

Biometrika Trust

A Note on the Use of Residuals for Detecting an Outlier in Linear Regression

Author(s): Ajit C. Tamhane

Source: *Biometrika*, Vol. 69, No. 2 (Aug., 1982), pp. 488-489

Published by: [Biometrika Trust](#)

Stable URL: <http://www.jstor.org/stable/2335429>

Accessed: 22/10/2010 10:16

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=bio>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Biometrika Trust is collaborating with JSTOR to digitize, preserve and extend access to *Biometrika*.

<http://www.jstor.org>

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 ε 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 $r \leq n$, β is an $r \times 1$ vector of unknown parameters and ε 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 σ^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 σ^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

$$z_i = d_i/\sqrt{\text{var}(d_i)} \quad (i = 1, \dots, n). \quad (1)$$

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 $E(\varepsilon_i) \neq 0$ for some i ($i = 1, \dots, n$), 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 $\delta_1 = \dots = \delta_{n-1} = 0$. 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 $k > 0$,

$$\text{pr}(|z_n^*| > k) \geq \text{pr}(|z_i| > k) \tag{2}$$

for all z ,

$$\text{pr}(|z_n^*| > k) > \text{pr}(|z_i^*| > k) \quad (i = 1, \dots, n-1), \tag{3}$$

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 $E(d) = B'MP\delta$, and $\text{cov}(d) = B'MB = C$, say. Also write $\gamma_i = E(z_i) = (B'MP\delta)_i / \sqrt{c_{ii}}$, 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 $i = 1, \dots, n$, 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 $|\gamma_i|$ ($i = 1, \dots, n$) is true, e.g. under the normality assumption for ε .

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 $i = 1, \dots, n$

$$\gamma_i = \frac{(CQ\delta)_i}{\sqrt{(b'_i M b_i)}} = \frac{\delta_n c'_i q_n}{\sqrt{(b'_i M b_i)}} = \frac{\delta_n b'_i M p_n}{\sqrt{(b'_i M b_i)}}$$

when $\delta_1 = \dots = \delta_{n-1} = 0$ 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 $i = 1, \dots, n$

$$\gamma_i^* = \frac{\delta_n c'^*_i q_n^*}{\sqrt{(p'_i M p_i)}} = \frac{\delta_n c^*_{in}}{\sqrt{(p'_i M p_i)}} = \frac{\delta_n p'_i M p_n}{\sqrt{(p'_i M p_i)}}.$$

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

$$\sqrt{(p'_n M p_n)} \geq |b'_n M p_n| / \sqrt{(b'_n M b_n)}$$

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

$$\sqrt{(p'_n M p_n)} > |p'_i M p_n| / \sqrt{(p'_i M p_i)},$$

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 $d = Ae$ gives an optimal test vector if A is diagonal.