Biometrika (1989), **76**, 1, *pp.* 149–60
Printed in Great Britain

Robust estimation in the errors-in-variables model

By RUBEN H. ZAMAR
Department of Statistics, University of British Columbia, Vancouver, B.C., Canada, V6T 1W5

SUMMARY

Orthogonal regression analogues of M estimates of regression, called here orthogonal regression M estimates, are defined. These estimates are shown to be consistent at elliptical errors-in-variables models and robust, if the corresponding loss function is bounded. The orthogonal regression analogues of regression scale estimates, called here orthogonal regression S estimates, are considered as well. In particular, they provide a robust estimate for the scale of the orthogonal residuals, a crucial quantity in the computation of orthogonal regression M estimates. Finally we present an algorithm for computing orthogonal regression S and M estimates and the results of a small Monte Carlo experiment.

Some key words: Errors-in-variables; M estimate; Orthogonal regression; Robustness.

1. INTRODUCTION

This paper deals with robust estimation of the parameters of a linear model when all the variables are subject to errors. Given data consisting of $(p+1)$-tuples $x_1, \ldots, x_n$, the classical errors-in-variables model is

$$x_i = X_i + \varepsilon_i, \quad a_0' X_i = b_0, \tag{1}$$

where a_0 is a vector of length one and $X_1, \ldots, X_n, \varepsilon_1, \ldots, \varepsilon_n$ are nonobservable independent random vectors. Except for sign change, a_0 and b_0 are unique and equations among stochastic quantities are intended to hold almost surely.

It is also usually assumed that ε_i is normal, $E(\varepsilon_i) = 0$ and $\operatorname{cov}(\varepsilon_i) = \sigma^2 I$. The Gaussian maximum likelihood procedure under model (1) is the method of orthogonal regression. A detailed discussion of the errors-in-variables model and the method of orthogonal regression is given, for example, by Fuller (1987, Ch. 1-2).

Whether or not the assumption of normality holds, the method of orthogonal regression is taken to mean finding the unit vector $\hat{a}$ and the number $\hat{b}$ which solve

$$\min_{\|a\|=1, b} \sum (a' x_i - b)^2. \tag{2}$$

Notice that $(a'x_i - b)^2$ is the square of the orthogonal distance from the point x_i to the hyperplane $H(a, b) = \{y\colon a'y = b\}$. It can be shown that $\hat{a}$ is the principal component of the sample covariance matrix, associated to the smallest eigenvalue. Furthermore $\hat{b} = \hat{a}'\bar{x}$, where $\bar{x}$ is the sample mean.

Two additional features which make this method attractive in some practical situations are: (a) it treats all the variables in a symmetric way since it does not distinguish between response and explanatory variables; (b) under some regularity conditions, it is consistent and asymptotically normal. However, it has long been recognized that classical regression methods are nonrobust as they are very sensitive to some kinds of nonnormality of the data and to high leverage points in the design. Carroll & Gallo (1982) and Brown (1982) point to the lack of robustness as a very severe problem of the method of orthogonal regression.

This motivates the search for closely related robust alternatives. Carroll & Gallo (1982) introduce a consistent, asymptotically normal robust estimate which can be used whenever each design point is replicated exactly twice. Brown (1982) discusses the possibility of using reweighted orthogonal regression to estimate a straight line through the origin. However, if poor initial estimates of the true slope and 'error free' explanatory variables, $\hat{\beta}_0, \hat{X}_1, \ldots, \hat{X}_n$, are used to compute the weights, and some outliers remain unchecked, then the final estimates will be poor as well.

Based on some simulations, Brown concludes that robust linear regression may be preferable to robust orthogonal regression in the errors-in-variables set-up. As we shall see later, this conclusion can be reversed by using properly defined and properly computed orthogonal regression M estimates.

Orthogonal regression M estimates are defined as solutions of the minimization problem

$$\min_{\|a\|=1,b} \sum \rho\{(a'x_i - b)/S_n\}, \tag{3}$$

where ρ is some loss function designed to induce robustness in the resulting estimate and S_n is some estimate of the scale of the orthogonal residuals. Notice that (2) is the particular case of (3) when $\rho(t) = t^2$ and $S_n = 1$. We show later in § 4 that to obtain robust estimates ρ must be bounded and S_n must be robust. This is natural since the role of ρ is to downweight the influence of residuals which are large in comparison to S_n. Therefore, robust estimation of the scale of the orthogonal residuals is an important related problem.

Orthogonal regression S estimates are defined as follows: for a given unit vector a and number b, let $S(a, b)$ be the M scale of $a'x_1, \ldots, a'x_n$; that is, $S(a, b)$ is a solution of the equation

$$n^{-1} \sum \chi\{(a'x_i - b)/S\} = \beta,$$

where χ is even, continuous, nondecreasing on $[0, \infty)$, $\chi(0) = 0$ and $\lim \chi(t) = 1$ as $t \to \infty$. Furthermore, β is a constant, usually taken to be equal to $E\{\chi(Z)\}$, where Z is a standard normal random variable. For details on M scale see, for example, Huber (1981, p. 109). Orthogonal regression scale estimates $(\hat{a}, \hat{b})$ are implicitly defined as the minimizers of $S(a, b)$. Finally, $S_n = S(\hat{a}, \hat{b})$ is a robust estimate of the scale of the orthogonal regression residuals.

The idea of constructing robust estimates by means of minimizing a robust scale was first used by Rousseeuw (1982) in the context of linear regression; see also Rousseeuw & Yohai (1984). This idea was also exploited by Li & Chen (1985) to define robust estimates of multivariate scale and principal components.

The rest of the paper is organized as follows. A computing algorithm is presented in § 2. An application and examples are given in § 3. Asymptotic and robustness properties of orthogonal regression M estimates are discussed in § 4. In particular it is shown that these estimates are consistent at the errors-in-variables model if the distribution of the error, ε, in (1) is spherically symmetric, that is, the distribution of $a'\varepsilon$ is the same for all unit vectors a. It is also shown that they are robust provided the loss function ρ is bounded.

Unfortunately consistency does not hold in general. Indeed, in § 5 examples are presented in which the estimate is actually asymptotically biased. On the other hand, the asymptotic bias seems to be fairly small in all the cases considered.

An important assumption for our consistency proof is that $\text{cov}(A\varepsilon_i) = \sigma^2 I$ for some specified matrix A, which therefore can be assumed to be equal to the identity matrix, I, without loss of generality. On the other hand, if A cannot be specified and can be consistently and robustly estimated by A_n say, orthogonal regression M estimates can still be computed, using A_n instead of A; we conjecture that in this case consistency and robustness can be preserved. This, however, deserves further study.

2. Computing algorithm and Monte Carlo

The computing algorithm is laid out in four steps as follows.

Step 1. Given data $x_1, \ldots, x_n$, compute $y_i = x_i - m$, where m is some robust multivariate location estimate. A simple choice, adopted in our simulations and examples below, is the coordinate-wise median.

Step 2: *Reparameterization.* To avoid redundancy and to allow a simple differential approach, the unit vector a is expressed in polar coordinates; that is, $a = a(\theta)$, where $\theta = (\theta_1, \ldots, \theta_p)$ with $0 \leq \theta_j \leq \pi$, for $j = 1, \ldots, p-1$ and $0 \leq \theta_p \leq 2\pi$. More precisely, $a = a(\theta) = (a_1(\theta), \ldots, a_{p+1}(\theta))$ with

$$a_1(\theta) = \sin\theta_1 \ldots \sin\theta_p, \quad a_2(\theta) = \sin\theta_1 \ldots \sin\theta_{p-1}\cos\theta_p,$$

$$a_3(\theta) = \sin\theta_1 \ldots \sin\theta_{p-2}\cos\theta_{p-1}, \quad \ldots, \quad a_p(\theta) = \sin\theta_1\cos\theta_2, \quad a_{p+1}(\theta) = \cos\theta_1.$$

Observe that the symmetric treatment of the data and the compactness of the parameter space are preserved by the new parameterization.

Step 3: *Initial orthogonal regression S estimates.* Initial values $a_1(\hat{\theta})$ and robust scale $\hat{S}_n$ are found as follows. For each θ in

$$C = \{\theta: 0 \leq \theta_j \leq \pi, 0 \leq \theta_p \leq 2\pi, j = 1, \ldots, p-1\}$$

let $S(\theta)$ be the solution to

$$n^{-1} \sum \chi\{e_i(\theta)/S\} = \beta, \tag{4}$$

where $e_i(\theta) = a(\theta)' x_i$. The function χ and β must be chosen so that $S(\theta)$ is a smooth function of θ and $\hat{S}_n$ is robust. For simulations and examples, we use Tukey's loss function for χ, with tuning constant $c = 1{\cdot}56$, see (5) below, and $\beta = 0{\cdot}05$. Smoothness of $S(\theta)$ is necessary to ensure a relatively fast and stable minimization procedure. Robustness of $\hat{S}_n$ is necessary for the next step.

The minimization of $S(\theta)$ entails two steps, a grid search and a gradient search. First, let $C_1 = \{\theta_1, \ldots, \theta_N\}$ be a grid of δ-equispaced points in C and $S(\theta_k)$ the scale of the orthogonal residuals $e_i(\theta_k)$. By direct comparison, find the point $\theta^{(0)}$ in C_1 which minimizes $S(\theta_k)$. Secondly, the gradient, $\dot{S}(\theta) = (\dot{S}_1(\theta), \ldots \dot{S}_p(\theta))'$, of $S(\theta)$ can be obtained by differentiating (4). It is easy to see that

$$\dot{S}_j(\theta) = \frac{\sum \chi\{e_i(\theta)/S(\theta)\} z_{ij}(\theta)}{\sum \chi\{e_i(\theta)/S(\theta)\} e_i(\theta)},$$

where $z_{ij}(\theta) = (\partial/\partial\theta_j) e_i(\theta)$. The gradient search for a local minimum 'close' to $\theta^{(0)}$ is as follows. Given $\theta^{(m)}$ let

$$\Delta^{(m)} = (\dot{S}_1(\theta^{(m)}), \ldots, \dot{S}_p(\theta^{(m)}))', \quad \theta^{(m+1)} = \theta^{(m)} + \delta^{(m)} \Delta^{(m)},$$

where $0<\delta^{(m)}\leq 1$ is chosen so that $\|\theta^{(m+1)}-\theta^{(m)}\|<\delta$, the size of the grid C_1, and $S(\theta^{(m+1)})<S(\theta^{(m)})$. The iteration stops at step m_0 if $\|\Delta^{(m_0)}\|$ is smaller than some predefined $\varepsilon>0$. An important feature of this gradient search is that, at each step, it forces a reduction in the value of $S(\theta)$.

Step 4: *Computing the final estimates.* Based on an appropriate loss function ρ, such as Tukey's in (5) below with $c=4\cdot 7$, which we use in numerical computations, and the initial estimates $a_1(\hat{\theta})$ and $\hat{S}_n$ from Step 3, compute the final orthogonal regression M estimate $\hat{a}=a(\hat{\theta})$. The function ρ must be such that the corresponding estimate is fairly efficient at the pure 'target' model. For example, the proposed choice achieves 95% efficiency at the Gaussian errors-in-variables model.

The minimization of $M(\theta)=n^{-1}\Sigma\,\rho\{e_i(\theta)/\hat{S}_n\}$ is as in Step 3. It can be easily seen that the gradient in this case is $\dot{M}(\theta)=(\dot{M}_1(\theta),\ldots,\dot{M}_p(\theta))'$, where

$$\dot{M}_j(\theta)=n^{-1}\sum \psi\{e_i(\theta)/\hat{S}_n\}z_{ij}(\theta).$$

Finally let $\hat{b}=\text{median}\,\{\hat{a}'x_1,\ldots,\hat{a}'x_n\}$.

A modest Monte Carlo experiment was performed to investigate the small-sample behaviour of orthogonal regression M estimates. The particular orthogonal regression M estimate considered here uses Tukey's loss function,

$$\rho(t)=\min\{1, 3c^{-2}(t^2-c^{-2}t^4+3c^{-4}t^6)\}, \tag{5}$$

with $c=4\cdot 7$. This estimate is compared with its relative, the classical M estimate of regression based on the same loss function, and two nonrobust alternatives, the usual orthogonal regression and least-squares.

We generated 100 samples of size $n=20$ of pseudo-random variables x_i and y_i following a 5% contaminated Gaussian errors-in-variables model; that is,

$$x_i=X_i+u_i,\quad y_i=Y_i+v_i,\quad (1+\beta^2)^{-\frac{1}{2}}(Y_i-\beta X_i)=\alpha,$$

with

$$X_i\sim N(0,1),\quad u_i\sim \text{CN}(0\cdot 25, \sigma^2, 0\cdot 05),\quad v_i\sim \text{CN}(0\cdot 25, \tau^2, 0\cdot 05).$$

Here, $\text{CN}(\sigma_1^2, \sigma_2^2, \varepsilon)=(1-\varepsilon)N(0,\sigma_1^2)+\varepsilon N(0,\sigma_2^2)$.

In this simulation, $\alpha=0$ and β was chosen at random uniformly between -5 and 5 to take account of the fact that the effect of outliers on the competing estimates depends on the 'true' value of β.

Six estimates of β were computed:

(i) T_1, the classical least-squares estimate;

(ii) T_2, the classical orthogonal regression estimate;

(iii) T_3, the repeated medians estimate (Siegel, 1982), which has a breakdown point of $\frac{1}{2}$ and is often used as initial estimate in the linear regression set-up;

(iv) T_4, the M estimate of regression using (5) with $c=4\cdot 7$ and computed by the usual reweighted least-squares algorithm, with T_3 as initial estimate;

(v) T_5, the orthogonal regression S estimate using (5) with $c=1\cdot 56$ and $b=0\cdot 5$; and

(vi) T_6, the orthogonal regression M estimate using (5) with $c=4\cdot 7$.

Both T_5 and T_6 are computed by the algorithm described above; T_4 and T_6 are 95% efficient at the Gaussian linear regression and errors-in-variables models, respectively.

Since the errors-in-variables model is invariant under orthogonal transformations, so must be the criterion used to measure the performance of the estimates. We adopt

$$m = \sum_{j=1}^{100} \left\{ 1 - \frac{|1 + T_{ij}\beta_j|}{(1 + T_{ij}^2)^{\frac{1}{2}}(1 + \beta_j^2)^{\frac{1}{2}}} \right\} \quad (i = 1, \ldots, 6) \tag{6}$$

as a measure of performance.

Notice that the unit vector $a(\beta)$ yielding a hyperplane through the origin with slope β is $(1+\beta^2)^{\frac{1}{2}}(-\beta, 1)$. Therefore, the general term in (6) is equal to

$$\tfrac{1}{2} \min \{\|a(T_{ij}) - a(\beta_j)\|^2, \|a(T_{ij}) + a(\beta_j)\|^2\},$$

which is orthogonally invariant and between zero and one.

Table 1 summarizes the Monte Carlo results. The orthogonal regression M estimate outperforms the M estimate in all the sampling situations. Classical orthogonal regression is better than least-squares when $\sigma \geq \tau$, that is when the contamination in the x-coordinate is more severe. When the pure model is in force, $\sigma = \tau = 0{\cdot}5$, orthogonal regression is slightly better than its robust counterpart. Not surprisingly, however, it rapidly deteriorates in the presence of contamination.

Table 1. *Simulated performance measure m as in* (6) *for regression estimates* T_1, T_2, T_4 *and* T_6; T_1, *least-squares*; T_2, *orthogonal regression*; T_4, *M estimate*; T_6, *orthogonal regression M estimate. Sample size*, 20. *Number of replications*, 100

Contamination std dev.					
σ	τ	T_1	T_2	T_4	T_6
0·5	0·5	0·95	0·58	2·45	0·87*
0·5	2·0	1·21	1·99	2·20	0·56
0·5	5·0	3·09	12·30	2·69	0·84
2·0	0·5	2·70	0·94	2·25	0·70
2·0	2·0	2·49	1·33	2·61	0·81
2·0	5·0	3·95	7·21	3·73	0·85
5·0	0·5	14·28	5·60	2·92	1·03
5·0	2·0	13·99	8·31	4·30	1.09
5·0	5·0	15·40	9·69	2·85	1·19

* No-contamination case.

3. Application and examples

Robust orthogonal regression methods can also be used to identify multidimensional outliers in situations when classical methods are not very reliable, as, for example, when outliers occur in bunches and mask each other. In fact, robust orthogonal regression can help to find projections which are 'interesting' from the outlier-detection point of view, as shown below.

Let $x_1, \ldots, x_n$ be a sample of multivariate data. Consider the projection index

$$\frac{S^2(a'x_1, \ldots, a'x_n)}{\text{var}\,(a'x_1, \ldots, a'x_n)}, \tag{7}$$

where a is a unit vector and S is a robust scale estimate. The minimizing unit vector $\hat{a}_0$

gives the direction producing the largest difference between a robust and a nonrobust measure of dispersion. For a survey of projection pursuit techniques, see Huber (1984).

Since the minimum of (7) is invariant under affine linear transformations one can equally well work with the standardized data $y_i = V^{-\frac{1}{2}}(x_i - m)$. Here, m and V are the sample mean and covariance matrix. For the standardized data, the denominator of (7) is constant, equal to one, and the minimization problem reduces to that of searching for the direction of smallest robust spread, that is, a robust orthogonal regression S or M estimate.

Example 1. The data in Table 2 were created so that cases 1, 2, 19 and 20 are multivariate outliers but neither their natural nor principal component coordinates are unusually large. Indeed, the largest Mahalanobis distance for these data, 6·35 for case 19, is well below the 95th χ_3^2 percentage point, 7·81. On the other hand, when the standardized data $y_1, \ldots, y_n$ are projected on the direction of the corresponding orthogonal regression M estimate $\hat{a}_0$ with loss function (5), the outlying character of these cases is apparent; see Fig. 1.

Example 2. The data in Table 3 on simultaneous pairs of measurements of serum

Table 2. *Artificial data including four multivariate outliers, cases* 1, 2, 19 *and* 20

Case	X_1	X_2	X_3	Case	X_1	X_2	X_3
1	−95·7	4·8	32·6	11	6·6	0·6	−1·1
2	−85·2	7·1	28·7	12	12·5	1·5	9·6
3	−75·5	−7·9	−13·5	13	24·0	4·0	15·8
4	−63·4	−6·8	25·1	14	36·1	3·0	−20·1
5	−57·0	−5·4	−12·0	15	45·1	5·3	−2·9
6	−44·9	−4·1	−6·7	16	52·7	5·7	3·6
7	−35·9	−4·0	−25·2	17	64·9	6·6	−13·0
8	−24·3	−2·5	−12·9	18	75·0	7·6	20·8
9	−13·2	−1·8	−15·1	19	85·9	−5·7	10·3
10	−7·1	−1·7	−16·2	20	93·9	−4·1	11·7

Fig. 1

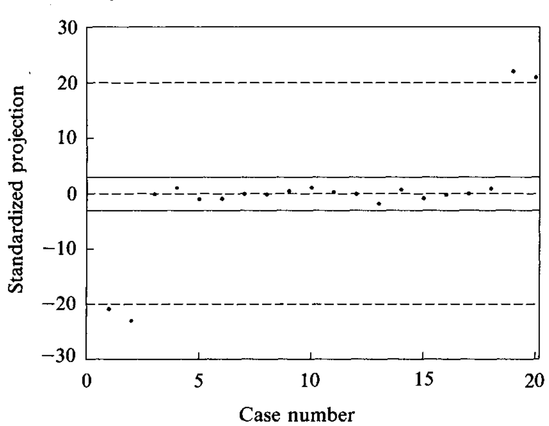

Fig. 2

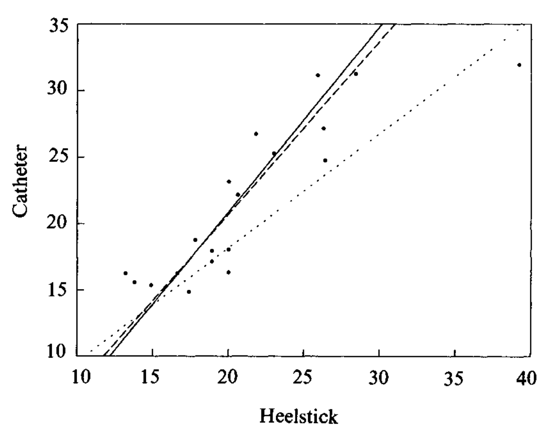

Fig. 1. Data from Table 2 standardized and projected on direction determined by orthogonal regression M estimate. Cases 1, 2, 19 and 20 more than 20 robust-scale units away from origin. Cases 3 to 18 within 3 robust-scale units from origin.

Fig. 2. Three different fits for data in Table 3. Solid line, orthogonal regression M estimate; dashed line, orthogonal regression estimate with case number two, outlying point on upper-right corner, removed; dotted line, orthogonal regression estimate, using all data.

Table 3. *Serum kanamycin levels in blood samples from umbilical catheter and heel venapuncture in samples from 20 babies*

Case	Heelstick	Catheter	Case	Heelstick	Catheter
1	23·0	25·2	11	26·4	24·8
2	33·2	26·0	12	21·8	26·8
3	16·6	16·3	13	14·9	15·4
4	26·3	27·2	14	17·4	14·9
5	20·0	23·2	15	20·0	18·1
6	20·0	18·1	16	13·2	16·3
7	20·6	22·2	17	28·4	31·3
8	18·9	17·2	18	25·9	31·2
9	17·8	18·8	19	18·9	18·0
10	20·0	16·4	20	13·8	15·6

kanamycin levels in blood samples drawn from 20 babies are given by Kelly (1984). The assumption that both measurements are subject to random errors with equal variances seems reasonable. To illustrate the behaviour of different estimates in the presence of outliers, case number 2 was changed from its original value (33·2, 26·0) to (39·2, 32·0). This is the outlying point in the upper right-hand corner of Fig. 2. Classical orthogonal regression gives the dotted line in Fig. 2, with slope 0·86 and intercept 0·97. If case 2 is deleted the corresponding orthogonal regression fit becomes the dashed line, with slope 1·30 and intercept −5·29. Observe the sensitivity of this method to the presence of just one outlier in the data. The solid line with slope 1·39 and intercept −6·91, corresponds to the orthogonal regression M estimate with loss function (5). This is very close to the classical fit without the outlier. The M estimate line with slope 0·85 and intercept 3·37 is not shown. Observe the similarity between the M and the classical orthogonal regression estimates of the slope.

4. Some asymptotic and robustness results

4·1. *Consistency of orthogonal regression M estimates*

The following theorem follows from Huber (1967, Th. 1). Details are provided in the author's University of Washington Ph.D. thesis.

Theorem 1. *Let $x_1, \ldots, x_n$ be independent, identically distributed random vectors with common distribution F. Suppose*: (*a*) *ρ is continuous, nonnegative and nondecreasing on $[0, \infty)$*; (*b*) *there exists $0 < s < \infty$ such that $s_n \to s$ almost surely $[F]$ as $n \to \infty$*; (*c*) *there exists a vector (a_1, b_1), $\|a_1\| = 1$, which uniquely, up to sign changes, minimizes $E_F[\rho\{(a'x - b)/s\}]$ among all unit vectors a and real numbers b. Let*

$$A_n = \inf_{\|a\|=1, b} n^{-1} \sum_{i=1}^{n} \rho\{(a'x_i - b)/s\}.$$

If the sequence $(\hat{a}_n, \hat{b}_n)$, $\|\hat{a}_n\| = 1$, satisfies $n^{-1} \sum \rho\{(\hat{a}_n' x_i - \hat{b})/s_n\} - A_n \to 0$ almost surely $[F]$, then $(\hat{a}_n, \hat{b}_n) \to (a_1, b_1)$ almost surely $[F]$.

If the common distribution F is given by model (1), one may ask under what conditions the orthogonal regression M estimate is consistent; that is, under what conditions $(a_1, b_1) = (a_0, b_0)$. The following corollary gives a sufficient condition.

Corollary 1. *Suppose that x_i is as in model* (1) *and*: (*i*) *the distribution of ε_i is spherically symmetric*; (*ii*) *the density, h, of $a'\varepsilon_i$ is unimodal and continuous*; (*iii*)

$E\{\rho(a_0'\varepsilon_i)\}<\infty$; and (iv) there exists $\delta>0$ such that ρ and h are strictly monotone on $[0, \delta)$. If assumptions (a) and (b) of Theorem 1 hold, then $(\hat{a}_n, \hat{b}_n)\rightarrow(a_0, b_0)$ almost surely $[F]$.

4·2. *Contaminated errors-in-variable distributions*

We now study the asymptotic behaviour of orthogonal regression M estimates when the underlying distribution belongs to the family of contamination distribution functions

$$F=(1-\varepsilon)F_0+\varepsilon H, \tag{8}$$

where F_0 is the distribution of the data under the Gaussian errors-in-variable model, see (1), H is an arbitrary distribution on $R^{(p+1)}$ and $0<\varepsilon<0{\cdot}5$. Equation (8) provides a simple way to model a sampling distribution which generates occasional aberrant data values, which typically appear in applications.

An orthogonal M estimate, $\hat{a}_n$, can be viewed as a functional $\hat{a}(.)$ defined on a subset of distribution functions. In particular $\hat{a}_n=\hat{a}(F_n)$, where F_n is the empirical distribution of the data. Assume that $\hat{a}(F_0)=a_0$ and $\hat{a}(.)$ is continuous at all F in (8) so that $\hat{a}(F_n)\rightarrow\hat{a}(F)$. See Theorem 1 and Corollary 1.

The asymptotic bias of $\hat{a}(.)$ at F, $B(\hat{a}, F)$, is defined as

$$B(\hat{a}, F)=\tfrac{1}{2}\min\{\|\hat{a}(F)-\hat{a}_0\|^2, \|\hat{a}(F)+\hat{a}_0\|^2\}=1-|a_0'\hat{a}(F)|. \tag{9}$$

To assess the bias it is necessary that $\hat{a}(F)$ and a_0 have the same direction. Notice that both $\hat{a}(F)$ and $-\hat{a}(F)$ define the same hyperplane. This justifies the minimum in (9). Also notice that (9) respects the orthogonal invariance of the errors-in-variables model.

A useful measure of the degree of bias-robustness of an estimate is its maximum asymptotic bias over a contamination family. A robust estimate is expected to be stable in a neighbourhood of the target model and therefore to have a relatively small maximum asymptotic bias. Clearly, the maximum asymptotic bias of an estimate is a function of ε, the fraction of contamination, so one is led to consider maximum bias curves. Two important robustness concepts, the breakdown point and the gross-error sensitivity, are derived from such curves. The breakdown point of an estimate is the smallest value of ε for which the maximum asymptotic bias attains its theoretical maximum, usually equal to ∞. The gross-error sensitivity is the derivative of the maximum bias curve at $\varepsilon=0$. It can be used, when finite, to obtain a local linear approximation to the maximum asymptotic bias curve, near zero. For discussion of these and other robustness concepts, see Hampel et al. (1986, Ch. 2).

Let V be the covariance matrix of X in (1), $0=\lambda_0<\lambda_1\leqslant\lambda_2\leqslant\ldots\leqslant\lambda_p$ the eigenvalues of V and $a_0, a_1, \ldots, a_p$, the corresponding eigenvectors. The author, in his Ph.D. thesis, shows that for an orthogonal M estimate, $\hat{a}$, with bounded ρ function the maximum asymptotic bias,

$$B(\hat{a})=\sup_F B(\hat{a}, F),$$

is given by

$$B(\hat{a})=\begin{cases} h_\rho^{-1}(\delta) & (h_\rho(1)>\delta), \\ 1 & \text{otherwise}, \end{cases} \tag{10}$$

where

$$\delta=\varepsilon/(1-\varepsilon), \quad h_\rho(\alpha)=E_{F_0}[\rho\{a(\alpha)'x\}-\rho(a_0'x)],$$
$$a(\alpha)=(1-\alpha)a_0+\{1-(1-\alpha)^2\}^{\frac{1}{2}}a_1.$$

Since ρ is bounded it can be assumed, without loss of generality, that, as $t \to \infty$, $\lim \rho(t) = 1$. It can also be shown that if ρ is unbounded then $B(\hat{a}) = 1$, for all $\varepsilon > 0$.

The asymptotic breakdown point, ε^*, of an orthogonal M estimate $\hat{a}$ is defined as

$$\varepsilon^* = \inf\{\varepsilon : B(\hat{a}) = 1\}.$$

From (10) it follows that for $\hat{a}$ based on a bounded ρ,

$$\varepsilon^* = \inf\{\varepsilon : h_\rho(1) \leqslant \varepsilon/(1-\varepsilon)\}. \tag{11}$$

On the other hand, $\varepsilon^* = 0$ for all orthogonal M estimate based on an unbounded ρ. Thus, for example, the classical orthogonal regression estimate with $\rho(t) = \frac{1}{2}t^2$ and the orthogonal M estimate based on Huber's favourite ρ-function $\rho(t) = \min(\frac{1}{2}t^2, c|t| + \frac{1}{2}c^2)$, $c > 0$, have $\varepsilon^* = 0$.

The remainder of this section is devoted to the computation of the influence curve of orthogonal regression M estimates. Once more, the reader is referred to Hampel et al. (1986, Ch. 1) for a general definition and detailed discussion of influence curve.

For simplicity, we only consider here the case when the target model is the Gaussian errors-in-variables model (1) with $p = 1$, $b_0 = 0$ and $\sigma^2 = 1$. The general case is treated in the author's thesis. The influence curve for the classical orthogonal regression estimate is derived by Kelly (1984).

In the simple case treated here, the orthogonal regression M function $\hat{a}(F)$ can be viewed as a composite function $g\{\hat{\theta}(F)\}$, where $\hat{\theta}(F)$ is defined as the minimizer of $E_F\{\rho(x_1 \sin\theta + x_2 \cos\theta)\}$ and $g(\theta) = (\sin\theta, \cos\theta)'$. It follows that $\hat{\theta}(F)$ satisfies the estimating equation

$$E_F[\psi\{e(x, \theta)\}z(x, \theta)] = 0, \tag{12}$$

where

$$e(x, \theta) = x_1 \sin\theta + x_2 \cos\theta, \quad z(x, \theta) = x_1 \cos\theta - x_2 \sin\theta.$$

From (12) we can easily derive the influence curve of $\hat{\theta}$ at $y = (y_1, y_2)$ and F_0,

$$\text{IC}(y, \hat{\theta}, F_0) = K(F_0)\psi\{e(y, \theta)\}z(y, \theta), \tag{13}$$

where

$$K(F_0) = E_{F_0}[\psi\{e(\theta_0)\}e(\theta_0) - \psi'\{e(\theta_0)\}z^2(\theta_0)].$$

Finally, since $\hat{a} = g(\hat{\theta}) = (\cos\hat{\theta}, \sin\hat{\theta})'$, it follows from (13) that

$$\text{IC}(y, \hat{a}, F_0) = (\cos\theta_0, \sin\theta_0)'\text{IC}(y, \hat{\theta}, F_0). \tag{14}$$

The asymptotic variances for $\hat{\theta}$ and $\hat{a}$ can be derived from (13) and (14) in the usual way.

5. Concluding remarks

The case where the errors ε_i in model (1) are independent and identically distributed is of considerable practical interest. In this case, the joint distribution of ε is spherically symmetric if and only if the common marginal distribution is Gaussian. In the non-Gaussian case, Corollary 1 does not apply and the orthogonal regression M estimate may be asymptotically biased. However, numerical computations indicate that the asymptotic bias is small. For example, consider the orthogonal regression M estimate using (5) with $c = 4{\cdot}7$. We found that, if the distribution of x and y is as in § 2 except that u and v are independent, Student's t random variables with k degrees of freedom, the asymptotic bias B is no larger than 0·005. The worst case, $B = 0{\cdot}005$, corresponds to $k = 4$. The Student's t distributions used in our computations were scaled so that the 'signal-to-noise ratio', r^2, measured in terms of the square of the median of the absolute deviations, is approximately equal to two. The use of a robust measure of dispersion is justified on intuitive grounds and also by the fact that Student's t distributions with $k \leqslant 2$ do not have finite variance. The asymptotic bias is quite small, despite the poor signal-to-noise ratio used in our computations. The maximum asymptotic bias over ε-contaminated neighbourhoods, computed using (10), is small too, even for moderately large values of ε. Also the breakdown point is fairly large as shown in Table 4. Therefore we can conclude that this estimate enjoys a fair degree of bias-robustness as is consistent with the Monte Carlo results in Table 1.

Table 4. *Breakdown point, ε^*, of the orthogonal regression M estimate using* (5) *with* $c = 4{\cdot}7$, *for several values of the signal-to-noise ratio* r^2

r^2	ε^*	r^2	ε^*
1·0	0·082	4·0	0·205
1·5	0·112	5·0	0·228
2·0	0·137	6·0	0·246
2·5	0·158	8·0	0·273
3·0	0·176	10·0	0·293

$\varepsilon^* \geqslant 0{\cdot}20$ for $r \geqslant 2$.

It follows from (14) that the influence curve of orthogonal regression M estimates is unbounded; there are sequences $\{y_n\}$ for which $\psi\{e(y_n, \theta_0)\}z(y_n, \theta_0) \to \infty$ as $n \to \infty$. Therefore, the gross-error sensitivity of these estimates is infinite and cannot be used to approximate their maximum asymptotic bias. This emphasizes the importance of formulae (10) and (11) which give the exact asymptotic bias and breakdown point for these estimates. Note that boundedness of the influence curve is neither a necessary nor sufficient condition for bias-robustness. For example, in the regression set-up, so called generalized M estimates have bounded influence curves and their breakdown points shrink to zero as the dimension of the factor space increases. On the other hand, S estimates of regression have unbounded influence curves but breakdown-point $\frac{1}{2}$, independently from the dimension of the factor space.

However, it may be of interest to study the bias-robustness properties of bounded influence estimates in the errors-in-variables set-up. In the linear regression case, bounded influence estimates are defined for example as the solution of

$$E[\psi\{(y - \beta' x)\|x\|\}x/\|x\|] = 0.$$

See Hampel et al. (1986, p. 315) for a detailed discussion of bounded influence curve estimation in the regression set-up.

In the orthogonal set-up one may be tempted to define bounded influence estimates as the solution to the estimating equation

$$E[\psi\{a'x\|z(a)\|\}z(a)/\|z(a)\|] = 0, \tag{15}$$

where $z(a) = (I - aa')x$ and where I is the identity matrix. The problem with this approach is that even at a spherical error-in-variables model, (15) has several 'wrong' roots in addition to the 'right' one. For example, in the classical orthogonal regression case, $\psi(t)=t$, any eigenvector of the covariance matrix of x will solve (15) and we are only interested in the one with the smallest eigenvalue.

One way to overcome this difficulty might be the following: for each $\|c\|=1$ define $\gamma(c)$ as the solution to the minimization problem

$$\min_{\|a\|=1} E[\rho\{a'x\|z(c)\|\}\|z(a)\|^{-2}]=0.$$

Then, a fixed point $c_0=\gamma(c_0)$ is the 'right' root of (15). An algorithm to compute c_0 may be given by the recursion $c^{m+1}=\gamma(c^m)$. A simpler method is to use only one-step recursion, starting from a robust initial estimate, for example an orthogonal regression M estimate with a bounded ρ.

Acknowledgements

This work was partially funded by Office of Naval Research and National Science and Engineering Research Council. The author wants to thank Professors Douglas R. Martin, Victor J. Yohai and James V. Zidek for useful comments and critiques of the earlier version of this paper.

Appendix

Proof of Corollary 1

LEMMA 1. *Suppose that: (a) the random variable Y has a unimodal and continuous density f; (b) the loss function ρ is continuous, nonnegative and nondecreasing on $[0,\infty)$; (c) there exists $\delta>0$ such that ρ and f are strictly monotone on $[0,\delta)$. If*

$$g(t)=\int\{\rho(y-t)-\rho(y)\}f(y)\,dy,$$

then, $g(t)>0$ for all $t\neq 0$.

Proof. Let $t>0$. By (a) and (b)

$$\begin{aligned}g(t)&=\int_{-\infty}^{+\infty}\{\rho(y-t)-\rho(y)\}f(y)\,dy\\&=\int_{-\infty}^{t/2}\{\rho(y-t)-\rho(y)\}f(y)\,dy+\int_{t/2}^{\infty}\{\rho(y-t)-\rho(y)\}f(y)\,dy\\&=\int_{t/2}^{\infty}\{\rho(-y)-\rho(t-y)\}f(t-y)\,dy-\int_{t/2}^{\infty}\{\rho(y)-\rho(y-t)\}f(y)\,dy\\&=\int_{t/2}^{\infty}\{\rho(y)-\rho(y-t)\}\{f(y-t)-f(y)\}\,dy\geqslant 0.\end{aligned}$$

Notice that $\rho(y)-\rho(y-t)\leqslant 0$ and $f(y-t)-f(y)\leqslant 0$ for all $y\geqslant\frac{1}{2}t$.

Finally, by (c), for $y=t$,

$$\{\rho(y)-\rho(y-t)\}\{f(y-t)-f(y)\}=\{\rho(t)-\rho(0)\}\{f(0)-f(t)\}<0,$$

and the lemma follows from continuity of ρ and f. □

Proof of Corollary 1. It suffices to show that assumption (c) of Theorem 1 holds with $(a_1,b_1)=(a_0,b_0)$. Assume, without loss of generality, that $s=1$ and $E_F\{\rho(a'x-b)-\rho(a_0'x-b_0)\}<\infty$. Notice that by (iii), if this last quantity is not finite then (c) of Theorem 1 trivially holds. By (i) the above expectation is equal to $E_F\{\rho(a'x-b)-\rho(a'\varepsilon)\}$ which can be written as

$$\int_{R^{p+1}}\int_R [\rho\{y-(b-a'z)\}-\rho(y)]h(y)\,dy\,dF_X(z). \quad \text{(A1)}$$

By (ii), (iv) and Lemma 1, the inner integral is strictly positive for all $(b-a'z)\neq 0$. Therefore, if (A1) is equal to zero, $b-a'X=0$ almost surely $[F_X]$ and $(a, b)=\pm(a_0, b_0)$. □

References

BROWN, M. (1982). Robust line estimation with errors in both variables. *J. Am. Statist. Assoc.* **77**, 71–9.

FULLER, W. A. (1987). *Measurements Errors Models.* New York: Wiley.

CARROLL, R. & GALLO, P. (1982). Some aspects of robustness in the functional errors-in-variables model. *Comm. Statist.* A **11**, 2573–85.

HAMPEL, F. R. & RONCHETTI, R. M. & ROUSSEEUW, P. J. & STHAHEL, W. A. (1986). *Robust Statistics. The Approach Based on Influence Functions.* New York: Wiley.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. 5th Berkeley Symp.* **1**, 221–33.

HUBER, P. J. (1981). *Robust Statistics.* New York: Wiley.

HUBER, P. J. (1984). Projection pursuit (with discussion). *Ann. Statist.* **13**, 435–75.

KELLY, G. (1984). The influence function in the errors in variables problem. *Ann. Statist.* **12**, 87–100.

LI, G. & CHEN, Z. (1985). Projection-pursuit approach to robust dispersion matrices and principal components: primary theory and Monte Carlo. *J. Am. Statist. Assoc.* **80**, 759–66.

ROUSSEEUW, P. J. (1982). Least median of squares regression. *J. Am. Statist. Assoc.* **79**, 871–80.

ROUSSEEUW, P. J. & YOHAI, V. J. (1984). Robust regression by means of *S*-estimates. In *Robust and Nonlinear Time Series*, Ed. W. Franke, W. Hardle and R. D. MARTIN, pp. 256–72. New York: Springer-Verlag.

SIEGEL, A. F. (1982). Robust regression using repeated medians. *Biometrika* **69**, 242–4.

[*Received September* 1987. *Revised July* 1988]