**A note on the use of residuals for detecting an outlier in linear regression**

By AJIT C. TAMHANE

*Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois, U.S.A.*

**Summary**

Consider the usual linear regression model \( y = X\beta + \varepsilon \), where the vector \( \varepsilon \) has \( E(\varepsilon) = 0 \), \( \text{cov} (\varepsilon) = \sigma^2 V \), where \( V \) is known. Let \( e = y - \hat{y} \) be the least squares residual vector. It is shown that a test based on the transformed residual vector \( d^* = V^{-1} e \) has, in the class of linear transformations of \( e \), certain optimal power properties for detecting the presence of a single outlier when the label of the outlier observation is unknown. The outlier model considered here is that of shift in location.

Some key words: Linear regression; Outlier; Power; Residual.

Consider the usual full rank linear regression model

\[ y = X\beta + \varepsilon, \]

where \( y \) is an \( n \times 1 \) vector of dependent variables, \( X \) is an \( n \times r \) matrix of nonstochastic regressors, with \( r \leq n \), \( \beta \) is an \( r \times 1 \) vector of unknown parameters and \( \varepsilon \) is an \( n \times 1 \) vector of random errors with \( E(\varepsilon) = 0 \) and \( \text{cov} (\varepsilon) = \sigma^2 V \), where \( V \) is a known symmetric positive-definite matrix and \( \sigma^2 \) is an unknown positive scalar.

The least squares residual vector \( e \) is given by

\[ e = y - \hat{y} = y - X\hat{\beta} = \{I - X(X'V^{-1}X)^{-1}X'V^{-1}\} y. \]

Standardized residuals \( z_i = \{e_i/\sqrt{\text{var}(e_i)}\} \) are often used to detect outlier observations or gross errors. In this note we show that the transformed residual vector \( V^{-1} e \) has certain optimal power properties for detecting a single outlier when the experimenter is unaware that there is exactly one outlier present. Thus the usual tests based on \( e \) are less powerful for this situation.

To avoid the complicated distribution problems associated with studentized residuals and obtain the power results in an uncluttered and distribution-free manner, we shall assume that \( \sigma^2 \) is known and hence can be taken to be unity.

We consider the class of all linearly transformed residual vectors \( d = Ae \), where \( A \) is an \( n \times n \) nonsingular, nonrandom matrix. The outlier detection procedure will be as follows. Define a test vector \( z \) based on \( d \) by

\[ z_i = d_i/\sqrt{\text{var}(d_i)} \quad (i = 1, \ldots, n). \]  

Then declare the \( i \)th observation an outlier if \( |z_i| > k \), where \( k \) is a suitably chosen positive constant.

We consider an outlier model in which \( E(e_i) \neq 0 \) for some \( i (i = 1, \ldots, n) \), where the label \( i \) of the outlier observation is, of course, unknown to the experimenter. Without loss of generality we may take the \( n \)th observation to be an outlier. Thus let \( E(\varepsilon) = \delta \), where \( \delta_n \neq 0 \) but \( \delta_1 = \ldots = \delta_{n-1} = 0 \). Under this assumption we define an optimal test vector \( z^* \), or equivalently the corresponding \( d^* \) since \( z^* \) and \( d^* \) are related by (1), for detecting
the outlier as follows: \( z^* \), or equivalently the corresponding \( d^* \), is said to be an optimal test vector for the test \(| z_i | > k \) for detecting a single outlier if for all \( k > 0 \),

\[
\text{pr} (| z^*_n | > k) \geq \text{pr} (| z_n | > k)
\]  

(2)

for all \( z \),

\[
\text{pr} (| z^*_n | > k) > \text{pr} (| z^*_i | > k) \quad (i = 1, \ldots, n-1),
\]  

(3)

with a strict inequality in (2) for at least some \( z \).

Thus \( z^* \) has the property that the correct observation is declared an outlier with the highest possible probability. Preparatory to stating the main result we introduce some additional notation: let \( P \) be an \( n \times n \) nonsingular matrix such that

\[
P'P = V^{-1}, \quad B' = AP^{-1}, \quad M = I - PX(X'V^{-1}X)^{-1}X'P',
\]

where \( I \) is an \( n \times n \) identity matrix. Then it is easy to show that \( E(d) = B'M\delta \), and \( \text{cov} (d) = B'MB = C \), say. Also write \( \gamma_i = E(z_i) = (B'M\delta)_i/\sqrt{c_{ii}} \), where \( c_{ii} \) is the \( i \)th diagonal entry of \( C \). Now we state our main result.

**Theorem.** If for fixed \( k > 0 \), \( \text{pr} (| z_i | > k) \) is an increasing function of \( | \gamma_i | \) for \( i = 1, \ldots, n \), then the optimal test vector for detecting a single outlier is given by \( d^* = V^{-1} e \), that is \( A^* = V^{-1} \).

Note that the assumption that \( \text{pr} (z_i > k) \) is an increasing function of \( | \gamma_i | \) \((i = 1, \ldots, n)\) is true, e.g. under the normality assumption for \( \varepsilon \).

**Proof.** Let \( Q = B^{-1} P \) and let \( p_i, q_i, b_i \) and \( c_i \) be the \( i \)th column vectors of \( P, Q, B, \) and \( C \) respectively. Then for \( i = 1, \ldots, n \)

\[
\gamma_i = \frac{(CQ\delta)_i}{\sqrt{(b_i'Mb_i)}} = \frac{\delta_n c_i^* q_n}{\sqrt{(b_i'Mb_i)}} = \frac{\delta_n b_i'Mp_n}{\sqrt{(b_i'Mb_i)}}
\]

when \( \delta_1 = \ldots = \delta_{n-1} = 0 \) and \( \delta_n \neq 0 \). Next

\[
A^* = V^{-1} \Rightarrow B^* = P'^{-1} V^{-1} = P'^{-1} P' P = P, \quad Q^* = B'^{-1} P = I.
\]

Therefore again for \( i = 1, \ldots, n \)

\[
\gamma_i^* = \frac{\delta_n c_i^* q_n^*}{\sqrt{(p_i'Mp_n)}} = \frac{\delta_n c_i^* q_n^*}{\sqrt{(p_i'Mp_n)}} = \frac{\delta_n p_i'Mp_n}{\sqrt{(p_i'Mp_n)}}.
\]

To show (2) it suffices to show that \( | \gamma_i^* | \geq | \gamma_n | \), that is

\[
\sqrt{(p_i'Mp_n)} \geq | b_i'Mp_n | / \sqrt{(b_i'Mb_n)}
\]

which follows by the Cauchy–Schwarz inequality. Next, to show (3) it suffices to show that \( | \gamma_i^* | > | \gamma_i^* | \) for \( 1 \leq i \leq n-1 \), that is

\[
\sqrt{(p_i'Mp_n)} > | p_i'Mp_n | / \sqrt{(p_i'Mp_i)},
\]

which also follows by the Cauchy–Schwarz inequality; the strict inequality holds because \( P \) is nonsingular.

An obvious corollary is that if \( V \) is a diagonal matrix, then any \( d = A e \) gives an optimal test vector if \( A \) is diagonal.

[Received July 1981. Revised September 1981]