

On the number of outliers in data from a linear model

P.R. FREEMAN
University of Leicester

SUMMARY

This paper reviews models for the occurrence of outliers in data from the linear model. The Bayesian analyses are all closely similar in form, but differ in the way they treat suspected outliers. The models are compared on Darwin's data and one of them is used on data from a 2^5 factorial experiment.

The question of how many outliers are present involves comparison of models with different numbers of parameters. A solution using proper priors on all parameters is given. On two trial datasets it is found to be insensitive to choice of priors on all except the parameters representing the amount of contamination in the outliers. Here, choice of even a slightly "wrong" prior can be very misleading. Moreover, it is difficult to choose an appropriate prior when contaminations can be both positive or negative.

Keywords: CONTAMINATION; LINEAR MODEL; MODEL DISCRIMINATION; OUTLIERS; PROPER PRIORS; SPURIOUS OBSERVATIONS.

1. A VARIETY OF MODELS

Consider the common problem in which a statistician would like to use a standard linear model to represent the generation of a dataset arising from some experiment. He has, however, some doubts about whether all the observations were generated by that model and feels there is a chance that some (hopefully, a few) observations will have been contaminated in some way. Recording errors, temporary changes in experimental conditions or the use of abnormal experimental units are the kinds of flaws he has in mind. In analysing the data he must therefore elaborate his simple linear model in some parsimonious way so as to guard himself against such gremlins and ensure inferences about the parameters of interest that are robust.

The word “outlier” will here be used to mean any observation that has not been generated by the mechanism that generated the majority of observations in the dataset. Note that we automatically assume that outliers are a small minority of the observations and that for each possible alternative we must use a different model for outlier generation.

In this section we shall briefly review three such models. We first establish some common notation.

We write the standard linear model as

$$y = \chi\beta + e \quad (1.1)$$

where y is $n \times 1$ and χ is $n \times p$.

If a particular subset $y_{i1} \dots y_{ir}$ of the y 's are suspected of being outliers, we partition the y vector into $y_{(r)}$ and $y_{(n-r)}$.

A simple application of Bayes theorem shows that β , the parameter of interest, has posterior distribution that can be written

$$p(\beta | y) = \sum w_{(r)} p_{(r)}(\beta | y)$$

where the summation extends over all 2^n possible partitions of y , $w_{(r)}$ denotes the posterior probability that the subset $y_{i1} \dots y_{ir}$ are indeed outliers and $p_{(r)}(\beta | y)$ the posterior density of β given that they are outliers. The presence of outliers is thus handled automatically. If a subset is particularly discrepant, the corresponding weight $w_{(r)}$ will be large and our ideas about β will allow for the discrepancies.

In each of the following three models, $p_{(r)}(\beta | y)$ turns out to be a p -variate Student's t distribution with mean $\beta_{(r)}$, dispersion matrix $B^{-1}_{(r)}$ and degrees of freedom $v_{(r)}$, say. It is the different ways in which they treat the suspect observations in arriving at the quantities, especially $\beta_{(r)}$, that is interesting. The posterior weights $w_{(r)}$ are complex, but for a *given* number of outliers we always get $w_{(r)}$ inversely proportional to some power of $s^2_{(r)}$, a kind of “residual sum of squares” from the analysis allowing for outliers.

We shall refer throughout to the standard least-squares values

$$\begin{aligned} \hat{\beta} &= (\chi' \chi)^{-1} \chi' y \\ s^2 &= (y - \chi \hat{\beta})' (y - \chi \hat{\beta}) \end{aligned}$$

Box and Tiao (1968) first considered this problem and their model (BT) assumes that each observation has probability $1-\alpha$ of being generated by the usual linear model and small probability α of coming from the same model but with error variance $k^2\sigma^2$ instead of just σ^2 . They took k and α as known and used the usual improper uniform prior on β and $\log \sigma$.

Here

$$\begin{aligned}\hat{\beta}_{(r)} &= [\chi'_{(n-r)}\chi_{(n-r)} + k^{-2}\chi'_{(r)}\chi_{(r)}]^{-1} [\chi'_{(n-r)}y_{(n-r)} + k^{-2}\chi'_{(r)}y_{(r)}] \\ s^2_{(r)} &= [y_{(n-r)} - \chi_{(n-r)}\hat{\beta}_{(r)}]' [y_{(n-r)} - \chi_{(n-r)}\hat{\beta}_{(r)}] + k^{-2}[y_{(r)} - \chi_{(r)}\hat{\beta}_{(r)}]' [y_{(r)} - \chi_{(r)}\hat{\beta}_{(r)}] \\ v_{(r)} &= n-p \\ B_{(r)} &= (n-p) [\chi'_{(n-r)}\chi_{(n-r)} + k^{-2}\chi'_{(r)}\chi_{(r)}]/s^2_{(r)} \\ \text{and } w_{(r)} &\propto \{\alpha/k(1-\alpha)\}^r |\chi'_{(n-r)}\chi_{(n-r)} + k^{-2}\chi'_{(r)}\chi_{(r)}|^{-1/2} s_{(r)}^{-(n-p)}\end{aligned}$$

Each suspected outlier is thus dealt with by dividing the y value and the corresponding row of the χ matrix through by k and then doing the usual least-squares analysis on this new dataset.

Additive, rather than multiplicative, contamination of the data was considered by Abraham and Box (1978). Their model (AB) was

$$y = \chi\beta + \delta Z + e$$

where Z is a vector each of whose n elements has probability α of being 1 and $1-\alpha$ of being 0. The amount of contamination δ is thus assumed to be the same for each outlier. Any particular Z vector written $Z_{(r)}$ say, corresponds to a subset of observations being outliers.

Taking α known and improper uniform prior on β, δ and $\log \sigma$ gives

$$\begin{aligned}\hat{\beta}_{(r)} &= [\chi' V_{(r)}\chi]^{-1} \chi' V_{(r)}y \text{ where } V_{(r)} = I - r^{-1}Z_{(r)}Z'_{(r)} \\ s^2_{(r)} &= [y - \chi\hat{\beta}_{(r)}]' V'_{(r)} [y - \chi\hat{\beta}_{(r)}] \\ v_{(r)} &= n-p-1 \\ B_{(r)} &= \frac{n-p-1}{s^2_{(r)}} \chi' V_{(r)}\chi \\ \text{and } w_{(r)} &\propto [\alpha/(1-\alpha)]^r r^{1/2} |\chi' V_{(r)}\chi|^{-1/2} s_{(r)}^{-(n-p-1)}\end{aligned}$$

This model thus copes with outliers by doing a weighted least squares analysis using the weighting matrix $V_{(r)}$.

Guttman, Dutter and Freeman (1978) consider additive contamination in a rather different way. Their model (GDF) is

$$y = \chi\beta + a + e$$

where a is a vector exactly r of whose elements are non-zero. They assume the value of r is known, but hedge their bets by doing separate analyses for $r = 0, 1, 2, \dots$. The non-zero elements of a are not forced to be equal, but form r extra unknown parameters which are duly given a uniform improper prior along with β and $\log \sigma$. We now get

$$\begin{aligned}\hat{\beta}_{(r)} &= [\chi'_{(n-r)} \chi_{(n-r)}]^{-1} \chi'_{(n-r)} y_{(n-r)} \\ s_{(r)}^2 &= [y_{(n-r)} - \chi_{(n-r)} \hat{\beta}_{(r)}]' [y_{(n-r)} - \chi_{(n-r)} \hat{\beta}_{(r)}] \\ v_{(r)} &= n-p-r \\ B_{(r)} &= \frac{n-p-r}{s_{(r)}^2} \chi'_{(n-r)} \chi_{(n-r)} \\ \text{and } w_{(r)} &\propto |\chi'_{(n-r)} \chi_{(n-r)}|^{-1/2} s_{(r)}^{-(n-p-r)}\end{aligned}$$

The effect of allowing “totally unknown” amounts of contamination is therefore the dramatic one of dropping suspect observations completely and doing a least-squares analysis on the others.

2. DARWIN'S DATA

All these papers apply their results to the famous set of data due to Darwin quoted by Fisher (1960) and eternally popular with students of outliers.

Here the $n = 15$ observations are

-67 -48 6 8 14 16 23 24 28 29 41 49 56 60 75

and β is the unknown population mean, so $p = 1$ and χ is a column vector of ones.

Box and Tiao display the posterior density of β when $\alpha = .05$ and $k = 5$. In identifying outliers, the largest posterior probabilities are as follows:

Outliers	:	None	y_1 and y_2	y_1 only	y_2 only	y_{15} only
Prior prob	:	.463	.0013	.024	.024	.24
Posterior prob	:	.462	.190	.175	.036	.016

If we condition on a fixed number of outliers, we have

$$w_{(r)} \propto S_{(r)}^{-(n-1)}$$

where $S_{(r)}^2 = \sum_{(n-r)} [y_i - \hat{\beta}_{(r)}]^2 + k^{-2} \sum_{(r)} [y_i - \hat{\beta}_{(r)}]^2$

$$\text{and } \hat{\beta}_{(r)} = \frac{\sum_{(n-r)} y_i + k^{-2} \sum_{(r)} y_i}{n-r + k^{-2}r}$$

in obvious notations.

The largest of these conditional probabilities, for $r = 1$ and 2, are given in the columns headed BT of table 1.

TABLE 1

Posterior probabilities, given one or two outliers, for Darwin's data

One outlier				Two outliers				
Observation number	BT	AB = GDF	Observation Pair	BT	Pair	GDF	Observation pair	AB
1	.588	.579	1,2	.785	1,2	.751	1,2	.646
2	.120	.120	1,15	.037	1,15	.037	1,3	.002
15	.053	.054	1,14	.016	1,14	.017	1,4	.002
14	.030	.031	1,13	.013	1,13	.014	1,5	.001
13	.027	.028	1,12	.011	1,12	.012	1,6	.001
12	.023	.023	1,3	.010	1,3	.011	14,15	.001
11	.020	.020	1,4	.010	1,4	.011	13,15	.001
3	.018	.019	1,11	.009	1,11	.010	1,7	.001
4	.018	.019	1,6	.008	1,6	.010		

A sensitivity analysis showed that the posterior mean and variance of β are hardly affected by large changes in the value of k . While changes in α are rather more crucial, there is still a fair amount of robustness and the results do not vary much as α ranges between .03 and .07.

In the Abraham and Box model

$$\chi' V_{(r)} \chi = n-r, \quad \hat{\beta}_{(r)} = \bar{y}_{(n-r)}$$

$$S_{(r)}^2 = \sum_{(r)} [y_i - \bar{y}_{(r)}]^2 + \sum_{(n-r)} [y_i - \bar{y}_{(n-r)}]^2$$

The first term here clearly arises as a consequence of the assumption of the same δ for each outlier, $\bar{y}_{(r)}$ being the natural estimate of $\beta + \delta$.

Conditionally on r ,

$$w_{(r)} \propto S_{(r)}^{-(n-2)}$$

Note that, since $p=1$, all suspect observations are ignored in forming $\hat{\beta}_{(r)}$, but contribute towards $S_{(r)}^2$, except when $r=1$. In that case these results coincide with those of the GDF model.

Abraham and Box give the posterior density of β for a range of α values, do a sensitivity analysis on the mean and variance of β as α changes, and quote conditional posterior probabilities $w_{(r)}$ for $r=1$ and 2, reproduced here in table 1.

In the Guttman, Dutter and Freeman model,

$$\begin{aligned}\hat{\beta}_{(r)} &= \bar{y}_{(n-r)} \\ S_{(r)}^2 &= \sum_{(n-r)} [y_i - \bar{y}_{(n-r)}]^2 \\ \text{and } w_{(r)} &\propto S_{(r)}^{-(n-1-r)}\end{aligned}$$

these latter being inherently conditional on fixed r .

As Table 1 shows for only one outlier all three models agree on observation 1(-67) as being by far the most likely candidate. All the central observations from 6 to 29 get almost identical posterior probabilities, as dropping any one of them makes very little difference to the sum of squares about the mean. For two outliers, however, the Abraham-Box model diverges from the others in that it cannot encompass the possibility that outliers might occur in both tails of the distribution. It also gives less posterior weight to the most obvious pair (-67, -48) and spreads the posterior probability pretty uniformly over all except three pairs. The model is clearly not a good one for identifying outliers and so must necessarily be weak at providing robust estimates of β under some circumstances.

3. A 2⁵ FACTORIAL EXPERIMENT

John (1978) discussed the results of a 2⁵ factorial experiment in two blocks with the ABCDE interaction confounded. Visual inspection of a plot of residuals against fitted values suggests that there might be two outliers. Having derived a suitable test statistic and simulated its sampling distribution, a significance level $\alpha = .117$ was obtained, from which it was concluded that there were not two outliers. Had a test for only one outlier been performed, however, the result would have been significant with $\alpha = .044$.

Besag (1979) reports that a robustified regression analysis, using Tukey's "exploratory data" approach, clearly shows the presence of one outlier, not two.

An analysis using the GDF model fitting main effects and first-order interactions confirms this approach. Table 2 shows that assuming one outlier gives posterior probability .734 to one of the observations, whereas the most likely pair only gets probability .147. The posterior mean of β changes markedly as we change from 0 to 1 outlier but hardly at all when we progress to 2 outliers. The sum of the posterior variances of the elements of β is again least for one outlier.

TABLE 2
Data on 2⁵ factorial experiment, from John (1978)

DATA

(1)	1.4	d	5.0	e	1.7	de	9.5
a	1.2	ad	9.0	ae	2.0	ade	5.9
b	3.6	bd	12.0	be	3.1	bde	12.6
ab	1.2	abd	5.4	abe	1.2	abde	6.3
c	1.5	cd	4.2	ce	1.9	cde	8.0
ac	1.4	acd	4.4	ace	1.2	acde	4.2
bc	1.5	bcd	9.3	bce	1.0	bcde	7.7
abc	1.6	abcd	2.8	abce	1.8	abcde	6.0

POSTERIOR PROBABILITIES

One outlier

.734	ad
.098	d
.010	bcd
.009	bce
.008	abcd
.008	abcde

Two outliers

.147	ad, acd
.090	d, ad
.050	ad, abcd
.047	ad, bcde
.040	ad, ace
.040	ad, abce

POSTERIOR MEAN AND VARIANCE OF β

N° Outliers	MEAN			VARIANCE		
	0	1	2	0	1	2
(1)	4.36	4.24	4.22	.087	.074	.082
a	-0.89	-1.04	-1.09	.087	.065	.067
b	0.46	0.58	0.60	.087	.074	.084
c	-0.71	-0.58	-0.58	.087	.074	.073
d	2.66	2.53	2.51	.087	.074	.082
e	0.27	0.40	0.42	.087	.074	.084
ab	-0.64	-0.49	-0.44	.087	.065	.067
ac	0.16	0.31	0.32	.087	.065	.063
ad	-0.63	-0.79	-0.83	.087	.066	.068
ae	-0.17	-0.01	0.03	.087	.066	.068
bc	-0.15	-0.28	-0.27	.087	.074	.071
bd	0.29	0.41	0.44	.087	.074	.085
be	-0.12	-0.25	-0.27	.087	.074	.085
cd	-0.49	-0.36	-0.36	.087	.074	.072
ce	0.05	-0.08	-0.07	.087	.074	.072
de	0.24	0.36	0.38	.087	.074	.084
			Total	1.393	1.139	1.206

4. HOW MANY OUTLIERS?

While this question is less interesting than the main one of the unknown value of β , there are some examples in which it is important to have a fairly clear answer. A central laboratory receiving routine radioimmunoassay readings from a number of medical centres, for example, needs not only to allow for outliers during analysis of the collected data, but also to note which centres are consistently producing relatively large numbers of outliers so that their experimental techniques can be kept up to scratch.

In answering the question we always have to be careful not to compare models with different numbers of parameters since if we do, using improper priors of different dimensionality, the posterior probabilities we obtain will be meaningless. Box and Tiao can safely derive the probabilities we quote in section 2 of 0, 1 or 2 outliers since their model always has $p+1$ parameters, independent of r .

To attempt to do the same for the AB model would be disastrous, however, as this has $p+1$ parameters when $r \neq 0$ but only p when $r=0$. The GDF model carries this problem further as each new outlier adds a new unknown parameter. A naive attempt to apply the formal analysis would merely lead to nearly all the posterior probability being heaped onto the largest number of outliers considered, since it can never do any harm to add more parameters. There is much current discussion about what is a fair penalty to expect a complex model to pay when comparing it with a parsimonious one, but as yet no general agreement. Akaike's (1973) very popular AIC criterion cannot be used here as the likelihood functions of all these models are themselves sums of 2^r or, (for GDF) rC_r terms each of which are products of normal distributions, so that the maximum likelihood estimates needed to evaluate the maximum of the likelihood functions are impossible to find analytically.

We propose here to sidestep this general question by pursuing the GDF model using proper priors throughout. While this automatically removes all doubt about whether the answers are right, it simultaneously introduces the need for a sensitivity analysis to see to what extent those answers depend on the particular priors used.

We first assign prior probability π_r to there being r outliers ($r=0, 1, \dots, n$, $\sum \pi_r = 1$) and refer to this as "model r ". Within this model we look at all rC_r possible partitions of the observations and assign prior probability $\pi_{(r)}$ to a particular subset being the outliers ($\sum_{(r)} \pi_{(r)} = 1$). Conditionally on this we now assign prior densities for the unknown parameters. We take β given σ^2 as p -variate normal with mean b_0 and dispersion matrix $\sigma^2 B_0$ and a given σ^2 as r -variate normal with mean a_0 and dispersion matrix $\sigma^2 A_0$. Finally we take $\nu\nu/\sigma^2$ as chi-square on ν degrees of freedom. We suppress the subscript (r) on the

quantities b_o, B_o, a_o, A_o, ν and ν partly for simplicity but mainly because in practice it is difficult to envisage how these could depend on the particular subset being considered.

Conditional on any given subset being outliers, the posterior for β is Student's t with

$$\hat{\beta}_{(r)} = B_{(r)}^{-1} d_{(r)}$$

$$\nu_{(r)} = n + \nu$$

$$B_{(r)} = B_0^{-1} + \chi'_{(n-r)} \chi_{(n-r)} + \chi'_{(r)} (A_0 + I)^{-1} \chi_{(r)}$$

$$\text{and } w_{(r)} \propto |B_{(r)}|^{-1/2} D_{(r)}^{-(n+\nu)/2} |A_0 + I|^{-1/2} \pi_{(r)}$$

$$\text{where } d_{(r)} = B_0^{-1} b_0 + \chi'_{(n-r)} y_{(n-r)} + \chi'_{(r)} (A_0 + I)^{-1} (y_{(r)} - a_0)$$

$$\text{and } D_{(r)} = b_0' B_0^{-1} b_0 + \nu \nu + y'_{(n-r)} y_{(n-r)} + (y_{(r)} - a_0)' (A_0 + I)^{-1} (y_{(r)} - a_0) - d_{(r)}' B_{(r)}^{-1} d_{(r)}.$$

The posterior mean of β , for example, given model r , is

$$E_r(\beta | y) = \frac{\sum_{(r)} w_{(r)} \hat{\beta}_{(r)}}{\sum_{(r)} w_{(r)}}$$

the sums being over all C_r possible partitions into r and $n - r$ observations.

The prior probability π_r that model r is true is changed into the posterior probability

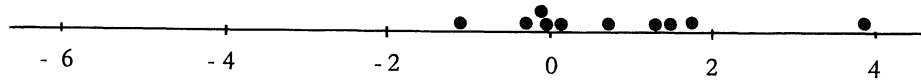
$$\pi'_r \propto \pi_r \sum_{(r)} |B_{(r)}|^{-1/2} D_{(r)}^{-(n+\nu)/2} |A_0 + I|^{-1/2} \pi_{(r)}.$$

We note that the effect of taking very vague, but proper, priors on a within each model is to throw all the posterior weight on $r = 0$, the simplest model, since then $|A_0 + I|^{-1/2}$ decreases geometrically in r . Like many "modern" results this simply rediscovers the work of Jeffreys (1961). The contrast with improper priors which put most posterior weight on the most complicated model is, however, so stark as to be worth mentioning again.

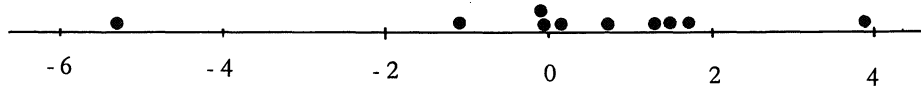
5. SOME TEST DATA

The trouble with proper priors, of course, is actually specifying them. A sensitivity analysis is essential to establish the influence of fairly large changes in the priors on the posterior statements. Darwin's data are not a very suitable set for seeing how well the above results perform as it is not at all clear how many outliers there really are. Accordingly, 10 random observations from $N(0,1)$ were taken. Dataset 1 was formed by adding 4 to one of the

observations, and dataset 2 by further adding -5 to another, see fig. 1. Any self-respecting method ought to be able to get the right answers in such clear-cut cases.



Dataset 1



Dataset 2

FIGURA 1

We assigned equal probability $\frac{1}{4}$ to the number of outliers r being 0, 1, 2 or 3 and assumed that within model r all $\binom{n}{r}$ subsets of r outliers were equally likely. We also took the elements of a to be identically and independently distributed $N(a_0, A_0\sigma^2)$, where a_0 and A_0 are now scalars, the same for each value for r .

Thinking firstly of dataset 1, we might agree that the “right” priors are

$$\beta \sim N(0, \sigma^2), a \sim N(4, A_0\sigma^2), 8\sigma^{-2} \sim \chi^2_{10}.$$

The last of these gives prior mode $= \frac{2}{3}$, mean $= 1$ and variance $\frac{1}{3}$ for σ^2 . We allow A_0 to vary between 10^{-4} and 10^4 since we know that this will crucially affect the answers. These come out to be as in fig. 2(a), that is that we get the clear, correct message that there is one outlier so long as A_0 is not too large. When we use the same priors on dataset 2, however, we get fig 2(b), which completely fails to detect the two outliers. This is hardly surprising since the prior on a is now highly inappropriate.

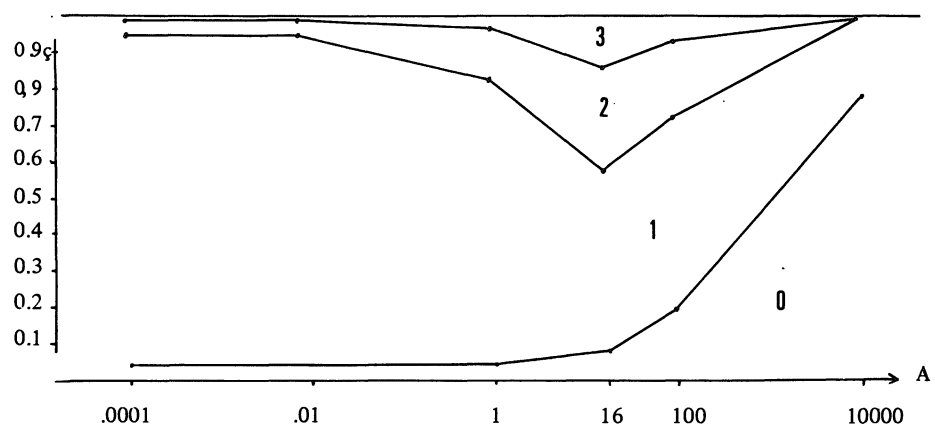


FIGURA 2a

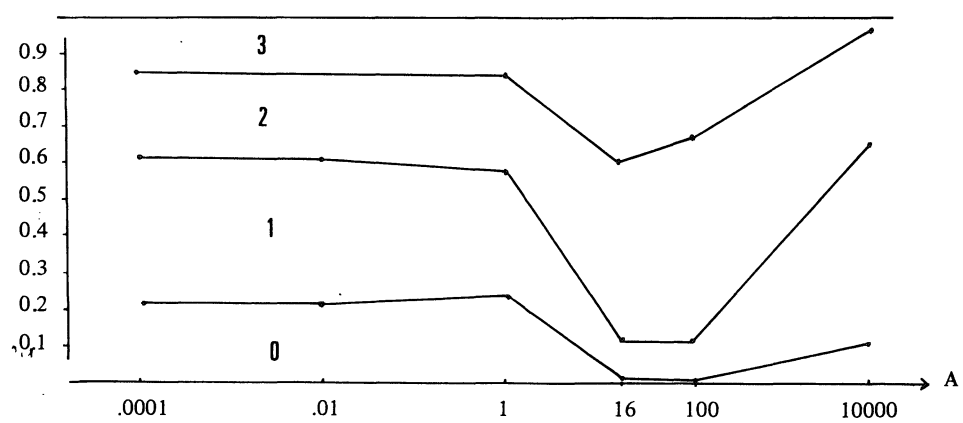


FIGURA 2b

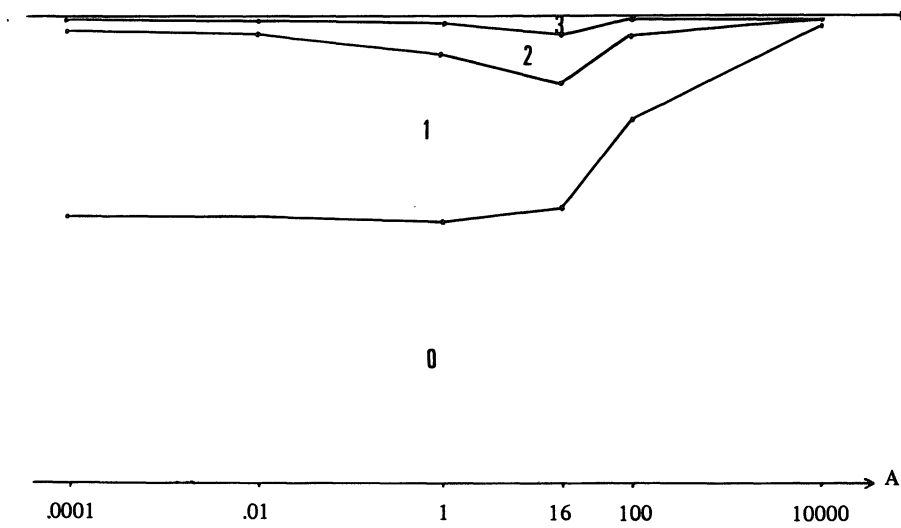


FIGURA 3a

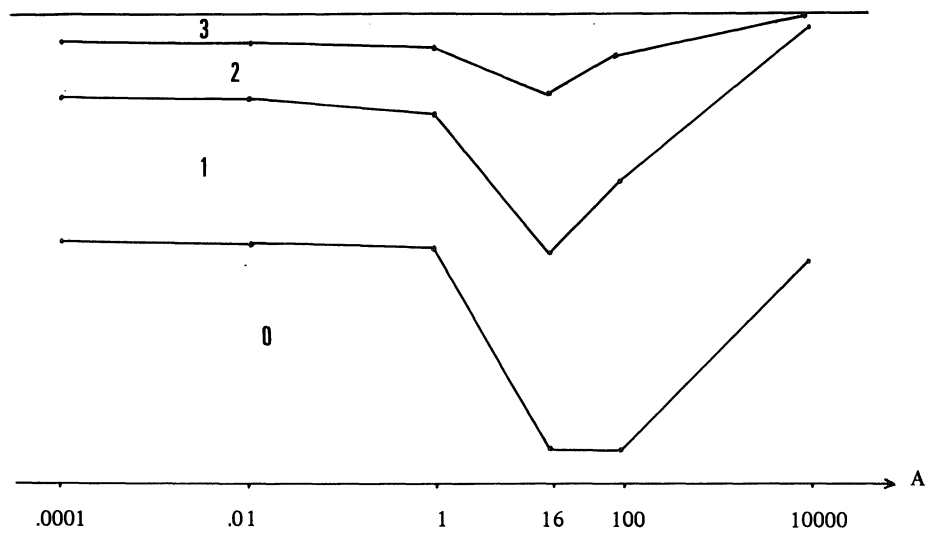


FIGURA 3b

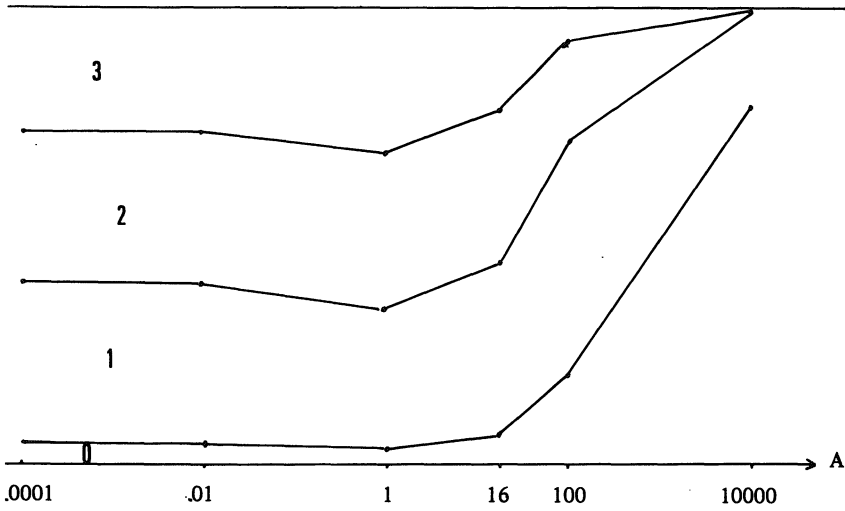


FIGURA 4a

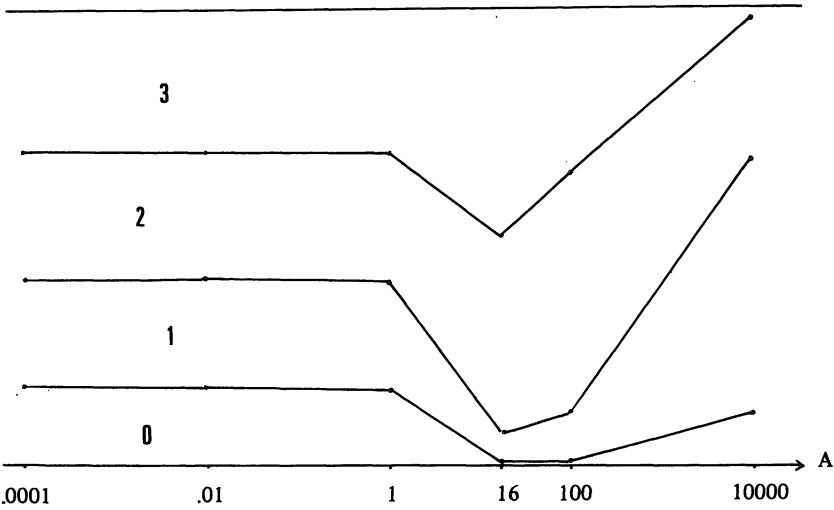


FIGURA 4b

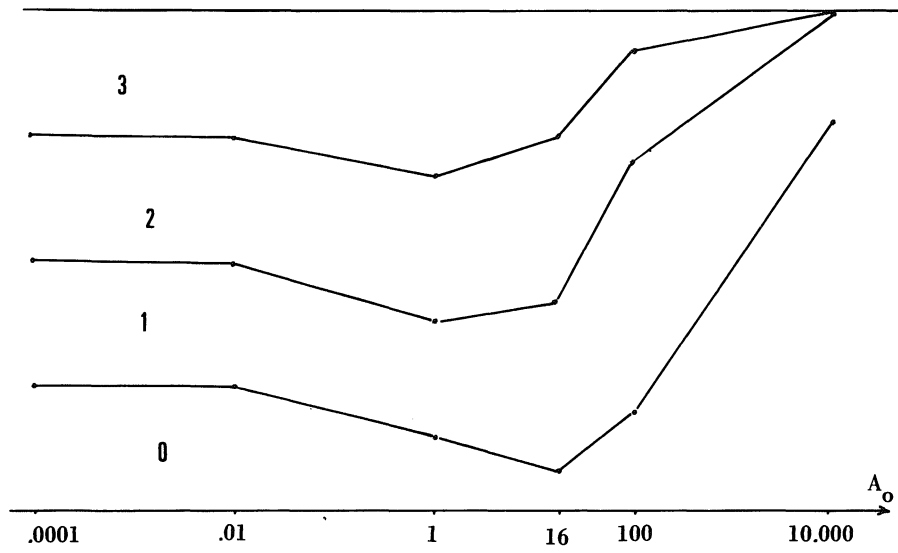


FIGURA 5a

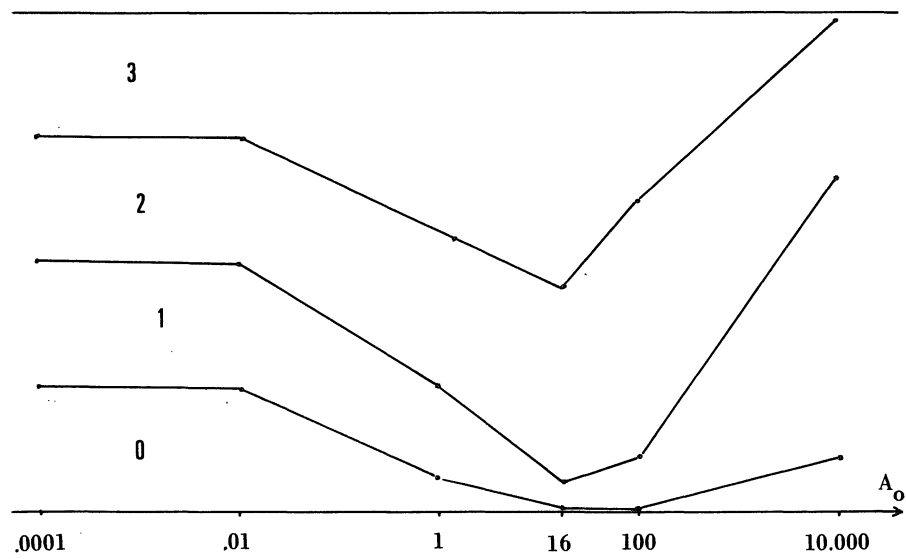
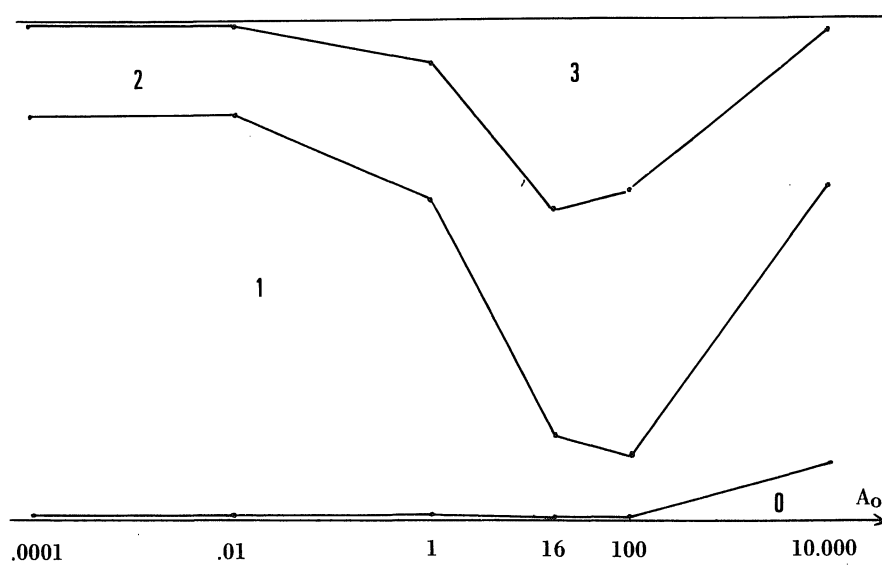
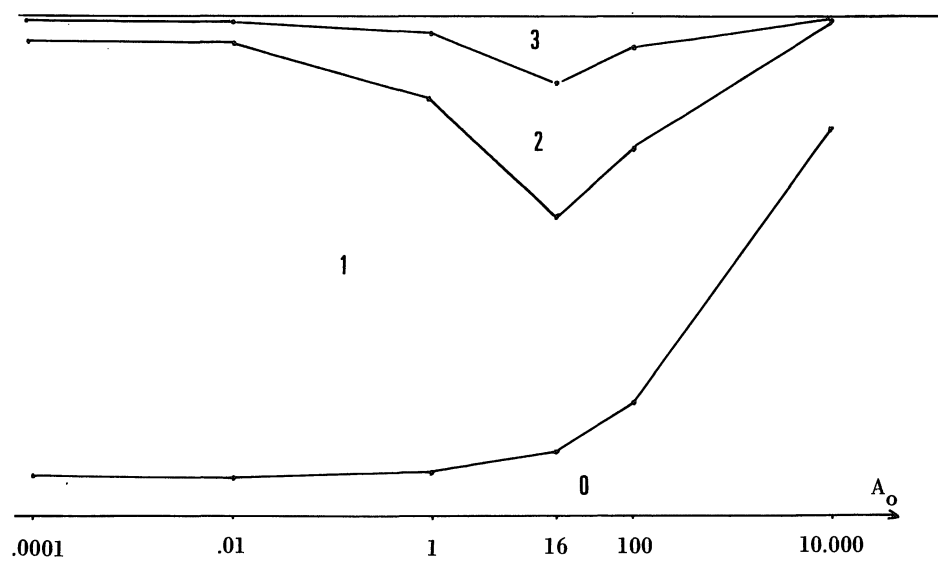


FIGURA 5b



Returning to dataset 1, if we use $N(0, 100\sigma^2)$ for β we get virtually the same result as we do also using $N(5, 100\sigma^2)$, but $N(5, \sigma^2)$ gives fig 3(a) in which we end up quite sure that $r = 0$ or 1 but fail to distinguish clearly between them. Replacing the prior for σ^2 by $2\sigma^{-2} \sim \chi_1^2$, having the same mode of $\frac{2}{3}$ but infinite mean and variance, makes the results slightly less sharp but substantially unaltered. If we get the prior mean of a wrong, though, the results are disastrous. Figs 4(a) and 5(a) show the effects of taking $a_0 = 2$ and $a_0 = 0$ respectively. The latter can be thought of as the closest the GDF model can get to the Box-Tiao philosophy. Not surprisingly, small values of A_0 give posterior probabilities exactly the same as the prior ones. As A_0 increases the probability of one outlier starts to build up but doesn't get near to being decisive before the inevitable slide towards no outlier sets in.

Turning to dataset 2, the corresponding results in figs 2(b), 3(b), 4(b) and 5(b) are all disappointing, especially the last. Taking zero prior mean with a large prior variance for a might have promised to model successfully the occurrence of "two-sided" outliers, but that large prior variance proves its downfall. A preference for two outliers is just starting to show when increasing A_0 pushes the probabilities down towards one and zero outliers. Another hopeful prior might be the mixture $\frac{1}{2}N(4, A_0\sigma^2) + \frac{1}{2}N(-4, A_0\sigma^2)$ but fig 6 shows that while this continues to pick out one outlier successfully, it has no better luck with two than any of its predecessors.

Perhaps this poor performance is not so disgraceful as it seems at first blush. Gentle (1979) reported simulation studies of his proposed frequentist-based outlier detection procedures. For twenty observations with p (the dimension of β) = 2 two outliers were correctly identified only 28% of the time. This rose to 74% for 40 observations and 82% with 60. One hope for our approach, then, might be to increase n in this fashion, but this would immediately create the usual combinatorial explosion and become prohibitively expensive on computer time. By their very nature all three models can only be used with small sample sizes unless a maximum of two outliers is contemplated.

6. DISCUSSION

The GDF model using proper priors can tentatively be claimed to be insensitive to choice of prior on σ^2 and β , so long as a too-precise wrong value of b_0 is not used. It is, however, very sensitive to choice of a_0 and care must be taken not to set A_0 too large. There is also at present no known prior structure that permits large positive and negative contaminations to show themselves simultaneously. On the other hand there is no set of improper priors that would generally be agreed to be appropriate for this problem. Perhaps some of the other papers at this conference will propose a way forward but it might

be that attempts like the AIC criterion to produce a standard way of answering a wide variety of questions regardless of their different contexts are doomed to failure.

Although the question 'How many outliers' may easily be dismissed as an unimportant one, so long as robust inferences about β and σ^2 are possible, I prefer to see it as just one manifestation of the model discrimination problem that is the biggest current challenge to Bayesian statisticians.

REFERENCES

- ABRAHAM, B. and BOX, G.E.P. (1978). Linear models and spurious observations. *Appl. Statist.* **27**, 131-8.
- AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory* (B.N. Petrov and F. Csaki, eds.) 267-281, Budapest, Akademia Kiado.
- BESAG, J. (1979). Exploratory data analysis. *Invited paper to R.S.S. Conference*, Oxford, April 2-6.
- BOX, G.E.P. and TIAO, G.C. (1968). A Bayesian approach to some outlier problems. *Biometrika* **55**, 119-29.
- FISHER, R.A. (1960). *The design of experiments* (7th ed.) Oliver and Boyd: Edinburgh.
- GENTLE J.E. (1978). Testing for outliers in linear regression. In *Contributions to survey sampling and applied statistics* (H.A. David ed.) 223-233. New York: Academic Press.
- GUTTMAN, I., DUTTER, R. and FREEMAN, P.R. (1978). Care and handling of univariate outliers in the general linear model to detect spuriousity - a Bayesian approach. *Technometrics* **20**, 187-193.
- JEFFREYS, H. (1961). *Theory of probability*. Oxford:University Press.
- JOHN, J.A. (1978). Outliers in factorial experiments. *Appl. Statist.* **27**, 111-9.