sigma_i^2 - \sigma_{i+1}^2) > 0 .
\]

That is,

\[
\sigma_{i+1}^2 A^2 x_i^2 x_{i+1}^2 - \sigma_i^2 A^2 x_i^2 x_{i+1}^2 + \sigma_{i+1}^2 (1 + 2A x_i^2)
\]

\[- \sigma_i^2 (1 + 2A x_i^2) + A^2 x_i^2 x_{i+1}^2 (\sigma_i^2 - \sigma_{i+1}^2) > 0 .
\]

Since \( x_i < x_{i+1} \), then the left side of the above final expression will be greater than

\[
\sigma_{i+1}^2 A^2 x_i^2 x_{i+1}^2 - \sigma_i^2 A^2 x_i^2 x_{i+1}^2 + \sigma_{i+1}^2 (1 + 2A x_i^2)
\]

\[- \sigma_i^2 (1 + 2A x_i^2) + A^2 x_i^2 x_{i+1}^2 (\sigma_i^2 - \sigma_{i+1}^2)
\]

\[= \sigma_{i+1}^2 (1 + 2A x_{i+1}^2) - \sigma_i^2 (1 + 2A x_i^2) > 0 . \]

Q.E.D.