*Geophysical Prospecting* **39**, 915–942, 1991

# MAKING AVO SECTIONS MORE ROBUST[1]

A. T. WALDEN[2]


## Abstract

WALDEN, A.T. 1991. Making AVO sections more robust. *Geophysical Prospecting* **39**, 915–942.

When large quantities of seismic data are involved it is impossible to examine all gathers by eye for AVO anomalies. The standard approach is to compute, for each amplitude profile (at a specific time) on each gather, the intercept and gradient of a straight-line fit to seismic amplitudes. These intercepts and gradients are each plotted as a sort of seismic section – an intercept section, and a gradient section.

Estimation of the intercept and gradient for a straight-line fit to each amplitude profile proceeds traditionally via least-squares. Two undesirable features can be hidden from the user by the fitting procedure, namely (i) the effect of outlying or uncharacteristic amplitudes on the intercept and gradient estimates, and (ii) complete breakdown of the straight-line model for the amplitudes, thus rendering meaningless the intercept and gradient estimates. It should be remembered that least-squares can always fit any sequence of numbers to any other sequence of numbers; checks are needed to show that the result is meaningful.

It is shown that statistically robust estimation methods can greatly limit the damage done by outlying amplitudes, and that a simple test on the model, the runs-statistic, is capable of detecting breakdown of the straight-line assumption. It is demonstrated using two seismic data sets that these two techniques, used in tandem, facilitate much better quality control of AVO intercept and gradient calculations.


## Introduction

Amplitude-versus-offset (AVO) analysis is a technique, now becoming well known and often routine, for interpreting the variation of reflection amplitudes with angle of incidence in terms of changes in lithology and/or fluid fill.

Shuey (1985) showed that for angles of incidence $\theta$ less than about 25°, the P-wave reflection amplitude from a planar interface between two elastic media can be approximated by

$$R(\theta) = R_0 + B \sin^2 \theta,$$

[1] Paper read at the 52nd EAEG meeting, Copenhagen, May–June 1990; received October 1990, revision accepted May 1991.

[2] Formerly: BP Exploration Co. Ltd, Britannic House, Moor Lane, London EC2Y 9BU, U.K. Now at: Department of Mathematics, Imperial College of Science, Technology and Medicine, Huxley Building, 180 Queens Gate, London SW7 2BZ, U.K.

where

$$R_0 = \left[\left(\frac{\Delta\rho}{\bar{\rho}} + \frac{\Delta V_p}{\bar{V}_p}\right)/2\right],$$

and

$$B = -\left[4\left(\frac{\bar{V}_s}{\bar{V}_p}\right)^2\left(\frac{\Delta V_s}{\bar{V}_s}\right) + 2\left(\frac{\bar{V}_s}{\bar{V}_p}\right)^2\left(\frac{\Delta\rho}{\bar{\rho}}\right) - \frac{1}{2}\left(\frac{\Delta V_p}{\bar{V}_p}\right)\right]. \quad (1)$$

Here $R_0$ is the normal incidence or zero-offset reflection coefficient, $\bar{V}_p$, $\bar{V}_s$ and $\bar{\rho}$ are the average P-wave velocity, S-wave velocity and bulk density of the media on either side of the interface, and $\Delta V_p$, $\Delta V_s$ and $\Delta\rho$ are the differences, assumed small, of the quantities across the boundary.

Estimates of $R_0$ from CMP gathers should prove superior to the ordinary stack in defining the true zero-offset reflection coefficients. Estimates of $B$ should enable AVO effects to be pinpointed within the reams of CMP gathers, and may be combined with other quantities for specific diagnostic purposes.

How can we compute $\sin^2\theta$? From Snell's law for a layered medium, the ratio of the sine of the incident angle to the upper velocity at an interface is constant along a raypath, having ray parameter $p$. The slope of the time–distance curve is given by $p$ (e.g. Yilmaz 1987, p. 430), and hence for a particular interface of interest,

$$\frac{dt(x)}{dx} = p = \frac{\sin\theta}{V_p}, \quad (2)$$

where $V_p$ is the interval velocity of the medium above the interface, and $x$ is the source–receiver offset. Assuming a layered medium, and an offset small compared to depth, the NMO equation

$$t^2(x) = t^2(0) + x^2/V_{\mathrm{rms}}^2$$

holds, where $V_{\mathrm{rms}}$ is the root-mean-square (rms) velocity and $t_0$ is the time at zero offset. Differentiating gives

$$\frac{dt(x)}{dx} = \frac{x}{t(x)V_{rms}^2} = \frac{\sin\theta}{V_p},$$

and hence

$$\sin^2\theta = \frac{x^2 V_p^2}{V_{\mathrm{rms}}^2[V_{\mathrm{rms}}^2 t^2(0) + x^2]}.$$

Some approximations exist which can simplify this expression for computational purposes.

Throughout this paper the term 'offset' will often be used as a substitute for the estimated $\sin^2\theta$, but in some places the formal difference between the two will be emphasized.

An accepted procedure for analysing AVO over large data sets is (a) to draw attention to places where there is a large AVO effect by plotting the gradients of linear regressions of amplitude against $\sin^2 \theta$ at each time level and for each gather, and (b) in regions pinpointed as interesting, to look at the gathers to try to assess the nature of the AVO effect. Sometimes the perceived effect is due to residual NMO (RNMO) or multiples. In addition the intercepts of the linear regressions are usually plotted as an idealized zero-offset section (Denham, Palmeira and Farrell 1985), which may then be matched to synthetics, or migrated.

This method of analysis might realistically be described as simple and dirty, since AVO effects might be better analysed by looking at time gates on gathers, rather than amplitude profiles. The former would use a tracker to follow an event and hence be less sensitive to RNMO and other forms of noise. Even the simple amplitude profile technique is heavy on computer time (typical lines using up to 6 h AP time for AVO processing) and if this was to be combined with intelligent tracking of events, computer time for such an intensive prestack method could prove prohibitive.

Two main problems arise in practice with the simple amplitude profile technique. Firstly, a few large outlying values might occur in the prestack amplitudes. These can be accommodated by using a *robust estimation* method. Secondly, due to the cumulative effects of many processing steps (or possibly due to physical complexity) the AVO results do not satisfy a straight-line model. For such time slices it is pointless fitting a straight line; in such cases we want to flag the amplitude profile as erroneous (with respect to the model) by use of a *test statistic*.

We shall look at these two areas for improving quality control of AVO intercept and gradient estimates.

## A Robust Estimation Approach

### *Introduction*

For a time $t$ and incident angles $\theta_{1,t}, \ldots, \theta_{N,t}$ for $N$ different offsets in a processed-for-AVO gather, we denote the straight-line relationship at time $t$ over the $N$ values of $\sin^2 \theta_{i,t}$ ('offsets') by

$$Y_{i;t} = \alpha_t + \beta_t Z_{i;t} + \varepsilon_{i;t}, \quad i = 1 \ldots N, \tag{3}$$

where $\{Y_{i;t}\}$ are the observed amplitudes, $\alpha_t$ is the intercept amplitude, $\beta_t$ the gradient, $\{Z_{i;t}\}$ the values of $\sin^2 \theta_{i,t}$ and $\{\varepsilon_{i;t}\}$ the random errors. Least-squares will be optimal for estimating the intercept and gradient when the error distribution is Gaussian. Let $a_t$ and $b_t$ be the least-squares estimates of $\alpha_t$ and $\beta_t$. Then a standard relationship in linear regression is

$$Na_t = \sum_{i=1}^{N} Y_{i;t} - b_t \sum_{i=1}^{N} Z_{i;t}.$$

Here $\sum_{i=1}^{N} Y_{i;t}$ is the result of stacking across values of $\sin^2 \theta$, i.e. across offsets, and $\sum_{i=1}^{N} Z_{i;t}$, the sum of the $\sin^2 \theta$'s, will always be positive. $a_t$ is the intercept amplitude. Hence, whenever $b_t < 0$,

$$a_t > \sum Y_{i;t}/N, \tag{4}$$

i.e. the value of the intercept section for the amplitude profile will exceed the conventional mean stack value (when gains are equal). Similarly, whenever $b_t > 0$,

$$a_t < \sum Y_{i;\,t}/N.$$

This situation is illustrated in Fig. 1. The top plot (a) shows a random sample of 100 points from the Gaussian distribution. The fitted regression line is only just non-horizontal (due to random effects), and its intercept differs very little from the mean of the points (arrow). If we substitute the first ten points by a set of larger *positive*

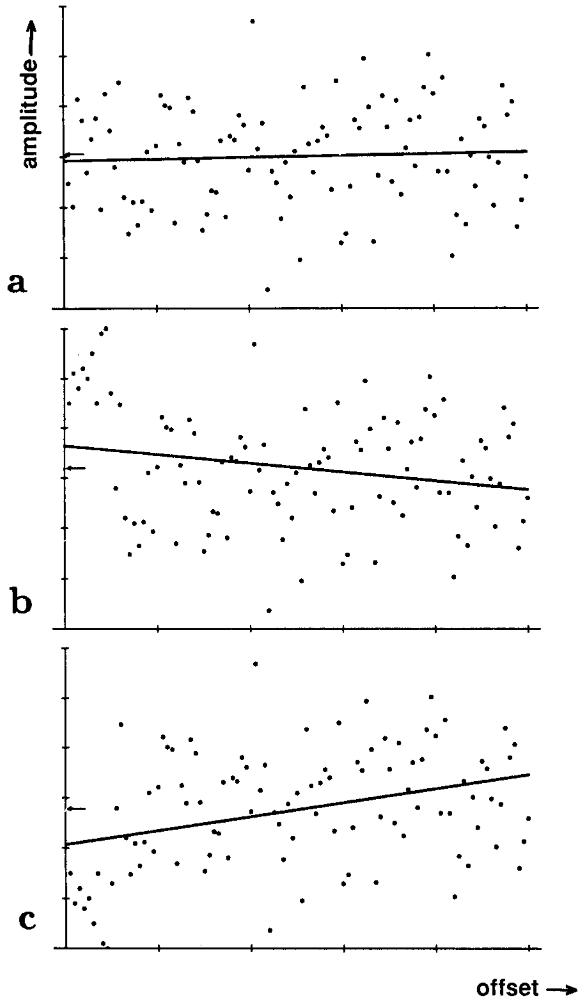



FIG. 1. Effect of outliers. In (a), with no outliers, the gradient is practically zero, while in (b) too-large values at the near offsets generate a spurious negative gradient, and in (c) too-small values generate a spurious positive gradient. The arrows show the mean of the points, which may be contrasted with the intercept of the straight line.

values (outliers) as in (b), the intercept of the new regression line is now considerably different from the mean of the points (arrow). (c) illustrates the same situation for large *negative* outliers replacing the first ten points. If the true amplitude of the event at the position of the multiple is constant with offset, then the remnant multiple at near offsets is likely to generate a non-zero gradient, as shown in Fig. 1. For the example in Fig. 2, the data has had *f–k* demultiple applied, which has little effect on near offsets. The multiple increases the amplitude of the near trace and so the

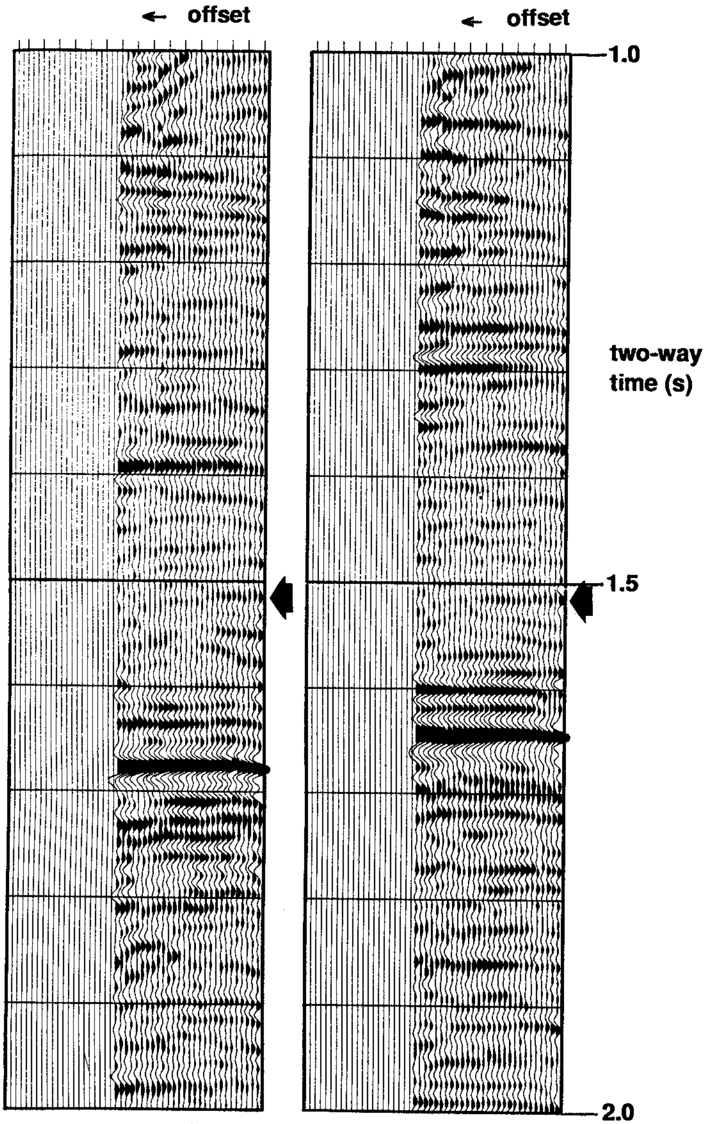



FIG. 2. Gathers showing the multiple event just below 1.5 s on the near offsets

gradient $b_t$ will be negative (Fig. 1b), and (4) will result, i.e. the intercept section will have an amplitude at $t$ more influenced by the multiple than would the conventional stack.

We have illustrated that least-squares estimation can give undesirable results in the presence of a few contaminated data points. In statistical terminology least-squares regression is said to be non-robust to contamination by a small percentage of outliers in the sample. Over the last 15 years or so, techniques have been developed to make regression more robust to contamination, by replacing least-squares by more sophisticated but generally more time-consuming approaches. Robust statistics is not popular with many statisticians who correctly point out that automated robust procedures rarely yield results superior to those from a careful study of a data set by a skilled analyst who can perhaps identify outliers by more subtle means. The arguments largely mirror those in geophysics between almost automated interpretation and the skill of an experienced interpreter. For AVO analysis of large quantities of data, the view expressed by Hocking (1983) is relevant:

"The concept of automated protection is reassuring to some analysts, especially those without the luxury of time required for a thorough diagnostic study."

Hence, we describe an approach to making amplitude profile AVO analysis less susceptible to error contamination. Firstly, a brief review of the current least-squares approach is useful.

## *AVO: The least-squares approach*

The intercept and gradient at time $t$, namely $\alpha_t$ and $\beta_t$, are obtained by minimizing

$$\sum_{i=1}^{N} w_i(Y_{i;t} - \alpha_t - \beta_t Z_{i;t})^2, \tag{5}$$

where $w_i$ is a weight for channel (offset) $i$. In practice the $\{w_i\}$ are usually set at unity for $i = 1, \ldots, N$, but might be varied in order to reduce the effect of offset-dependent noise. We shall retain the $\{w_i\}$ for latter use. In mute zones $N$ will of course only be a part of the full number of channels. Note that the $\{w_i, i = 1, \ldots, N\}$ are the same for each time.

The least-squares parameter estimates take the form of simple sums across offsets (see Appendix A) and are tailor-made for array processor implementation (Fig. 3). This is the fastest way to do the calculations, i.e. load as many offsets of a gather as possible into the array processor, accumulate the various sums, load the next set of offsets and so on, until the gather has been processed. The intercept and gradient are then computed from the accumulated sums.

The standard deviations of the parameter estimates are estimated (Appendix A), and may be used to control the output of various functions of $a_t$ and $b_t$ which might be of interest. For example, for the intercept section, the following is output:

$$\begin{cases} a_t, & \text{if } |a_t| \geq f_1\, \hat{\sigma}(a_t), \\ 0, & \text{if } |a_t| < f_1\, \hat{\sigma}(a_t), \end{cases}$$

Weights

$[w_1, w_2, \ldots, w_N]$

$$
\begin{array}{cccccc}
 & \text{Gather 'offsets'} & \text{Gather amplitudes} & \text{Accumulators} & & \text{Intercepts \& Gradients} \\
t \rightarrow & \begin{bmatrix} \\ Z_{1;t}, Z_{2;t}, \ldots, Z_{N;t} \\ \\ \end{bmatrix} & \begin{bmatrix} \\ Y_{1;t}, Y_{2;t}, \ldots, Y_{N;t} \\ \\ \end{bmatrix} \rightarrow & \begin{bmatrix} \\ \sum w_i Z_{i;t}^2 \\ \\ \end{bmatrix} \begin{bmatrix} \\ \sum w_i Y_{i;t} \\ \\ \end{bmatrix} & \xrightarrow{\text{etc.}} & \begin{bmatrix} \\ a_t, b_t \\ \\ \end{bmatrix}
\end{array}
$$

FIG. 3. Schematic showing how the least-squares estimates of intercept and slope for amplitude profiles at different times are tailor-made for array processor computation.

and for the gradient section, the gradient times sign of the (weighted) stack is output as a dual polarity display:

$$
\begin{cases} b_t \operatorname{sgn}\left(\sum w_i Y_{i;t}\right), & \text{if } |b_t| \geq f_2\, \hat{\sigma}(b_t), \\ 0, & \text{if } |b_t| < f_2\, \hat{\sigma}(b_t), \end{cases} \tag{6}
$$

where $f_1$ and $f_2$ are factors controllable by the processor. For example, non-zero values might only be output when the absolute value of the parameter estimate exceeds say 4 of its estimated standard deviations ($f_1, f_2 = 4$). An example interpretation of the gradient times the sign of the (weighted) stack is as follows: if the sign of the stack is negative, and the gradient positive, the overall result is negative, indicating a decreasing (in the absolute sense) amplitude with offset; if the sign of the stack is positive, and the gradient positive, the overall result is positive, indicating an increasing (in the absolute sense) amplitude with offset.

The larger the parameter estimate (intercept/gradient) as a proportion of its standard deviation, the more confident we are that the parameter estimate is non-zero. Hence increasing the factor merely results in those parameter estimates in which we have more confidence in their being non-zero being retained in the output. The idea is based on the ideal linear model with no outliers and no departure from the straight-line assumption. The use of the factor does not mean that we are properly taking care of the effect of outlying values or serious departures from the assumed straight-line model.

## *AVO: The robust estimation approach*

A well-established method for making regression analysis robust consists of two steps (see e.g. Andrews 1974; Hill and Holland 1977): (a) calculate an initial estimate of the regression parameters using a robust but perhaps rather crude method; (b) use the residuals from (a) as input to an M-estimate ('maximum-likelihood') regression analysis.

It is essential that a preliminary estimate of regression parameters obtained from step (a) is not found using least-squares since the residuals from outlying points will not be large enough, and so spoil step (b).

A suitable preliminary estimate is given by regression by medians. Computational details are discussed in Appendix B. This gives a series of preliminary residuals from the estimated fit. These are then used to generate weights for step (b), as described in Appendix C, and final estimates of $\alpha_t$ and $\beta_t$ calculated as the solution to the weighted least-squares problem:

$$\min \sum_{i=1}^{N} w_{i;\,t}(Y_{i;\,t} - \alpha_t - \beta_t Z_{i;\,t})^2. \tag{7}$$

Note that here the weights will be amplitude-profile dependent (varying at different levels in a gather). The robust approach does not readily lend itself to full array processor implementation; it was found to double the CPU time approximately, while halving the AP time.

It is important to realize that the term *robust* here means the ability to cope with departures from the assumed Gaussian random errors in (3). This is quite different from estimation *stability* which improves with more offsets if the statistical assumptions are not violated.

## A Test for Breakdown of the Straight-Line Model

### *Introduction*

The appearance of a few large outliers in an amplitude profile is equivalent to modification of the distribution of the errors in the model (3) away from Gaussian, probably towards a long-tailed distribution. As will be demonstrated by data examples the problem can be adequately dealt with using the robust approach developed above. Far more worrying is where the model of (3) no longer holds in the sense that the amplitudes are not linearly related to $\sin^2 \theta$. This breakdown of the model renders estimating an intercept and/or gradient meaningless.

### *The runs-statistic Z*

A frequent characteristic of cases where a straight-line fit is not appropriate is that the resulting residuals (differences between the fitted line and the data points) have long runs of the same sign, positive or negative. Figure 4a shows a V-shaped scatter of points plus the line fitted by least-squares, while Fig. 4b shows a sinusoidal scatter of points plus the fitted line. By looking for such runs it should therefore be possibe to pin-point the times and CMP positions where the amplitude profiles do not satisfy the straight-line assumption. Appendix D gives some details of the construction of the runs-test statistic $Z$ which has Gaussian (normal) distribution with mean zero and variance unity. Large positive and negative values of $Z$ indicate departure from the assumption of random errors about the straight line (a

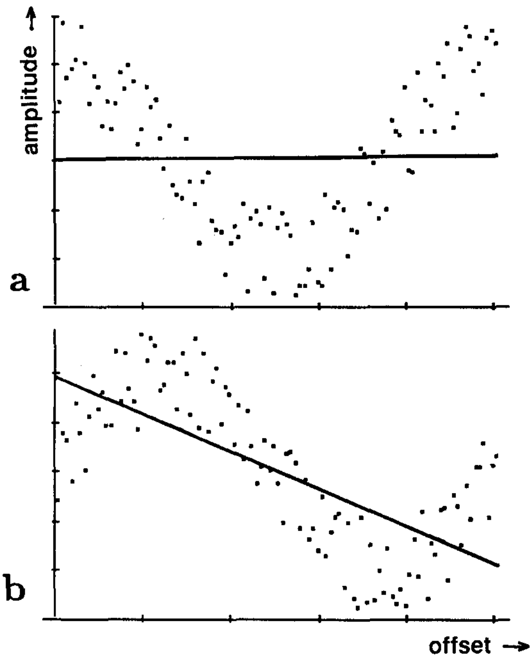



FIG. 4. Examples of data sets for which a straight-line fit is clearly inappropriate. (a) A V-shaped scatter of points, (b) a sinusoidal scatter. The heavy lines are the least-squares straight-line fits.

breakdown of the model). Large negative values of $Z$ indicate too few runs of positive or negative residuals, i.e. positive serial correlation. This is indeed what is found in practice, and is illustrated in Figs 4a and b.

## DATA EXAMPLES

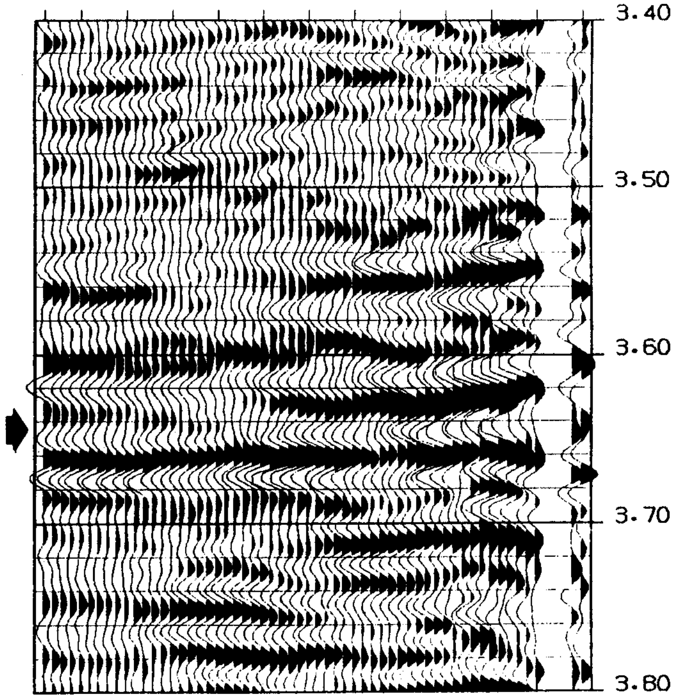
### *Introduction*

The data examined in these two examples were chosen primarily to illustrate important points regarding estimation of AVO intercept and gradient sections. Further processing of the data is proposed, so that what is shown here is not necessarily the best or final product.

### *Data Set 1: Least-squares versus robust estimation*

The first set of seismic data processed with both least-squares and robust methods was from a structurally extremely complex area. Some 940 CMP gathers were processed using both techniques. The time interval of interest was 3.4–5.4 s.

Differences of the gradient sections were computed, and areas of large differences pin-pointed for further investigation. One such area occurred in the CMP range 760–780, and for the time interval 3.4–3.8 s. Examples are given in Figs 5a and b for



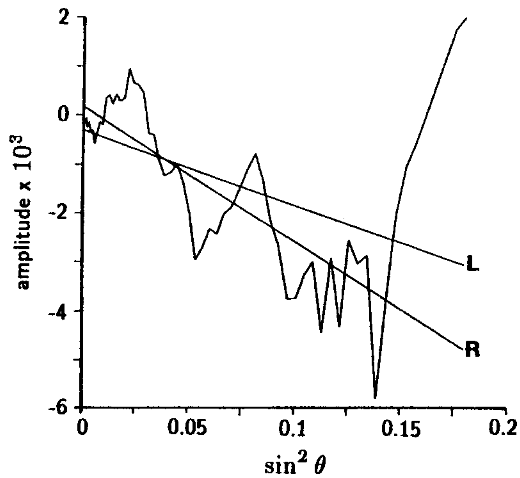




FIG. 5a.

FIG. 5. Comparison of AVO intercept and slopes estimated by least-squares and by the robust technique on seismic data set 1. The upper panel shows a section of the gather and the lower panel shows rescaled amplitudes at $t = 3.644$ s plotted against $\sin^2\theta$; the marked lines are fitted by least-squares (L) and robustly (R). The CMPs are (a) 770 and (b) 772.

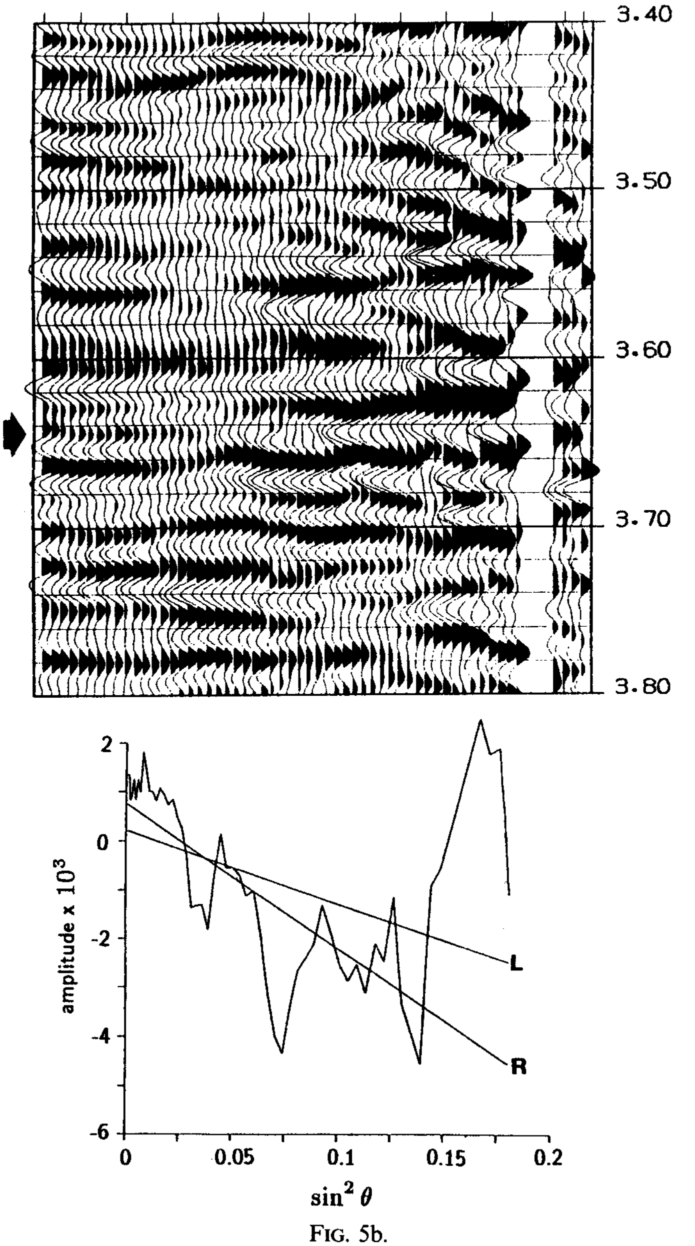



FIG. 5b.

CMPs 770 and 772 respectively. In both figures, the CMP gather is displayed together with the amplitudes versus $\sin^2 \theta$, and the fitted least-squares and robust regressions. Transforming offset to $\sin^2 \theta$ means that the near-offset points become closer together, and the far-offset points will exert considerable leverage in the straight-line fit. As can be seen, the far-offset amplitudes are precisely those that are most erratic and clearly in error. The decrease in amplitude with offset, which

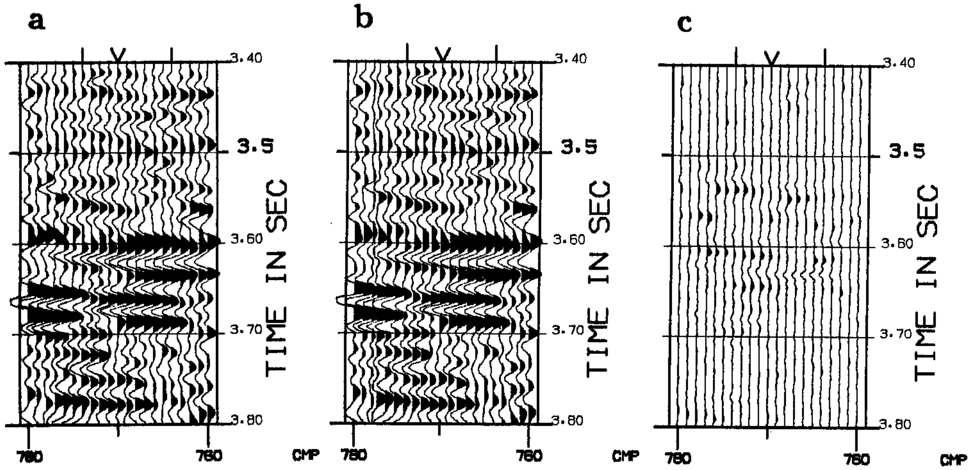



FIG. 6. A detailed view of the intercept section for CMPs 760–780 and time interval 3.4–3.8 s for data set 1: (a) computed robustly, (b) by least-squares, and (c) the difference on the same scale as (a) and (b).

occurs out to about 2200 m offset, is detected much more accurately by the robust method than by least-squares. Examination of the reason for such an AVO effect might negate any physical interpretation, but the robust method clearly identified the effect, and drew our attention to it.

Figure 6c shows the difference in the corresponding estimated intercept sections between that estimated robustly, Fig. 6a, and by least-squares, Fig. 6b. The scale on Fig. 6c is the same as that of Figs 6a and b. The factor $f_1$ was set at zero. The differences are not large (as would be expected from Fig. 5 where the intercept is seen to be relatively stable) but if the intercept section was used in matching to a synthetic, or for migration, such differences could become cumulatively important.

The difference in the estimated gradient sections between the robust (Fig. 7a) and the least-squares estimate (Fig. 7b) is given in Fig. 7c (on the same scale). The factor $f_2$ was set at 6 to help clarify the figure. The differences here are substantial, reflecting the considerable dependence of the gradient estimates on the technique used (Fig. 5).

### *Data Set 2: Serious breakdown of the straight-line model*

The second data set examined, from the North Sea, was of particular interest due to a strong AVO anomaly. The gradient section for a panel of interest is shown in Fig. 8a (using the robust method) and Fig. 8b (using least-squares). The difference (on the same scale) is shown in Fig. 8c. The factor $f_2$ was set at zero for these plots. A strong AVO anomaly is apparent between 1.45–1.5 s and CMPs 1660–1690 in Figs 8a and b. There is little difference between the robust and least-squares estimates.

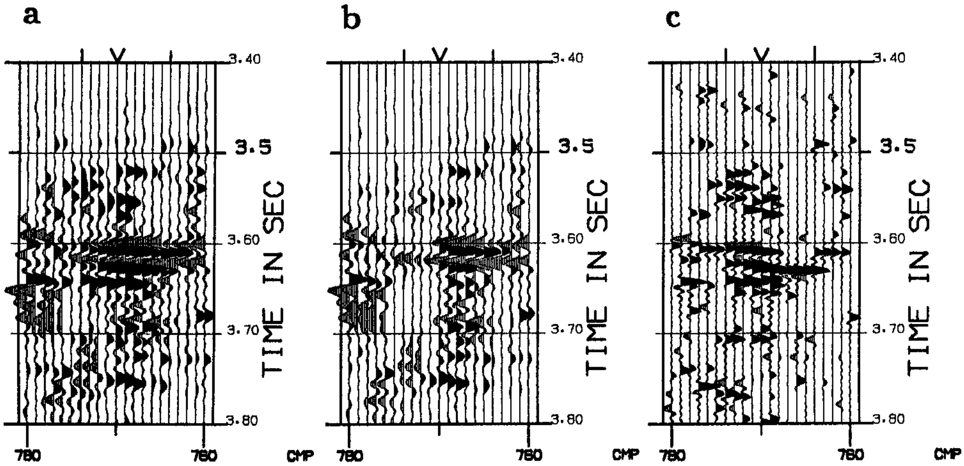



FIG. 7. A detailed view of the gradient section for CMPs 760–780 and time interval 3.4–3.8 s for data set 1: (a) computed robustly, (b) by least-squares, and (c) the difference on the same scale as (a) and (b).

The conclusion could be drawn that because there is little difference between the least-squares and robust results, the data quality must be 'high'. Such a conclusion is, in fact, erroneous. It can quite readily happen with real data that the profile amplitudes cannot, by any stretch of the imagination, be modelled by a straight line plus random noise. Figure 9 shows part of the gather from CMP 1705 of Fig. 8. The amplitude profiles at times 1.488 and 1.584 s are shown above the gather; the first takes the form of a shallow V-shape, while the second looks like a valley with mountains on either side. For these amplitude profiles, both regression methods find only a slight gradient (Figs 10a and b), and the fitted lines are quite meaningless.

A notable characteristic of the fitted lines is that the resulting residuals have long runs of the same sign. The runs-statistic is clearly appropriate here. The fit in Fig. 10a gives $Z = -4.8$, while Fig. 10b gives $Z = -4.4$. Figure 11a shows the gradient section ($f_2 = 0$) from robust estimation, while Fig. 11b shows the same, but values are plotted only if the corresponding $|Z| < 3$. The difference is shown in Fig. 11c, and is in fact a highlight of positions where the model is very poor (since the choice of 3 as a cut-off corresponds to 3 standard deviations from the mean). In order to illustrate that small values of $Z$ correspond to reasonably fitting models, Figs 12a and b give fits corresponding to $Z = 0.04$ and $Z = 0.9$ respectively.

## CONCLUSIONS

The current standard method of estimating AVO intercept and gradient sections leaves a lot to be desired in terms of quality control. Two main problems can arise. A few large outliers can occur in a set of amplitude profiles, leading to serious distortion of the least-squares straight-line estimate. Robust regression techniques

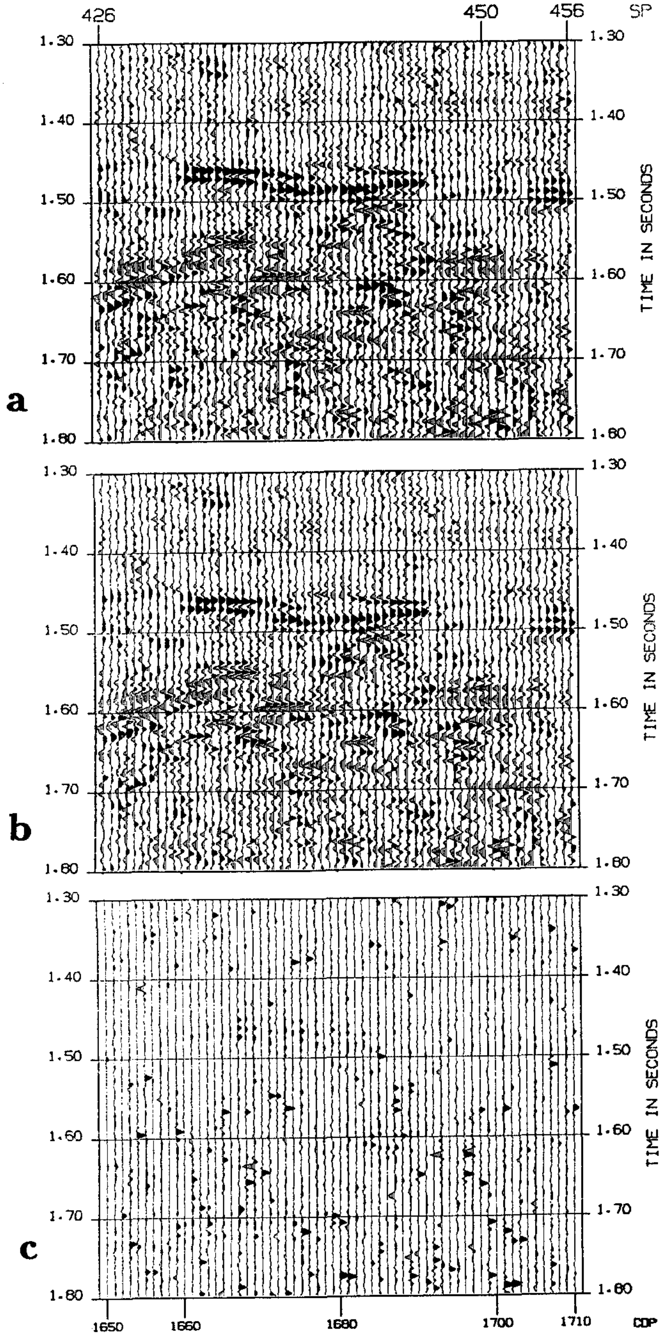



FIG. 8. Gradient section for panel of interest for data set 2: (a) computed robustly, (b) by least-squares, and (c) the difference on the same scale as (a) and (b).

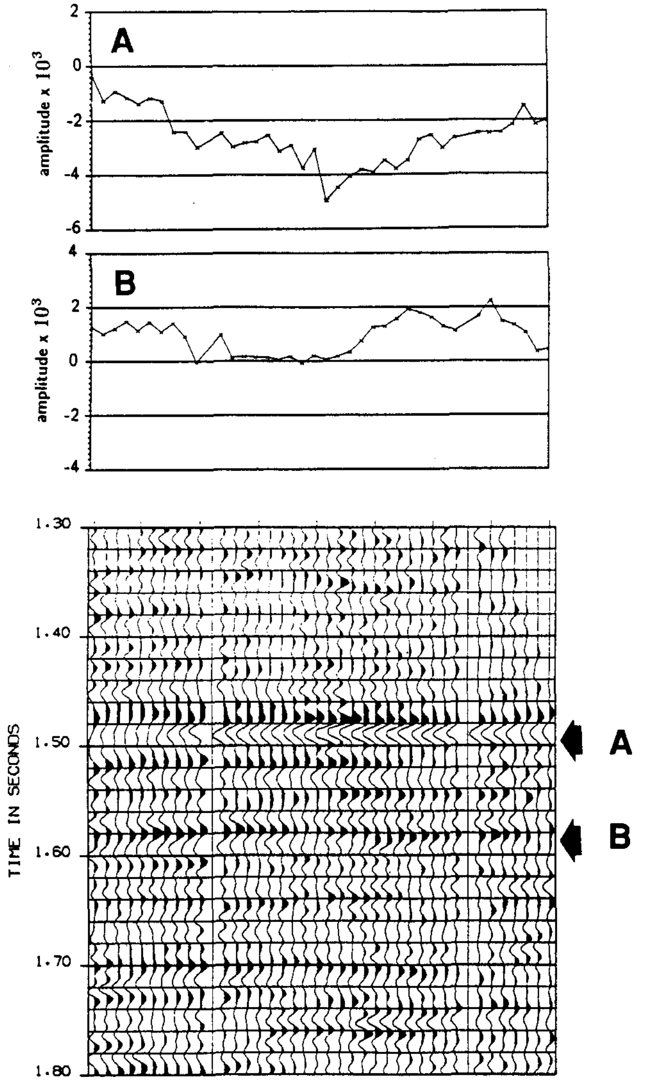



FIG. 9. The bottom panel shows a section of the gather for CMP 1705 (data set 2) and the upper panels show rescaled amplitudes at $t$ = 1.488 and 1.584 s plotted against offset.

can take care of such problems. Secondly, complete breakdown of the straight-line model can occur. Under such circumstances, the estimated intercept and gradient are meaningless. This eventuality can be successfully 'trapped' in many cases by use of the runs-statistic. Breakdown of the straight-line model may be of diagnostic importance if not due solely to processing problems, and hence highlighting where it occurs is also useful for that reason.

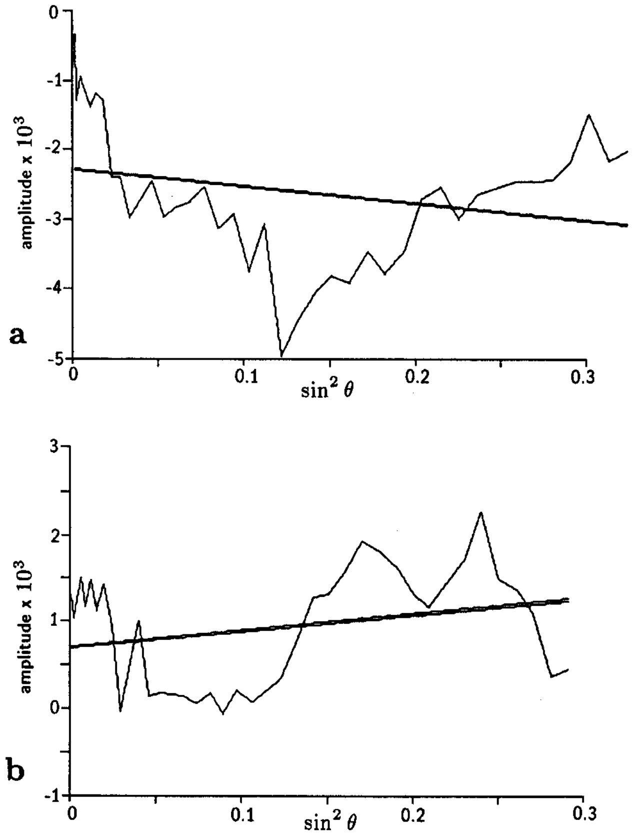



FIG. 10. AVO intercept and slopes estimated by least-squares and by the robust technique on seismic data set 2, CMP 1705 at (a) $t = 1.488$ s, for which $Z = -4.8$, and (b) $t = 1.584$ s for which $Z = -4.4$. The two methods give such similar fitted lines that they have not been distinguished. Note the straight lines are essentially unrelated to the variation of amplitude with offset.

These observations suggest the following strategy towards AVO processing of large amounts of data:

(i) Carry out a preliminary AVO processing run using the robust estimation method combined with output of results only if the runs-statistic $|Z| > 3$. This will produce a plot of areas where the model breaks down.

(ii) Look at the areas thus identified, make any corrections to the processing scheme found necessary or possible, and run the AVO module with robust estimation. (If breakdown of the model cannot be explained from processing considerations, seek other explanations). Robust estimation will protect against long-tailed distributions of errors, which can arise even when the model is correct.

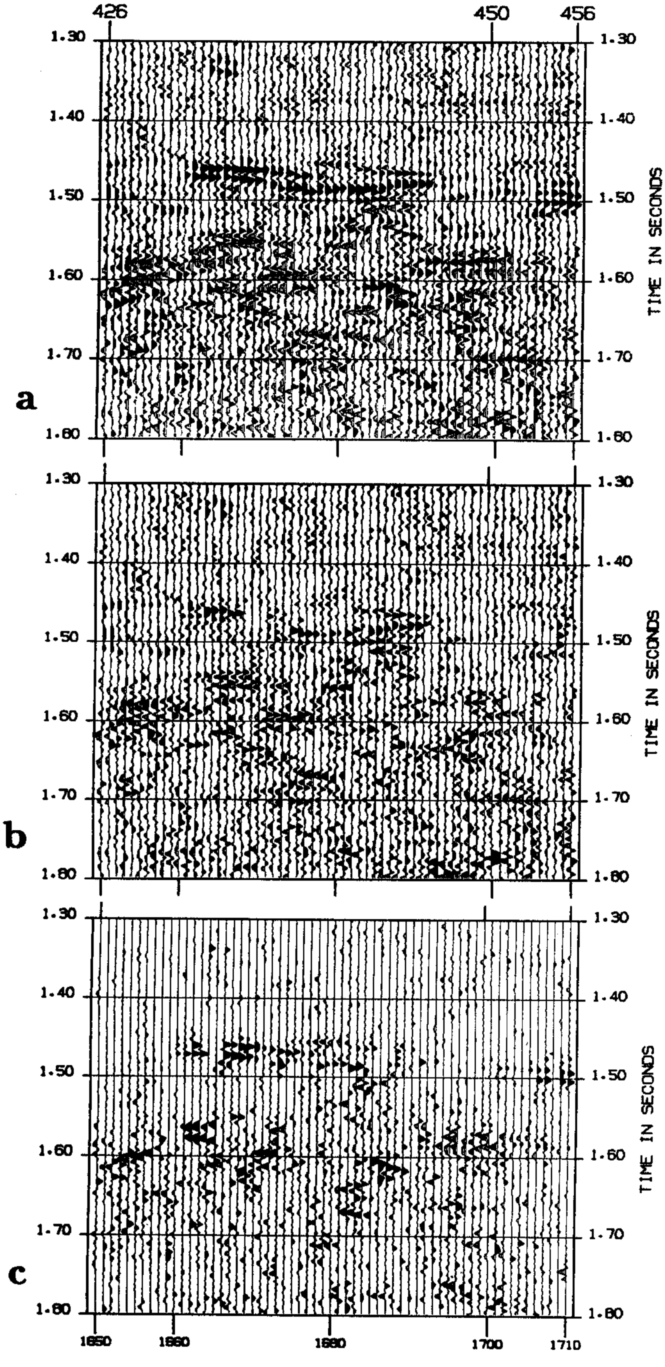



FIG. 11. Gradient section for panel of interest for data set 2: (a) computed robustly, (b) computed robustly and only plotted if the runs-statistic $Z < 3$, and (c) the difference on the same scale as (a) and (b).

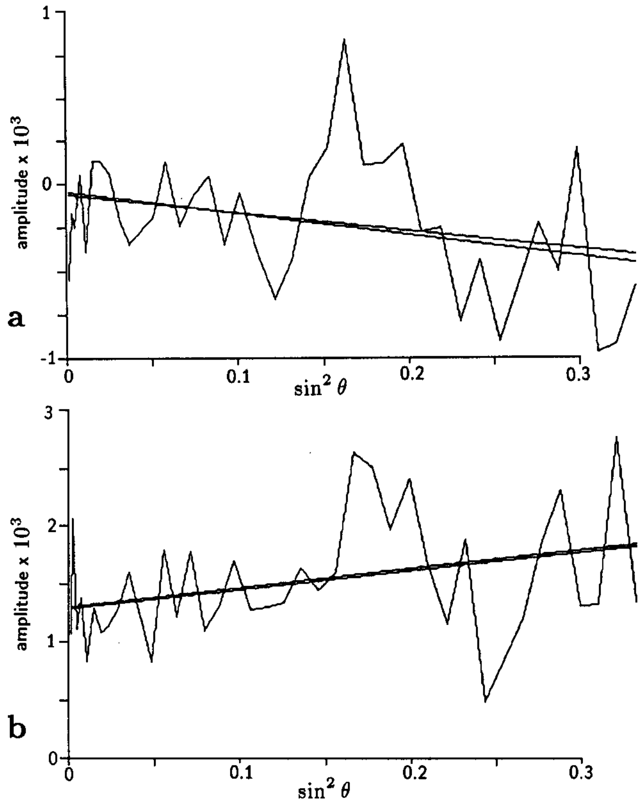



FIG. 12. AVO intercept and slopes estimated by least-squares and by the robust technique on seismic data set 2, CMP 1667 at (a) $t = 1.424$ s, for which $Z = 0.04$, and (b) $t = 1.456$ s for which $Z = 0.9$. The two methods again give such similar fitted lines that they have not been distinguished, but this time the straight-line model is perfectly reasonable.


## ACKNOWLEDGEMENT

The author thanks the Chairman and Board of Directors of The British Petroleum Company plc for permission to publish the paper.


## APPENDIX A
## THE LEAST-SQUARES EQUATIONS

The least-squares estimates $a_t$ of $\alpha_t$ and $b_t$ of $\beta_t$ from the minimization of (5), with weights $w_i$, are given by

$$a_t = \left[\left(\sum w_i Z_{i;t}\right)\left(\sum w_i Z_{i;t} Y_{i;t}\right) - \left(\sum w_i Z_{i;t}^2\right)\left(\sum w_i Y_{i;t}\right)\right] \Big/ D,$$

$$b_t = \left[\left(\sum w_i Z_{i;t}\right)\left(\sum w_i Y_{i;t}\right) - \left(\sum w_i\right)\left(\sum w_i Z_{i;t} Y_{i;t}\right)\right] \Big/ D,$$

and

$$D = \left[\left(\sum w_i Z_{i;t}\right)^2 - \left(\sum w_i\right)\left(\sum w_i Z_{i;t}^2\right)\right].$$

The usual weighted least-squares equations give

$$\hat{\sigma}^2(a_t) = -[1/(N-2)]\left(\sum w_i r_{i;t}^2\right)\left(\sum w_i Z_{i;t}^2\right)\Big/ D,$$

$$\hat{\sigma}^2(b_t) = -[1/(N-2)]\left(\sum w_i r_{i;t}^2\right)\left(\sum w_i\right)\Big/ D,$$

where $\hat{\sigma}^2(a_t)$ is the estimated variance of $a_t$, and likewise for $b_t$, and

$$r_{i;t} = Y_{i;t} - a_t - b_t Z_{i;t}$$

is the $i$th residual at time $t$.

## Appendix B
## Regression by Medians

Computing preliminary parameter estimates by medians was suggested by Mood (1950) and Andrews (1974). The data points $(Z_{i;t}, Y_{i;t})$ are divided into two groups $L_t = \{(Z_{i;t}, Y_{i;t}) : Z_{i;t} < \text{med } Z_{i;t}\}$ and $R_t = \{(Z_{i;t}, Y_{i;t} : Z_{i;t} > \text{med } Z_{i;t}\}$, i.e. points whose $Z$-coordinate is to the left of the median of the $Z_{i;t}$ and points whose $Z$-coordinate is to the right. If $N$ is odd the central point is not used. Figure 13 illustrates the groupings for even and odd $N$. Next, let

$$\tilde{Z}_{L;t} = \text{med } \{Z_{i;t} : Z_{i;t} \text{ in } L_t\},$$

$$\tilde{Z}_{R;t} = \text{med } \{Z_{i;t} : Z_{i;t} \text{ in } R_t\},$$

$$\tilde{Y}_{L;t} = \text{med } \{Y_{i;t} : Z_{i;t} \text{ in } L_t\},$$

$$\tilde{Y}_{R;t} = \text{med } \{Y_{i;t} : Z_{i;t} \text{ in } R_t\},$$

i.e. $\tilde{Z}_{L;t}$ is the median of the $Z$-coordinates in $L_t$, $\tilde{Y}_{L;t}$ is the median of the $Y$-coordinates whose $Z$-coordinate is in $L_t$, etc.

The gradient is then estimated by

$$\tilde{\beta}_t^{(1)} = (\tilde{Y}_{R;t} - \tilde{Y}_{L;t})/(\tilde{Z}_{R;t} - \tilde{Z}_{L;t})$$

as illustrated in Fig. 13. An iterative approach can be used to further refine the estimate. Define

$$U_{i;t} = Y_{i;t} - \tilde{\beta}_t^{(1)} Z_{i;t}$$
$$= \alpha_t + (\beta_t - \beta_t^{(1)})Z_{i;t} + \varepsilon_{i;t}.$$

The procedure is now repeated on the $(Z_{i;t}, U_{i;t})$ data and a slope estimate, $\tilde{\beta}_t^{(2)}$ say, obtained. The estimate of $\beta_t$ is now $\beta_t = \tilde{\beta}_t^{(1)} + \tilde{\beta}_t^{(2)}$. It is probably not worth further

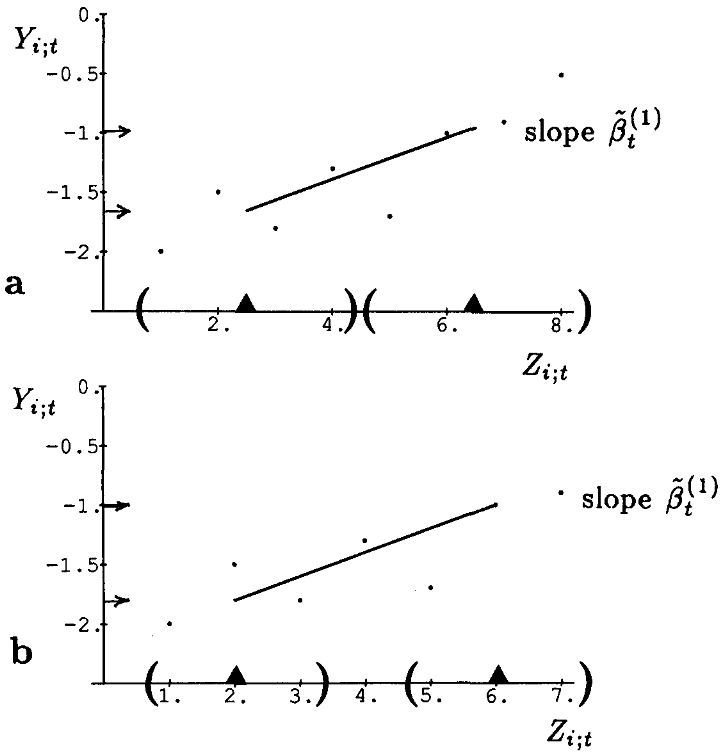



FIG. 13. Illustration of the method of gradient estimation by medians described in Appendix B for (a) an even number of points, i.e. 8, and (b) an odd number, i.e. 7. The round brackets show the $Z$-coordinates of the points in the left and right groupings, and the triangles mark the median of the $Z$-coordinates of the left and right groupings, i.e. $\tilde{Z}_{L;t}$ and $\tilde{Z}_{R;t}$ respectively. The slopes of the solid lines give $\tilde{\beta}_t^{(1)}$.

iteration, hence a series of residuals about $\alpha_t$ is defined:

$$\begin{aligned}\tilde{S}_{i;t} &= Y_{i;t} - \tilde{\beta}_t Z_{i;t}\\ &= \alpha_t + (\beta_t - \tilde{\beta}_t)Z_{i;t} + \varepsilon_{i;t}.\end{aligned}$$

Since $(\beta_t - \tilde{\beta}_t)$ should be small, and $\varepsilon_{i;t}$ are random, an estimate of $\alpha_t$ is given by

$$\tilde{\alpha}_t = \text{med } \tilde{S}_{i;t}.$$

These steps are shown in Fig. 14. For the example data set, $\tilde{\beta}_t^{(1)} = 0.207$ and $\tilde{\beta}_t^{(2)} = -0.0017$.

Several comments should be made on the above algorithm. Firstly, the breakdown point of the median is 50%, i.e. up to 50% of a sample can be erroneous values without causing breakdown of the median estimate of location. Since medians are calculated for both the left and right sides of the AVO sample, this means that up to 25% contamination can be tolerated in each half of the offset

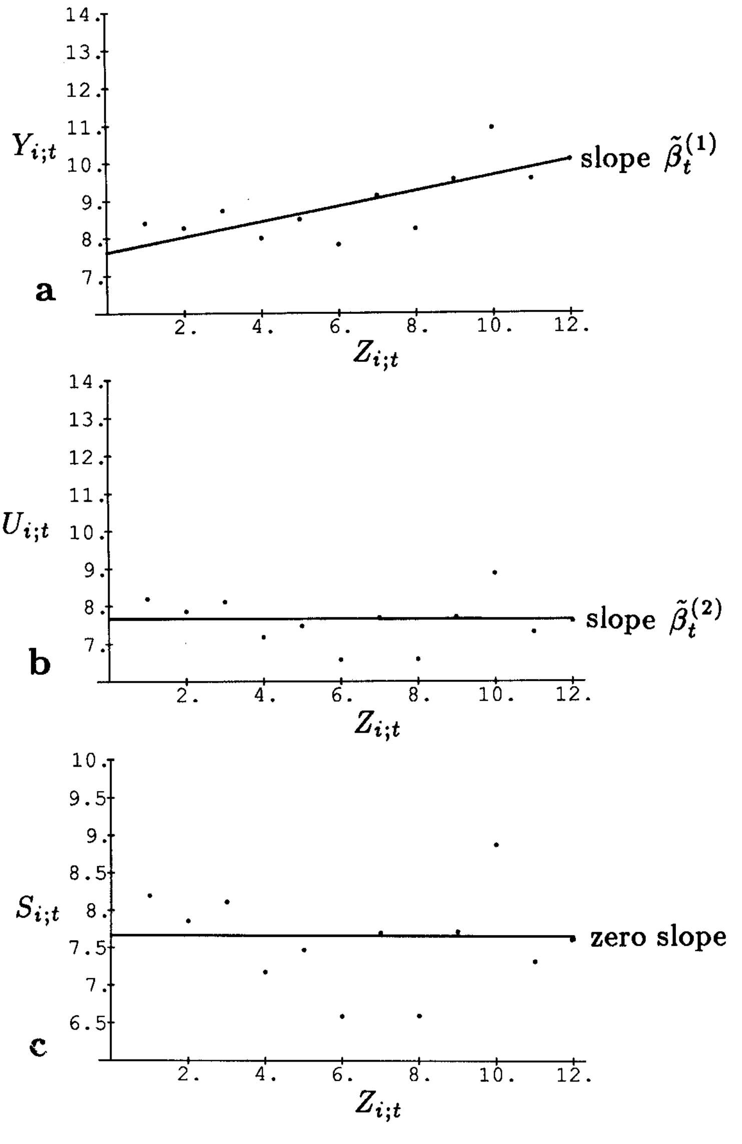



FIG. 14. Median regression using iterative refinement on 12 points. (a) A plot of some original $(Z_{i;t}, Y_{i;t})$ data; the solid line defines the slope $\tilde{\beta}_t^{(1)}$. (b) A plot of the $(Z_{i;t}, U_{i;t})$ data; the solid line defines the slope $\beta_t^{(2)}$. (c) A plot of the resulting $(Z_{i;t}, S_{i;t})$ data; the solid line has an almost exactly zero slope.

range. Secondly, it is not necessary to sort an entire array in order to find its median. The partial sorting algorithm due to Chambers (1971) proved about twice as fast as complete sorting using Quicksort (e.g. Singleton 1969) when only the median is required.

## Appendix C
## The M-estimation Step

Assuming we have obtained our preliminary estimate of $\alpha_t$, $\beta_t$, denoted $\tilde{\alpha}_t$, $\tilde{\beta}_t$, by medians as described in Appendix B above, we now carry out one step of refining procedure known as M-estimation (maximum-likelihood estimation). The classical solution to simple regression is least-squares without weights, i.e. we minimize

$$\sum_{i=1}^{N} (Y_{i;t} - \alpha_t - \beta_t Z_{i;t})^2. \tag{C1}$$

Differentiating gives a pair of equations to be solved:

$$\sum_{i=1}^{N} (Y_{i;t} - \alpha_t - \beta_t Z_{i;t}) = 0,$$

$$\sum_{i=1}^{N} (Y_{i;t} - \alpha_t - \beta_t Z_{i;t})Z_{i;t} = 0.$$

Denoting the $i$th residual by

$$r_{i;t} = Y_{i;t} - \alpha_t - \beta_t Z_{i;t},$$

then simple least-squares can be seen as solving

$$\sum r_{i;t} = 0,$$

$$\sum r_{i;t} Z_{i;t} = 0.$$

Note that $\{r_{i;t}\}$ can be rescaled without changing the equations. The classical solution to (C1) can be written instead as the minimization of

$$\sum_{i=1}^{N} \rho(r_{i;t}),$$

where $\rho(u) = u^2/2$ (The factor of 2 makes no difference to the minimization). Now, taking derivatives we obtain

$$\sum_{i=1}^{N} \psi(r_{i;t}) = 0,$$

$$\sum_{i=1}^{N} \psi(r_{i;t})Z_{i;t} = 0, \tag{C2}$$

where $\psi(u) = \mathrm{d}\rho(u)/\mathrm{d}u = u$. The functions $\rho(u)$ and $\psi(u)$ are shown in Figs 15a and b. In the terminology of robust statistics, $\psi(u)$ is called the 'influence function'; it measures the influence of the residuals $r_{i;t}$ on the parameter estimators. Clearly, $\psi(r_{i;t})$ increases without bound as $r_{i;t}$ increases (very large positive or very large negative), and thus least-squares is said to be non-robust.

Many alternatives to $\rho(u) = u^2/2$ have been derived. The best known is probably Huber's given by

$$\rho(u) = \begin{cases} u^2/2, & \text{if } |u| \le a, \\ a|u| - a^2/2, & \text{if } |u| > a, \end{cases}$$

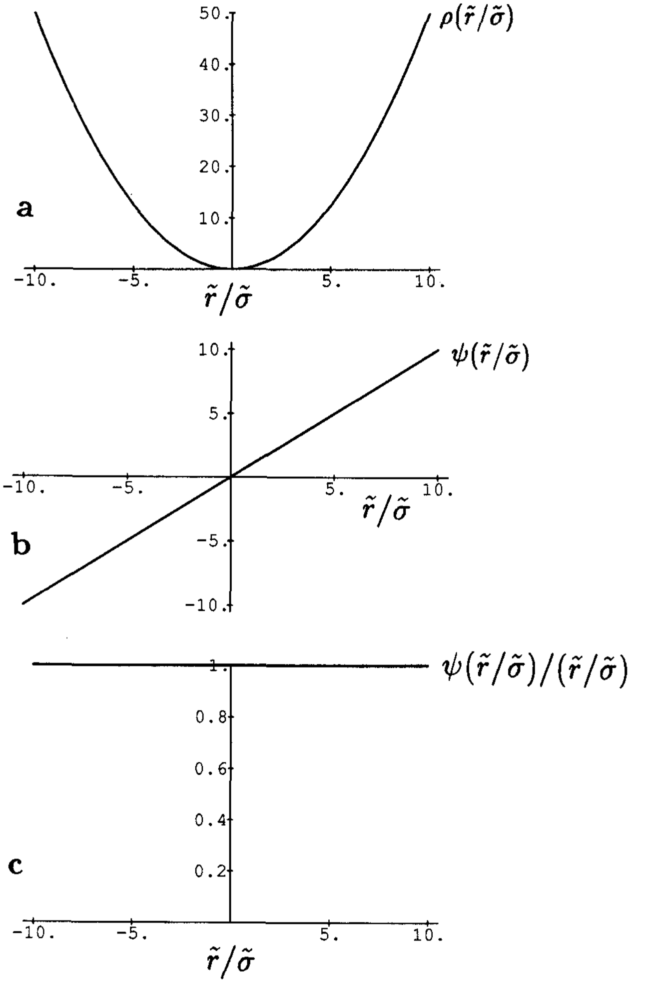



FIG. 15. Characteristic curves for least-squares. (a) The function to be minimized, $\rho(\tilde{r}/\tilde{\sigma})$, (b) its derivative, $\psi(\tilde{r}/\tilde{\sigma})$, and (c) the equivalent scaled weight function, $w\tilde{\sigma}$.

for which $\psi(u) = \mathrm{d}\rho(u)/\mathrm{d}u$ is given by

$$\psi(u) = \begin{cases} u, & \text{if } |u| \leq a, \\ a\,sgn(u), & \text{if } |u| > a, \end{cases}$$

illustrated in Figs 16a and b for varying $a$.

Huber's $\psi$ ensures that any residual exceeding a cut-off denoted by $a$, has a fixed influence on the parameter estimators, i.e. the influence will not grow with increasing residual. Nevertheless, arbitrarily large residuals will still have an influence.

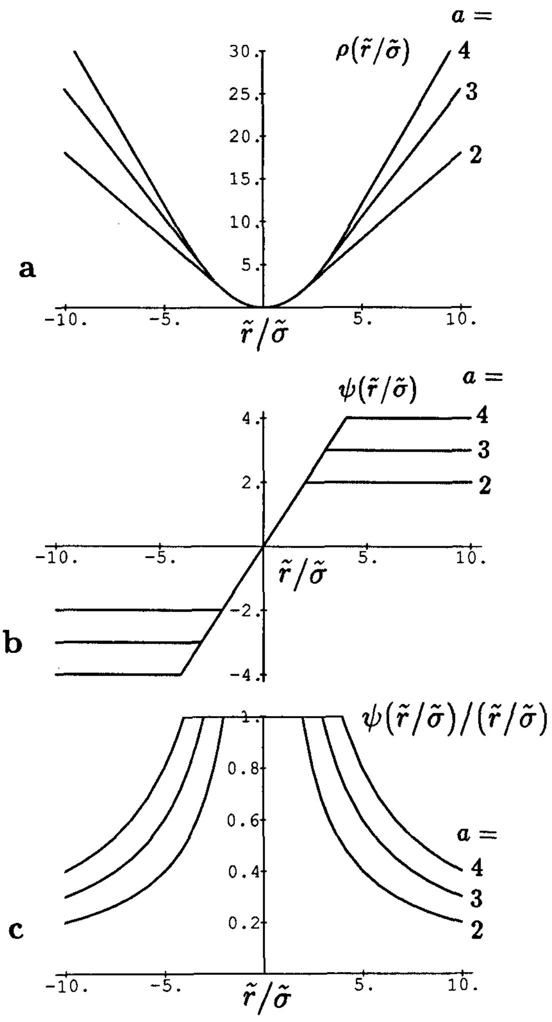



FIG. 16. Characteristic curves for Huber's method. (a) The function to be minimized, $\rho(\tilde{r}/\tilde{\sigma})$, (b) its derivative, $\psi(\tilde{r}/\tilde{\sigma})$, and (c) the equivalent scaled weight function, $w\tilde{\sigma}$. Three parameter values are considered, $a = 2, 3, 4$.

This can be overcome by using what is called a re-descending $\psi$ function, for which, past a certain point, residuals have no influence on parameter estimators. One such is used by Andrews (1974), for which

$$\rho(u) = \begin{cases} [1 - \cos(u)], & \text{if } |u| < \pi, \\ 0, & \text{if } |u| \geq \pi, \end{cases}$$

and

$$\psi(u) = \begin{cases} \sin(u), & \text{if } |u| < \pi, \\ 0, & \text{if } |u \geq \pi. \end{cases} \tag{C3}$$

These functions are shown in Figs 17a and b. The cut-off occurs at $\pi$.

Whereas least-squares is invariant to a rescaling of the residuals, the cut-off in Andrew's $\psi$ means that the corresponding equations of the form (C2) are scale dependent.

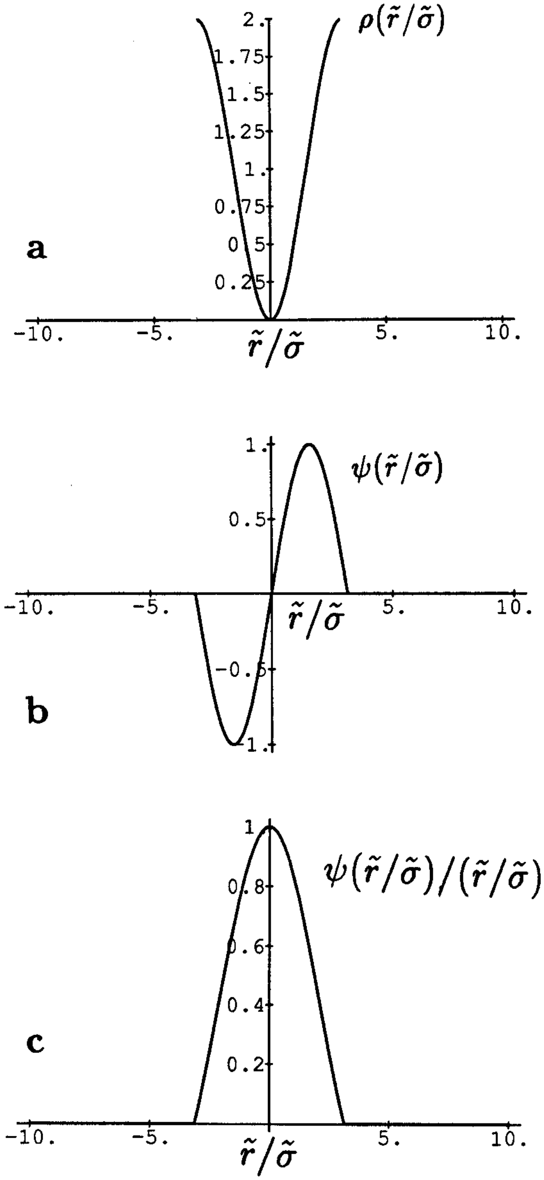



FIG. 17. Characteristic curves for Andrew's method. (a) The function to be minimized, $\rho(\tilde{r}/\tilde{\sigma})$, (b) its derivative, $\psi(\tilde{r}/\tilde{\sigma})$, and (c) the equivalent scaled weight function, $w\tilde{\sigma}$.

We need to replace (C2) by

$$\sum_{i=1}^{N} \psi\left(\frac{r_{i;t}}{\hat{\sigma}_t}\right) = 0,$$

$$\sum_{i=1}^{N} \psi\left(\frac{r_{i;t}}{\hat{\sigma}_t}\right) Z_{i;t} = 0. \tag{C4}$$

A scale estimate $\hat{\sigma}_t$ which gives an efficiency of 96% relative to the mean for Gaussian residuals where the problem is that of intercept estimation only (i.e. $\beta_t = 0$) was found by Andrews *et al.* (1972) to be

$$\hat{\sigma}_t = c \text{ med } \{ |r_{i;t}|, \quad i = 1, \ldots, N\},$$

with $c = 2.1$, and will be adopted here. (Efficiency of 96% means that the ratio, variance of the mean/variance of Andrew's $\psi$, is 0.96. Good robust statistics should have a high efficiency in the ideal case of Gaussian residuals.) Now (C4) can be written as

$$\sum_{i=1}^{N} \left[\psi\left(\frac{r_{i;t}}{\hat{\sigma}_t}\right) \Big/ r_{i;t}\right] r_{i;t} = 0,$$

$$\sum_{i=1}^{N} \left[\psi\left(\frac{r_{i;t}}{\hat{\sigma}_t}\right) \Big/ r_{i;t}\right] r_{i;t} Z_{i;t} = 0,$$

i.e., as

$$\sum_{i=1}^{N} w_{i;t} r_{i;t} = 0,$$

$$\sum_{i=1}^{N} w_{i;t} r_{i;t} Z_{i;t} = 0,$$

where $w_{i;t} = \{\psi(r_{i;t}/\hat{\sigma}_t)/r_{i;t}\}$ with $\psi$ defined in (C3), which corresponds to the minimization of

$$\sum_{i=1}^{N} w_{i;t}(Y_{i;t} - \alpha_t - \beta_t Z_{i;t})^2. \tag{C5}$$

Thus, if we knew the weights *a priori*, we could carry out the re-descending $\psi$ method using weighted least-squares. Unfortunately, the weights depend on the residuals which are unknown, but can be estimated using the preliminary estimates $\tilde{\alpha}_t$ and $\tilde{\beta}_t$:

$$\tilde{r}_{i;t} = Y_{i;t} - \tilde{\alpha}_t - \tilde{\beta}_t Z_{i;t}.$$

We can solve (C5) using standard weighted least-squares (Appendix A) where the weights are given by

$$w_{i;t} = \psi\left(\frac{\tilde{r}_{i;t}}{\tilde{\sigma}_t}\right) \Big/ \tilde{r}_{i;t}$$

and

$$\tilde{\sigma}_t = 2.1 \text{ med } \{ |\tilde{r}_{i;t}|, \quad i = 1, \ldots, N\}.$$

This completes the estimation of $\alpha_t$ and $\beta_t$. Clearly (C5) could be solved iteratively, by computing new residuals and up-dating the weights repeatedly but this is not guaranteed to converge (Andrews 1974; Hogg 1979) and is costly. A single step is often used, as here.

It is worth pointing out that so-called $L1$-norm estimation corresponds to a $\psi$ function which is constant negative for negative residuals, and constant positive for positive residuals. It therefore has an undesirable discontinuity at zero, and does not re-descend as does Andrew's $\psi$. It could, however, be used for computing preliminary parameter estimates instead of the regression by medians discussed in Appendix B.

## Appendix D
## The Runs-test Statistic

The residuals resulting from fitting a straight line to an amplitude profile in AVO analysis have a natural ordering attached to them, namely offset. By examining runs of positive and negative residuals we can see whether they are consistent with being random erorrs as required if the model is correct. To compute the statistic for an amplitude profile we need (Draper and Smith 1981, p. 159) the number of positive residuals ($n_1$), the number of negative residuals ($n_2$) and the number of runs of positive or negative residuals ($u$). For example, if for $N = 22$ the signs of the residuals were ordered as

$$(+\ +\ +)(-\ -)(+)(-\ -\ -)(+)(-)(+\ +\ +\ +)(-)(+)(-\ -\ -\ -)(+),$$

then we have $n_1 = 11$, $n_2 = 11$ and $u = 11$ runs indicated by the brackets. Provided that $n_1 > 10$ and $n_2 > 10$, then approximately,

$$Z = (u - \mu \pm 1/2)/\sigma$$

has a Gaussian (normally) distribution with mean zero and variance unity, where

$$\mu = \frac{2n_1 n_2}{n_1 + n_2} + 1$$

and

$$\sigma^2 = \frac{2n_1 n_2(2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)},$$

and the continuity correction of 1/2 takes a plus sign for a lower tail 'test' ($u < \mu$) and a minus sign for an upper tail test ($u > \mu$). Strictly, the test for runs is applicable only when the occurrences which produce the pattern of runs are independent. Following estimation of regression parameters, residuals are always slightly correlated; however, unless the number of points in the regression is small this is not an important consideration. Indeed since we require $n_1$, $n_2 > 10$, we need $N > 20$ anyway, which is certainly large enough to overcome the slight correlation problem.

## References


Andrews, D.F. 1974. A robust method for multiple linear regression. *Technometrics* **16**, 523–531.

Andrews, D.F., Bickell, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H. and Tukey, J.W. 1972. *Robust Estimates of Location: Survey and Advances*. Princeton University Press.

Chambers, J.M. 1971. Algorithm 410: Partial sorting. *Communications of the Association for Computer Machinery* **14**, 357–358.

Denham, L.R., Palmeira, R.A.R. and Farrell, R.C. 1985. The zero-offset stack. 55th SEG meeting, Washington DC, Expanded Abstracts, 624–625.

Draper, N. and Smith, H. 1981. *Applied Regression Analysis*, 2nd edn. John Wiley & Sons, Inc.

Hill, R.W. and Holland, P.W. 1977. Two robust alternatives to least-squares regression. *Journal of the American Statistical Association* **72**, 828–833.

Hocking, R.R. 1983. Developments in linear regression methodology: 1959–1982. *Technometrics* **25**, 219–230.

Hogg, R.V. 1979. An introduction to robust estimation. In: *Robustness in Statistics*, R. L. Launer and G. N. Wilkinson (eds), 1–17, Academic Press, Inc.

Mood, A.M. 1950. *Introduction to the Theory of Statistics*. McGraw-Hill Book Co.

Shuey, R.T. 1985. A simplification of the Zoeppritz equations. *Geophysics* **50**, 609–614.

Singleton, R.S. 1969. Algorithm 347: Sort. *Communications of the Association for Computer Machinery* **12**, 185–186.

Yilmaz, O. 1987. *Seismic Data Processing*. Society of Exploration Geophysicists.