ELSEVIER Computational Statistics & Data Analysis 33 (2000) 361–380

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

On robust linear regression with incomplete data

A.C. Atkinson *, Tsung-Chi Cheng

Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, UK

Received 1 February 1998; received in revised form 1 April 1999

Abstract

In this paper, we use recently developed methods of very robust regression to extend missing value techniques to data with several outliers. Simulation experiments reveal that additional outliers may be imputed if one ignores the outliers already in the data. The combination of the forward search algorithm for high breakdown point estimators and the EM algorithm or multiple imputation for missing values can avoid problems of this kind. Some multiple deletion diagnostics for linear regression with incomplete data are also discussed.

Keywords: EM algorithm; Forward search algorithm; High breakdown point; Least trimmed squares; Missing values; Multiple imputation; Regression diagnostics; Stalactite plot

1. Introduction

The study of missing values is one of the most important topics in applied statistics, especially in survey problems and medical and biological data. In this paper, we use recently developed methods of very robust regression to extend missing value techniques to data with several outliers. The usual assumption is that missing values are "missing at random" (MAR) (Rubin, 1976; see also Little and Rubin, 1987): the missing-data mechanism does not depend on X_{mis} (the set of missing values) though it may possibly depend on X_{obs} (the set of observed values). If the missingness mechanism does not depend on the parameters of the model, this assumption is called distinct. Moreover, if both MAR and distinctness hold, then the missing-data

* Corresponding author. Tel.: 00-44-171-955-7622; fax: 00-44-171-955-7416.
E-mail address: a.c.atkuson@lse.ac.uk (A.C. Atkinson)

0167-9473/00/$ - see front matter
PII: S 0167-9473(99)00061-4

mechanism is said to be ignorable (Little and Rubin, 1987; Rubin, 1987). Little (1992) suggests that model-based methods, such as maximum likelihood (ML), Bayesian methods and multiple imputation (MI), are best among the current methods for dealing with missing values.

The EM algorithm (Dempster et al., 1977) is an iterative computational method to get a maximum-likelihood estimate when the data can be conveniently viewed as incomplete. It has been widely used to cope with missing data problems. However, it does not provide the variance–covariance estimate of estimated parameters for the linear regression model with incomplete data. There are several approximate methods to get the asymptotic standard errors of ML estimators, for example, scoring or Newton algorithm, bootstrapping the sample, or constructing numerical approximations to the information matrix by the EM computations (see Little, 1992). Beale and Little (1975) gave an approximate formula for estimating the covariance matrix, which has quite stable performance through different missing patterns in a simulation study (Little, 1979). Rubin (1987) proposed multiple imputation, that requires multiply-imputed values for each missing value, resulting in multiple completed data sets. One of the advantages of this method is to avoid underestimation of the true variance.

Now consider the effects of outlying cases. The E-step of the EM algorithm which involves filling in missing values is based on the expected values of the data. Under the normal model of applying multiple imputation, the distribution of missing elements is defined by the multivariate normal linear regression of the missing variables on the observed variables (see Rubin and Schafer, 1990). However both of them are affected by outlying cases. We may therefore impute extra outliers if the existing outliers are ignored. The masking and swamping phenomena are more serious in incomplete data than those without missing values.

For the detection of multiple outliers from linear regression problems without missing data, it essentially needs the high breakdown estimators, such as the least median of squares (LMS) and least trimmed squares (LTS) (see Rousseeuw, 1984; Rousseeuw and Leroy, 1987). A problem with these high breakdown estimators is the lack of efficient algorithms. Several algorithms have been proposed recently (e.g. Atkinson, 1994; Hawkins, 1994; Atkinson and Cheng, 1999). Among these newly developed methods, Atkinson's (1994) forward search algorithm for the LMS is comparatively fast. Atkinson and Cheng (1999) adapt the forward search algorithm for the LTS, which maintains a high breakdown point, to resist the contamination of data, as well as to keep a high efficiency. For the outlier problems in missing values, Shih and Weisberg (1986) extended the distance measure (Cook and Weisberg, 1980) of assessing the influence of the ith case by deleting it from the model for incomplete data. Some quantities of multiple diagnostics will be discussed in this paper. However they are less attractive and limited in the problems of high dimension and large sample size. A procedure using estimators with a high breakdown point is then proposed to detect multiple outliers for the linear regression model with incomplete data. The main idea of the algorithm follows the forward search algorithm, that starts with a randomly selected subset of observations. The observations of the subset are incremented in such a way that outliers are likely to be excluded. If the data are

missing, the EM or MI is applied to estimate parameters which are related to the forward steps. Therefore, it can reveal the outlying cases as well as impute missing values excluding those outliers from the data. Before doing this, we shall give a comparison of estimation between the EM and MI for a linear regression model with missing values in the explanatory variables.

In Section 2, a brief outline of the EM algorithm and multiple imputation is presented, and then a simulation study shows the characteristics of these two methods. In Section 3, we first consider the multiple deletion regression diagnostic for incomplete data, and then propose a robust procedure that combines the forward search algorithm with the EM or MI to detect outliers from the linear regression model with incomplete data. Some examples are used to illustrate the algorithm in Section 4 and comments are given in Section 5.

2. Missing values: Imputation

In this section, we will briefly describe the idea of the EM algorithm and multiple imputation methods, and also give a comparison between them.

2.1. The EM algorithm

The EM algorithm is an iterative method for the computation of the maximiser of the posterior distribution of the observed data. Each iteration of the EM algorithm consists of two steps: expectation (E step) and maximisation (M step). First of all, we briefly review the basic terminology of the algorithm. The basic idea behind the EM algorithm is based on the procedure of augmenting the observed data y by a quantity z, which is referred to as latent data, and then computing and maximising the posterior expectation of $\log(p(\theta|y,z))$. One first computes the expectation of $\log(p(\theta|y,z))$ with respect to the conditional predictive distribution $p(z|y,\theta^{(i)})$, where $\theta^{(i)}$ is the current approximation to the mode of the observed posterior. This is known as the E step. Then, one obtains the maximiser of this conditional expectation at the M step. The conditional predictive distribution is then updated using the new maximiser and the algorithm is iterated until convergence. Specifically, let $\theta^{(i)}$ be the current estimate of the parameter θ. Given the current approximation to the maximiser of the observed estimate $\theta^{(i)}$, the E step of the EM algorithm is defined to compute the expected loglikelihood

$$Q(\theta|\theta^{(i)}) \int_z \log(p(\theta|z,y))\, p(z|\theta^{(i)},y)\, \mathrm{d}z.$$

The M step then consists of maximising this expected function with respect to θ to obtain the update $\theta^{(i+1)}$.

Consider the model,

$$Y = \mathscr{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \beta_0 + \boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{1}$$

where Y is an $n \times 1$ vector of response variables, $\mathscr{X}$ is an $n \times p$ matrix with $p-1$ explanatory variables, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})$ is the vector of regression coefficients, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of errors distributed $\mathrm{N}(\mathbf{0}, \sigma^2 I)$. For the complete data, $\hat{\boldsymbol{\beta}} = (\mathscr{X}^{\mathrm{T}}\mathscr{X})^{-1}\mathscr{X}^{\mathrm{T}}Y$ is the least-squares estimator, which however is spoiled by one or a few outliers. When data are incomplete, some values of the $n \times p$ data matrix $\boldsymbol{Z} = (Y, \boldsymbol{X}) = (z_{ij})$ are "missing at random". One approach to fitting the regression is to maximise the likelihood of an approximating joint distribution for $\boldsymbol{z} = (y, \boldsymbol{x}^{\mathrm{T}})$, where $\boldsymbol{z}$ is often the multinormal distribution, i.e.

$$\boldsymbol{z}_i = \begin{pmatrix} y_i \\ \boldsymbol{x}_i \end{pmatrix} \sim \mathrm{MN}_p \left[\begin{pmatrix} \mu_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{pmatrix} \right]. \tag{2}$$

Given the maximum-likelihood estimates (MLE) of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, the usual methods for obtaining estimates for the conditional distribution of $y|\boldsymbol{x}$ can be used.

Iterative procedures for computing the MLE of $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, using the EM algorithm are as follows. Let $\boldsymbol{\theta}_0$ be a starting value of $\boldsymbol{\theta}$, R_i be the vector of observed variables in the ith case, and σ_{jk} be the j-k entry of $\boldsymbol{\Sigma}$. The E step fills in the data matrix and estimates conditional covariances of the unobserved values given the observed ones. The filled-in values are merely conditional expectations: for each $i = 1, \ldots, n$ and $j = 1, \ldots, p$

$$\begin{aligned} \hat{z}_{ij,t} &= E(z_{ij}|R_i, \boldsymbol{\theta}_t), \quad z_{ij} \text{ not observed}, \\ &= z_{ij}, \qquad\qquad\quad z_{ij} \text{ observed}. \end{aligned} \tag{3}$$

Similarly, the conditional covariance matrix, $\hat{\boldsymbol{C}}_{i,t}$ for case i, given R_i and $\boldsymbol{\theta}_t$ has (j,k)th element

$$\begin{aligned} \hat{c}_{jk,t} &= cov(z_{ij}, z_{ik}|R_i, \boldsymbol{\theta}_t), \quad z_{ij}, z_{ik} \text{ not observed}, \\ &= 0 \qquad\qquad\qquad\quad \text{at least one of } z_{ij}, z_{ik} \text{ observed}. \end{aligned} \tag{4}$$

The M step obtains the estimates of μ and σ by

$$\begin{aligned} \hat{\mu}_{j,t+1} &= \frac{1}{n}\sum_{i=1}^{n} \hat{z}_{ij,t}, \qquad\qquad\qquad\qquad j = 1, \ldots, p, \\ \hat{\sigma}_{jk,t+1} &= \frac{1}{n}\sum_{i=1}^{n} \{(\hat{z}_{ij,t} - \hat{\mu}_{j,t})(\hat{z}_{ik,t} - \hat{\mu}_{k,t}) + \hat{c}_{jk,t}\}, \quad j,k = 1, \ldots, p. \end{aligned} \tag{5}$$

Iterate (3)–(5) until $\boldsymbol{\theta}_t$ converges. At convergence, we denote the fitted matrix by $\hat{\boldsymbol{Z}} = (Y, \hat{\boldsymbol{X}}) = (\hat{z}_{ij})$, the conditional covariance matrix for the ith case by $\hat{C}_i$. The MLEs of the regression parameters $\boldsymbol{\beta}$ and σ^2 can be obtained by the usual transformations:

$$\begin{aligned} \hat{\boldsymbol{\gamma}} &= \hat{\boldsymbol{\Sigma}}_{xx}^{-1}\hat{\boldsymbol{\Sigma}}_{xy}, \\ \hat{\beta}_0 &= \hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\beta}}^{\mathrm{T}}\hat{\boldsymbol{\mu}}_x, \\ \hat{\sigma}^2 &= \hat{\sigma}_y^2 - \hat{\boldsymbol{\Sigma}}_{yx}\hat{\boldsymbol{\Sigma}}_{xx}^{-1}\hat{\boldsymbol{\Sigma}}_{xy}. \end{aligned} \tag{6}$$

From (5) and (6), one can show that the convergent form of the MLE of γ can then be written as

$$\hat{\gamma} = (\hat{X}^{\mathrm{T}}\hat{X} + \hat{C})^{-1}\hat{X}^{\mathrm{T}}Y, \tag{7}$$

where $\hat{C} = \sum_{i=1}^{n} \hat{C}_i$ (Shih and Weisberg, 1986).

Little and Rubin (1987, pp. 141–145) have applied the sweep operator to get the above estimation results, which allows easy implementation. Unfortunately, it does not provide the covariance matrix of the estimated regression coefficients. Little (1979) gave an approximate method for estimating covariance matrix (see also Beale and Little, 1975),

$$A_w = \hat{\sigma}^2 S_w^{-1} = \hat{\sigma}^2(\hat{\mathcal{X}}^{\mathrm{T}}\hat{W}\hat{\mathcal{X}})^{-1}, \tag{8}$$

as an estimate of $Var(\hat{\beta})$ in the incomplete data problem, where $\hat{W}$ is a diagonal matrix with entries as follow:

$$w_i = \begin{cases} 1 & \text{for complete observations,} \\ \hat{\sigma}_y^2/\hat{\sigma}_{yi}^2 & \text{otherwise.} \end{cases}$$

Here $\hat{\sigma}_{yi}^2$ denotes the estimated residual variance of y given the observed part of x for the ith case and $\hat{\sigma}_y^2$ denotes the estimated residual variance of y given all independent variables.

2.2. Multiple imputation

Instead of imputing a single value for each missing value, multiple imputation is a technique designed to handle missing data, which fills in the missing values several times, then creating several completed data sets for analysis (see Rubin, 1987; Rubin and Schenker, 1986; Rubin, 1996; Schafer, 1997). Each data set is analysed separately using techniques designed for complete data, and the results are then combined in such a way that the variability due to imputation may be incorporated. In the notation of Rubin, let Y_{obs} be the set of observed values and Y_{mis} be the set of missing values. Then the posterior density of a population quantity Q can be written as

$$h(Q|Y_{obs}) = \int g(Q \mid Y_{obs}, Y_{mis}) f(Y_{mis} \mid Y_{obs})\, \mathrm{d}Y_{mis},$$

where $f(\cdot)$ is the posterior density of the missing values and $g(\cdot)$ is the complete-data posterior density of θ. Therefore, multiple imputations are simulated draws from the posterior distribution of the missing data.

The values of complete-data statistics $\hat{Q}$ and U calculated on the s completed data sets are $\hat{Q}_{*1}, \ldots, \hat{Q}_{*s}$ and $U_{*1}, \ldots, U_{*s}$. The repeated-imputation estimate is

$$\bar{Q}_s = \sum_{l=1}^{s} Q_{*l}/s, \tag{9}$$

and the associated variance–covariance of $\bar{Q}_s$ is

$$T_s = \bar{U}_s + \frac{s+1}{s} B_s, \tag{10}$$

where

$$\bar{U}_s = \sum_{l=1}^{s} U_{*l}/s = \text{ within-imputation variability} \tag{11}$$

and

$$B_s = \sum_{l=1}^{s} (Q_{*l} - \bar{Q}_s)(Q_{*l} - \bar{Q}_s)^{\mathrm{T}}/(s-1)$$
$$= \text{between-imputation variability.} \tag{12}$$

The large s repeated-imputation inference treats $(Q - \bar{Q}_s)$ as a normal distribution with variance–covariance matrix T_s. Letting $s = \infty$, we have

$$(Q - \bar{Q}_\infty) \sim \mathrm{N}(O, T_\infty),$$

where $T_\infty = \bar{U}_\infty + B_\infty$.

2.3. A simulation study comparing EM and MI

In this subsection, a simulation experiment is carried out to verify the characteristics of the EM algorithm and multiple imputation. Consider a model like (1), with the design matrix $\boldsymbol{X}$ generated from the multivariate normal distribution $\mathrm{MN}(\boldsymbol{O}, \boldsymbol{I}_p)$. All parameters of regression coefficients are assigned to 1, and $\varepsilon_i \sim \mathrm{N}(0,1)$. Once the data are generated, let 10%, 20%, 30% and 40% of the elements of the $\boldsymbol{X}$ matrix be randomly missing. Two kinds of data are generated: sample sizes $n = 100$ and 200 with dimension $p = 4$.

Glynn et al. (1993) give a short simulation study of multiple imputation applied to the linear regression model. They only consider missing values on the dependent variable Y. As mentioned by many authors (see Shih and Weisberg, 1986; Little, 1992), they will not convey any information on the estimation of regression coefficients if dependent variable Y is missing. This kind of situation is therefore excluded from the following studies. The data are firstly converted into the monotone missing pattern (see Little and Rubin, 1987; Schafer, 1997) when applying the EM and MI. The multiple imputation procedure we employed here however is slightly different from that of Schafer, (1997). The imputed values $\hat{\boldsymbol{X}}_I$ of the MI procedure are generated from a normal distribution with mean (3) and covariance (4), that is to say, after convergence of the EM algorithm

$$X_{mis,I} \sim \mathrm{MN}(\hat{X}_I, \hat{C}_I),$$

where I indicates the subset of observations with the same missing pattern. One or more imputed data sets will then be obtained. The usual least-squares regression analysis is carried out for each of the imputed data sets, yielding $\boldsymbol{b}_l$ and $\hat{\sigma}^2_{yl}$. The

multivariate analogous of (9)–(12) are

$$\begin{aligned}
&\boldsymbol{b}_l = (\hat{\mathscr{X}}_l^{\mathrm{T}}\hat{\mathscr{X}}_l)^{-1}\hat{\mathscr{X}}Y, \quad l = 1,\ldots,s,\\
&\bar{\boldsymbol{b}} = \sum_{l=1}^{s} \boldsymbol{b}_l/s,\\
&\boldsymbol{U} = \sum_{l=1}^{s} \hat{\sigma}_{yl}^2(\hat{\mathscr{X}}_l^{\mathrm{T}}\hat{\mathscr{X}}_l)^{-1}/s,\\
&\boldsymbol{B} = \sum_{l=1}^{s} (\boldsymbol{b}_l - \bar{\boldsymbol{b}})(\boldsymbol{b}_l - \bar{\boldsymbol{b}})^{\mathrm{T}}/(s-1),\\
&Var(\bar{\boldsymbol{b}}) = \boldsymbol{U} + \frac{s+1}{s}\boldsymbol{B}.
\end{aligned} \tag{13}$$

Two, three and 10 repeat imputations are considered in the study. Following the simulation study of Little (1979), for each problem 300 data sets are used to calculate the pivotal quantity

$$PQ(\hat{\beta}_j) = (\hat{\beta}_j - \beta_j)/\sqrt{Var(\hat{\beta}_j)}.$$

As large sample theory is applicable, $PQ(\hat{\beta}_j)$ should have approximately a standard normal distribution. Tables 1 and 2 are the mean sums of squares ($MSPQ$) of $PQ(\hat{\beta}_j)$ for sample sizes 100 and 200, respectively. Under the assumption that the pivotal quantities are standard normal deviates, the $MSPQ$ has expected value 1.

Table 1
The mean sums of squares of 300 simulated data sets of sample size $n = 100$ and $p = 4$ containing different proportions of missing values (the values in parentheses are numbers of repeat imputation samplings in multiple imputation)

Proportion of missing values (%)	Methods	Regression coefficients β_1	β_2	β_3	β_0
	EM	1.385	1.133	1.154	1.059
10	MI(2)	1.252	1.090	1.113	1.012
	MI(5)	1.207	1.046	1.037	0.990
	MI(10)	1.220	1.034	1.076	0.955
	EM	1.499	1.479	1.535	1.445
20	MI(2)	1.271	1.404	1.379	1.348
	MI(5)	1.172	1.289	1.326	1.268
	MI(10)	1.236	1.266	1.331	1.234
	EM	1.704	1.791	1.782	1.563
30	MI(2)	1.402	1.581	1.662	1.441
	MI(5)	1.294	1.485	1.523	1.284
	MI(10)	1.305	1.566	1.496	1.315
	EM	2.562	2.573	2.549	2.398
40	MI(2)	2.372	2.441	2.351	2.263
	MI(5)	1.867	2.228	2.265	1.966
	MI(10)	1.780	2.016	2.175	1.871

Table 2
As Table 1 with sample size $n = 200$

Proportion of missing values (%)	Methods	Regression coefficients β_1	β_2	β_3	β_0
	EM	1.239	1.108	1.058	1.211
10	MI(2)	1.152	1.002	0.984	1.180
	MI(5)	1.070	1.005	0.967	1.108
	MI(10)	1.075	1.017	0.976	1.094
	EM	1.328	1.349	1.270	1.241
20	MI(2)	1.288	1.215	1.127	1.150
	MI(5)	1.091	1.113	1.084	1.115
	MI(10)	1.085	1.148	1.097	1.077
	EM	1.784	1.469	1.557	1.570
30	MI(2)	1.464	1.440	1.381	1.499
	MI(5)	1.317	1.297	1.306	1.299
	MI(10)	1.366	1.212	1.259	1.236
	EM	2.349	2.071	2.166	2.253
40	MI(2)	1.813	1.996	2.106	1.992
	MI(5)	1.632	1.646	1.924	1.800
	MI(10)	1.653	1.731	1.710	1.831

From Tables 1 and 2, we see that all the *MSPQ* values from the MI are closer to 1 than those from the EM algorithm. This implies that MI provides a good approximation of the covariance matrix of coefficient estimates. The imputations repeated 5 and 10 times in MI have better results than do only two imputations. When the proportion of missing values is small, 5 times seem to be enough for most cases. However when the dimension and proportion of missing values increase, more imputations lead to better results.

3. Robust procedure for incomplete linear regression

In this section, we are concerned with the detection of outliers from a linear regression model with incomplete data. Firstly, simulated data are used to illustrate the problem of filling in missing values while ignoring existing outliers. Then some aspects of multiple deletion diagnostics adapted from complete data are presented. Finally, a robust procedure is proposed to deal with problems of this kind.

3.1. Additional outliers are imputed when existing outliers are ignored

Table 3 shows simulated complete data from model (1) with sample size 30, dimension 3, and including two outliers. The design matrix of the good data is generated from $\mathrm{MN}(\mathbf{0}, \boldsymbol{I}), \varepsilon \overset{iid}{\sim} \mathrm{N}(0,1)$, and the model is $y = 2 + x_1 + 0.5x_2$, whereas

Table 3
Simulated data with two outliers, cases 29 and 30

Case	X_1	X_2	Y
1	−0.2530425	0.78629518	2.12037182
2	−1.3424740	−0.09442905	0.59034646
3	1.0774149	0.27445057	3.27688575
4	−0.7256962	0.16148840	1.32058346
5	−1.3008882	1.54671621	1.48515952
6	1.5066662	−1.07856095	2.91094041
7	−2.5234580	0.03951394	−0.47175393
8	−0.3100128	0.21172406	1.80243886
9	0.4973584	−1.41039443	1.83068776
10	−0.5289536	−1.32829022	0.79587841
11	0.3795035	0.83916545	2.76672387
12	−1.5931098	−0.62407494	0.09427933
13	0.7042318	1.27642202	3.41215062
14	−1.8497643	−0.81236935	−0.25570947
15	−0.2748396	1.13541174	2.31134653
16	−0.1045025	−0.10942876	1.88417208
17	0.8271669	0.03965216	2.78954291
18	−0.6508313	0.85837823	1.77749276
19	−0.4010624	−0.54199392	1.38752222
20	0.9175199	0.43785310	3.09088349
21	−1.3967365	0.65102905	0.93447423
22	0.7595087	−0.34827709	2.65531683
23	−1.1566393	0.78649443	1.28391695
24	1.5474712	−0.58440167	3.25499225
25	0.5085918	0.58523446	2.83696365
26	−2.1660635	−1.58705235	−1.01987839
27	−0.6296006	−0.98754519	0.80522650
28	1.3586968	−0.09049257	3.25004196
29	4.3663468	4.63854504	0.42494628
30	5.9602356	4.84597874	2.11098456

the bad data are from

$$\begin{pmatrix} y \\ \boldsymbol{x}_i \end{pmatrix} \sim \mathrm{MN}\left(\begin{pmatrix} 2 \\ 5 \\ 5 \end{pmatrix}, \begin{pmatrix} 0.5 & 0.0 & 0.0 \\ & 0.5 & 0.0 \\ & & 0.5 \end{pmatrix} \right).$$

The scatter plots are shown in Fig. 1(a)–(c).

With the forward search algorithm of Atkinson (1994), the algorithm normally starts with a randomly sampled subset with $m = p + 1$ cases from the data. Then the subset is augmented based on the ordered residuals calculated by the current subset, until all cases are included. Therefore outliers are likely to be excluded during the processes of subset augmentation. Using the same idea, the forward search algorithm for the LTS provides a very robust method to detect multiple outliers as well as parameter estimates with a high efficiency for the linear regression model (see Atkinson and Cheng, 1999).

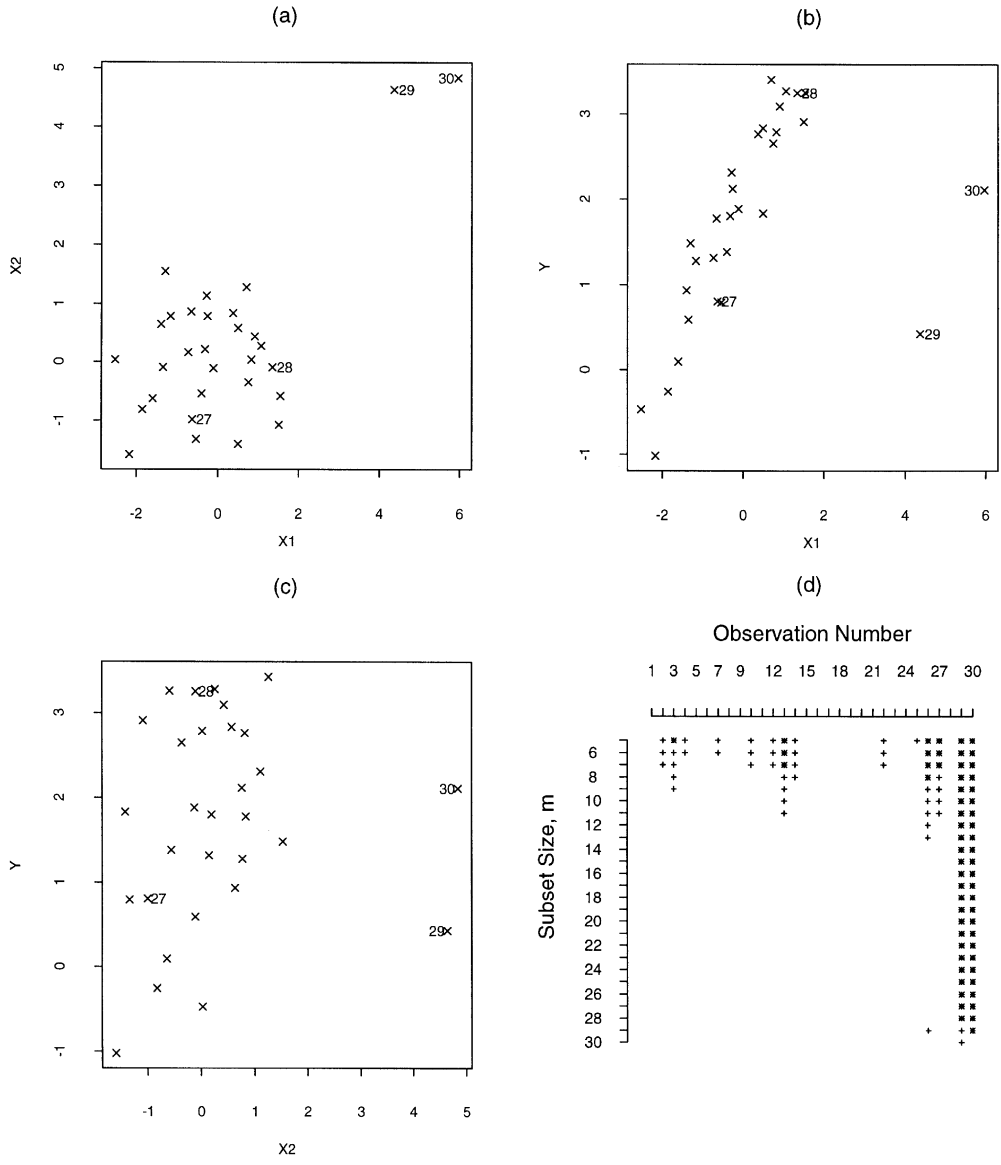

Fig. 1. Simulated data. (a) $X2$ vs. $X1$; (b) Y vs. $X1$; (c) Y vs. $X2$; (d) A stalactite plot using the forward searches for the LTS.

Very robust fitting using LTS fitted to 80% of the data and the forward search algorithm reveal that cases 29 and 30 are extreme outliers. Fig. 1(d) is a stalactite plot, showing those cases identified as outliers during the forward search. 100 random searches for the algorithms were used as suggested in Atkinson (1994). In the plot the values of the scaled absolute studentised residuals greater than 3 are indicated by $*$, and those greater than 2 by $+$. To illustrate the effect of these outliers on missing data procedures, we assume that cases 27 and 28 are missing at random as shown

Table 4
Cases 27 and 28 assumed missing at random (the values in parentheses are values imputed by the EM algorithm)

Case	X_1	X_2	Y
27	−0.6296006	NA (0.194)	0.80522650
28	NA (0.703)	−0.09049257	3.25004196
29	4.3663468	4.63854504	0.42494628
30	5.9602356	4.84597874	2.11098456

in Table 4. Fig. 2(a)–(c) are the scatter plots of the imputed data from the EM algorithm. Although it is not obvious from the scatterplots, the two missing cases 27 and 28 are also identified as outliers when the forward search algorithm is applied with 80% of data fitted in the LTS criterion. The stalactite plot of the best result from 100 searches is shown in Fig. 2(d). This example shows that we may easily impute "bad" values to missing observations if we do not take care of the existing outliers.

3.2. Multiple deletion diagnostics in incomplete data

The approach to missing data problems when outliers are present of Shih and Weisberg (1986) involves the use of regression diagnostics. One way to the regression diagnostics is to identify those cases that give the largest change in a specific aspect of an analysis when one or more cases are removed from data. From now on, we let $\boldsymbol{X} = \mathscr{X}$ to make the notations be the same as discussed in the regression literature. For fitting with complete data, the influence of the ith case can be assessed by the distance measure

$$D_i = (\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})^{\mathrm{T}} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{X} (\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}}) / p\hat{\sigma}^2.$$

Cook and Weisberg (1980) give a detailed discussion of this distance for data without missing values. Shih and Weisberg (1986) developed Cook's distance for the effect of "deleting one case at a time" from incomplete data, which is defined to be

$$D_i(\hat{\boldsymbol{W}}) = (\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})^{\mathrm{T}} \hat{\boldsymbol{X}}^{\mathrm{T}} \hat{\boldsymbol{W}} \hat{\boldsymbol{X}} (\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}}) / p\hat{\sigma}^2,$$

where $\hat{\boldsymbol{\beta}}_{(i)}$ is the coefficient estimate after deleting case i, $i = 1, \ldots, n$, and $\hat{\boldsymbol{W}}$ is given in (8). Observations with the largest values are referred to as (potentially) influential. However, the method may fail to identify the right influential observations if there exist multiple outliers in the complete data set. Multiple deletion diagnostics are then considered. For complete data, the Cook's distance of multiple deletion (see Cook and Weisberg, 1980; Atkinson, 1985, p. 221) is

$$D_I = (\hat{\boldsymbol{\beta}}_{(I)} - \hat{\boldsymbol{\beta}})^{\mathrm{T}} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{X} (\hat{\boldsymbol{\beta}}_{(I)} - \hat{\boldsymbol{\beta}}) / p\hat{\sigma}^2,$$

where $\hat{\boldsymbol{\beta}}_{(I)}$ is the coefficient estimate when the set I of observations is deleted. The analogue of Cook's distance with incomplete data is, in a similar way, defined as

$$D_I(\hat{\boldsymbol{W}}) = (\hat{\boldsymbol{\beta}}_{(I)} - \hat{\boldsymbol{\beta}})^{\mathrm{T}} \hat{\boldsymbol{X}}^{\mathrm{T}} \hat{\boldsymbol{W}} \hat{\boldsymbol{X}} (\hat{\boldsymbol{\beta}}_{(I)} - \hat{\boldsymbol{\beta}}) / p\hat{\sigma}^2. \tag{14}$$

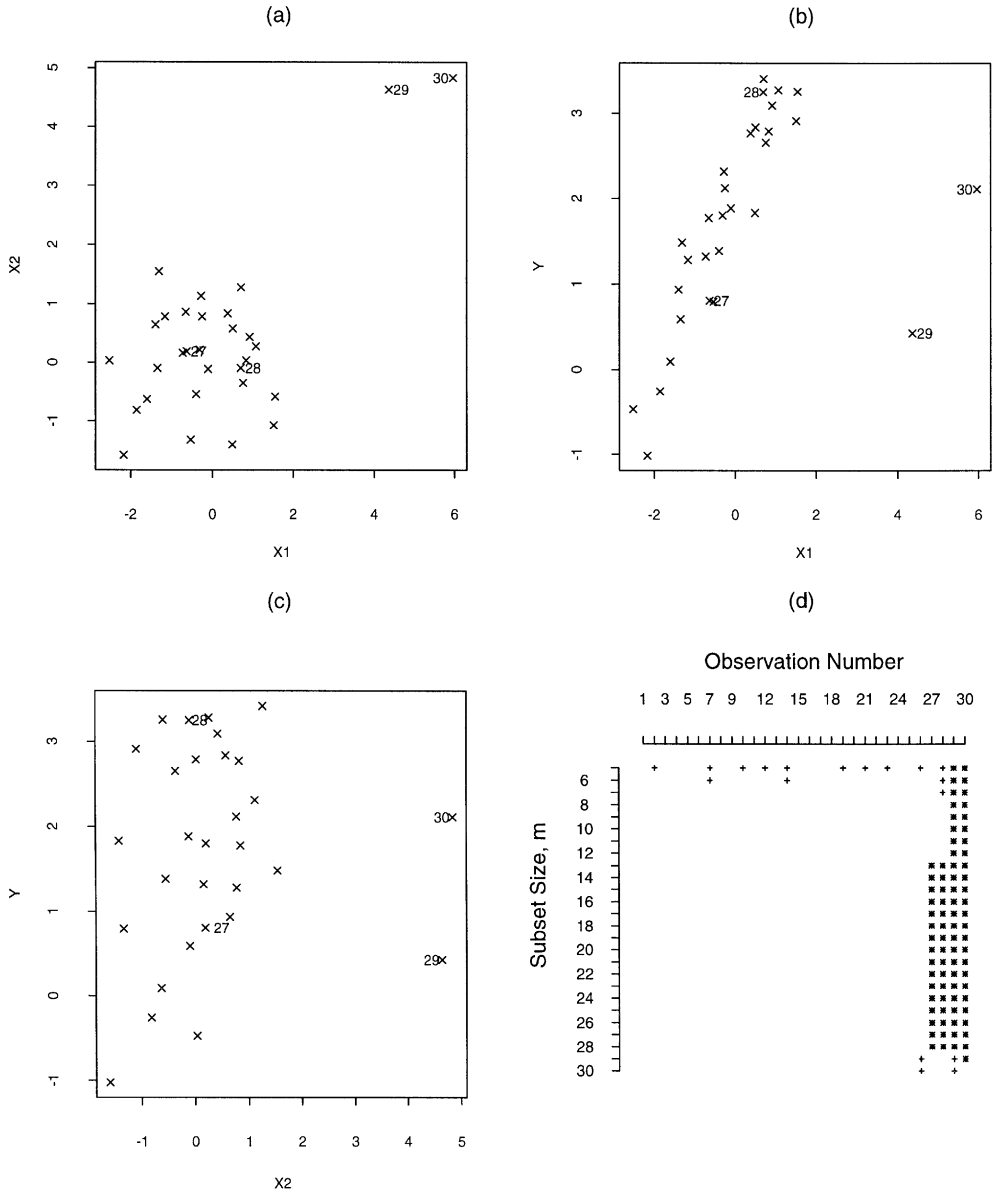

Fig. 2. Extra outliers are imputed if existing outliers are ignored when using the EM. (a) $X2$ vs. $X1$; (b) Y vs. $X1$; (c) Y vs. $X2$; (d) A stalactite plot.

We propose the following quantity:

$$R_I = \frac{\mathrm{RSS} - \mathrm{RSS}_{(I)}}{\mathrm{RSS}}, \tag{15}$$

where RSS is the residual sum of squares of all data, whereas $\mathrm{RSS}_{(I)}$ is that from deleting I observations. If R_I is relatively small or near 0, the error sum of squares is due to random sources rather than the I cases. Conversely, R_I being larger implies

that the I cases have more effect on the error, which indicates these observations are outliers.

A related idea which considers the change in volume of the confidence region for $\boldsymbol{\beta}$ measured by the determinant of $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ is (see Belsley et al., 1980, p. 38)

$$Q_I = \frac{\mathrm{RSS}_{(I)} \cdot \det(\boldsymbol{X}_{(I)}^{\mathrm{T}}\boldsymbol{X}_{(I)})}{\mathrm{RSS} \cdot \det(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})}.$$

Similarly, in incomplete data, we can consider

$$Q_I(\hat{\boldsymbol{W}}) = \frac{\mathrm{RSS}_{(I)} \cdot \det(\hat{\boldsymbol{X}}_{(I)}^{\mathrm{T}}\hat{\boldsymbol{W}}_{(I)}\hat{\boldsymbol{X}}_{(I)})}{\mathrm{RSS} \cdot \det(\hat{\boldsymbol{X}}^{\mathrm{T}}\hat{\boldsymbol{W}}\hat{\boldsymbol{X}})}. \tag{16}$$

All of the above quantities are presented using the approximate covariance of parameter estimates of Little's EM algorithm. If the MI is considered, the only change is that we use (13) instead of the function of $\hat{\boldsymbol{W}}$.

Unfortunately, all the multiple deletion quantities have the same difficulties of being computationally intensive and not easy to summarise or present. The calculation also requires large storage in computer memory, increasingly so with higher dimensions and larger sample sizes. For robust statistics and identification of outliers in complete data, the forward search algorithm provides a feasible tool to find some high breakdown estimators. The results can also be conveniently presented in the stalactite plot (Atkinson, 1994). In the next section, we propose an extension of this procedure to the detection of multiple outliers for incomplete data, that combines the forward search algorithm with EM or MI.

3.3. The forward search algorithm for incomplete data

In the robust approach, we also assume that $\boldsymbol{Z}$ follows a multivariate normal distribution as in (2). The EM algorithm is first applied to impute missing values as in the previous section. Then, the procedure of detection of outliers is carried out from the forward search algorithm for the least trimmed squares estimator. Hence the method considered here, which is a combination of the EM algorithm to fill in missing values and the forward search algorithm to detect outliers. The details are described as follows.

With the forward search algorithm of Atkinson (1994), the algorithm starts with a randomly sampled $m = p + 1$ cases from the data, and a starting value $(\hat{\boldsymbol{\mu}}_m^{(0)}, \hat{\boldsymbol{\Sigma}}_m^{(0)})$ is given. After convergence of the EM algorithm (3)–(5), we have the estimated mean and covariance matrix of the m cases, $(\hat{\boldsymbol{\mu}}_m, \hat{\boldsymbol{\Sigma}}_m)$, imputed values $\hat{\boldsymbol{X}}$ and regression coefficient estimates from (6) or (7),

$$\hat{\boldsymbol{\beta}}(m) = (\hat{\boldsymbol{X}}_m^{\mathrm{T}}\hat{\boldsymbol{X}}_m + \hat{\boldsymbol{C}}_m)^{-1}\hat{\boldsymbol{X}}_m^{\mathrm{T}}Y_m,$$

where $\hat{\boldsymbol{X}}_m$ and Y_m are the design matrix and response variable corresponding to the m observations, and $\hat{\boldsymbol{C}}_m = \sum_{i\in m}\hat{C}_i$. For given m the residuals from this estimate are $e_{i,m} = y_i - \hat{\boldsymbol{x}}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}(m)$ $(i = 1, \ldots, n)$. We then order them

$$e_{(1),m}^2 \leq e_{(2),m}^2 \leq \cdots \leq e_{(n),m}^2. \tag{17}$$

The LTS criterion requires the ordered residuals (17) to obtain the variance estimate

$$\tilde{\sigma}_q^2(m) = \sum_{i=1}^{q} e_{(i),m}^2/(q-p),$$

where the choice of q has been discussed in Atkinson and Cheng (1999). Each forward search involves successively augmenting the subset until $m=n$ and yields a series of values of $\tilde{\sigma}_q^2(m)$, the minimum value $\tilde{\sigma}_{q,j}^2$, defining the performance of the jth search. The overall estimate of σ^2 from the searches is $\tilde{\sigma}_q^2 = \min_j \tilde{\sigma}_{q,j}^2$. The forward search however uses scaled residuals for observations not included in the subset. For the m included observations, the least-squares residuals $e_{i,m}^2$ are used. But, for the $n-m$ observations not included in the fit, the least-squares residuals are scaled by the variance of prediction. Let $\mathscr{M}$ be the subset with m cases. The ordering is thus on the n squared residuals r_i^2 defined by

$$\begin{aligned} r_i^2 &= e_{i,m}^2, \quad i \in \mathscr{M}, \\ r_i^2 &= e_{i,m}^2/(1+d_i), \quad i \notin \mathscr{M}, \end{aligned} \tag{18}$$

with $d_i = \hat{\boldsymbol{x}}_i^{\mathrm{T}}(\hat{\boldsymbol{X}}_m^{\mathrm{T}}\hat{\boldsymbol{X}}_m + \hat{\boldsymbol{C}}_m)^{-1}\hat{\boldsymbol{x}}_i$. The subset size is incremented to $m+1$, based on the rule of ordering of residuals (18). Usually one observation is added in the subset, but sometimes two or more observations are introduced when one or more must leave. Given a new starting point, $(\hat{\boldsymbol{\mu}}_{m+1}^{(0)}, \hat{\boldsymbol{\Sigma}}_{m+1}^{(0)})$, of a new iterative EM algorithm based on the $m+1$ cases, the forward step restarts again to obtain $(\hat{\boldsymbol{\mu}}_{m+1}, \hat{\boldsymbol{\Sigma}}_{m+1})$, the filled-in values and the required estimates. The procedure remains the same as that in Atkinson and Cheng, except that the EM algorithm is used to fill in missing values. The important fact of the forward search algorithm is that outlier cases are likely to be excluded from the forward processes, resulting in reasonable imputed values, based on good data. To calibrate the residuals, the forward procedure is also run on synthetic data with the same structure of explanatory variables $\boldsymbol{X}$ as the data but with Y simulated from a standard normal distribution. To save computational time, the missing values of $\boldsymbol{X}$ may be imputed by the median of each variable. The purposes of this step are to reduce the bias in the estimate σ_q^2 and to scale the residuals used in the detection of outliers as presented in the stalactite plot (Atkinson, 1994). The residuals used for plotting are the scaled studentised residuals

$$t_i = \frac{e_{i,m}\bar{\sigma}_q(m)}{\tilde{\sigma}_q(m)\sqrt{(1-h_i)}}, \quad i \in \mathscr{M},$$

where h_i is the hat matrix diagonal of the same form as d_i but for $i \in \mathscr{M}$, and

$$t_i = \frac{e_{i,m}\bar{\sigma}_q(m)}{\tilde{\sigma}_q(m)\sqrt{(1+d_i)}}, \quad i \notin \mathscr{M},$$

where $\bar{\sigma}_q^2(m)$ is the average of $\tilde{\sigma}_q^2(m)$ for each m, from 100 simulations of forward searches.

If multiple imputation replaces the EM step in the algorithm, the only difference is the estimation of regression coefficients using $\bar{\boldsymbol{b}}$ of (13) instead of (7), when the

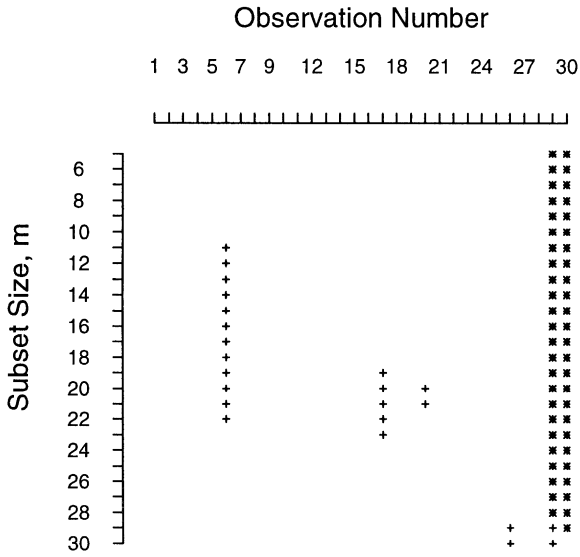

Fig. 3. Simulated data. A stalactite plot of the best solution from 100 searches using the robust procedure described in subsection 3.3.

corresponding residuals and hat matrix are

$$e_{il} = y_i - \hat{\boldsymbol{x}}_{i,l}^{\mathrm{T}} \boldsymbol{b}_l \quad i = 1, \ldots, n \text{ and } l = 1, \ldots, s$$

$$\bar{e}_i = \sum_{i=1}^{s} e_{il}/s,$$

$$d_{il} = \hat{\boldsymbol{x}}_{i,l}^{\mathrm{T}} (\hat{\boldsymbol{X}}_{m,l}^{\mathrm{T}} \hat{\boldsymbol{X}}_{m,l})^{-1} \hat{\boldsymbol{x}}_i,$$

$$\bar{d}_i = \sum_{i=1}^{s} d_{il}/s.$$

4. Examples

Two data sets are used to show the performance of the algorithm.

4.1. Simulated data

We first applied the above combined algorithm to the simulated data in Section 3.1. A hundred forward searches were implemented as usual. Fig. 3 is the best solution among them, which reveals the same outliers as Fig. 1(d). This demonstrates that the algorithm can impute reasonable results as well as detect outliers.

4.2. Clinical trial data

The data were collected from a clinical trial on 34 male patients. There are three explanatory variables, body weight (WT) in kg, serum creatinine (SC) concentration

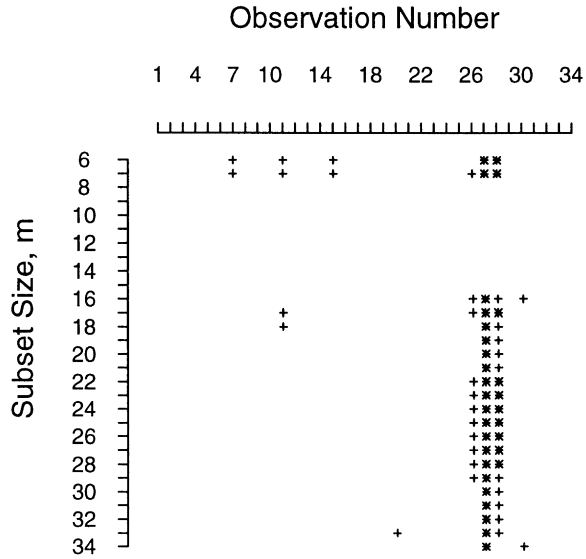

Fig. 4. Clinical trial data. A stalactite plot of the best solution from 100 forward searches.

in mg/deciliter and age in years. The response variable is endogenous creatinine (CR). Shih and Weisberg (1986) used this data set to present their method for assessing influence in multiple linear regression with incomplete data. A typical model recommended in many pharmacokinetics textbooks is (see Shih and Weisberg, 1986),

$$E(\log(\mathrm{CR})) = \beta_0 + \beta_1 \log(\mathrm{WT}) + \beta_2 \log(\mathrm{SC}) + \beta_3 \log(140 - \mathrm{age}).$$

They applied the method of "deleting one case at a time" of Cook's distances in complete data to incomplete data, and found that case 27 is influential and case 30 has a relatively large influence, but not significantly so. Liu (1995) also employed the data which were modified to the monotone missing pattern, to demonstrate a Bayesian imputation method using multivariate t distributions. We here use Liu's modified data to present the algorithm of Section 3.3.

Fig. 4 is the stalactite plot of the best solution of the 100 searches, which shows that cases 26–28 are indicated as outliers using 90% of the data to fit the LTS criterion. Both the EM and MI with five imputations give very similar results, because of the small proportion of missing values. The swamping and masking phenomena can be easily revealed by the stalactite plot. The final step of the forward search includes all cases, so that the estimate is the MLE (7), leading to the same outliers as Shih and Weisberg. Hence the cases 27 and 30 are of interest in these data.

Other proportions of the data can be used in the LTS fit, as Atkinson and Cheng (1999) show. As long as outliers are excluded, the higher the proportion of the data fitted in the LTS, the higher the efficiency of parameter estimates and the more stable the resulting detection of outliers. Different proportions of the data fitted to the LTS are also considered here. For 80% of the data fitted, only two of 100 searches found the optimum and revealed the right outliers, whereas for 90% of the data, the leading

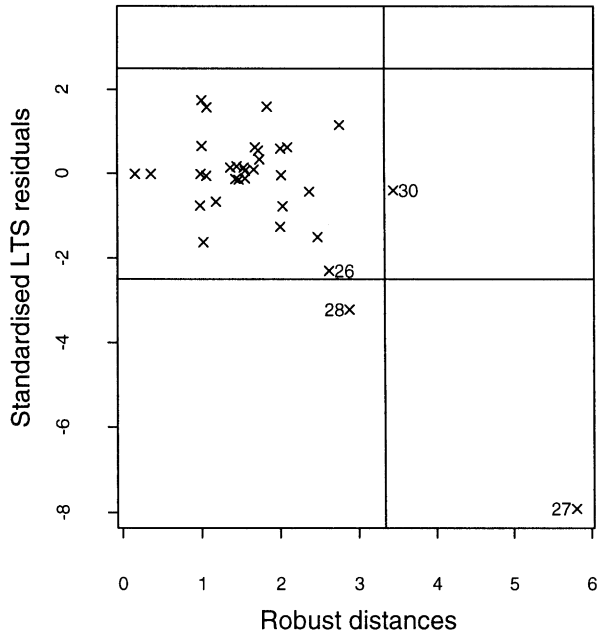

Fig. 5. Diagnostics plot for clinical trial data.

three optima, 5 of the 100 searches, have results as in Fig. 2. Fitting either 80% or 90% of the data leads to quite similar results and reveals that cases 26–28 are outliers. For 70% of data fitted the LTS yields results with more outliers including cases 26–28.

Rousseeuw and van Zomeren (1990) first proposed a plot of Studentised residuals (from the LMS) against the robust distances (from the minimum volume ellipsoid estimator, MVE) of the $\boldsymbol{X}$ matrix, from which the different types of outliers can be classified. The cutoffs values are indicated ± 2.5 and $\sqrt{\chi^2_{p,0.975}}$ by horizontal and vertical lines. Fig. 5 is the plot of Studentised residuals (from the LTS) against the robust distances,

$$(\hat{\boldsymbol{x}}_k - \bar{\boldsymbol{x}}(m))^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_{xx}^{-1}(m)(\hat{\boldsymbol{x}}_k - \bar{\boldsymbol{x}}(m)),$$

where $\bar{\boldsymbol{x}}(m)$ and $\hat{\boldsymbol{\Sigma}}_{xx}(m)$ are the sample mean and sample covariance corresponding to the design matrix at the forward step m, i.e., the subset with m cases. Under good behaviour of the forward search algorithm, they are essentially the minimum covariance determinant (MCD) estimators of multivariate location and shape. We can see cases 27 and 28 are extreme outliers, and case 26 is a mild one, whereas case 30 is an outlier from the $\boldsymbol{X}$ matrix, but it is not a regression outlier. The plot has been suggested independently by Rousseeuw and Van Driessen (1998) and called the distance–distance plot (D–D plot).

We now contrast this robust analysis with the application of the diagnostic quantities discussed in Section 3.2 for these data. Table 5 shows that the values of $D_I(\hat{\boldsymbol{W}})$ are greater than 5.5 for all combinations of deleting three cases. Both R_I and $Q_I(\hat{\boldsymbol{W}})$

Table 5
(Clinical trial data) Multiple deletion diagnostics of cases with $D_I(\hat{W})$ $(I = 3)$ larger than 5.5

Deleted cases			$D_I(\hat{W})$ (14)	R_I (15)	$Q_I(\hat{W})$ (16)
1	27	28	5.784	0.702	0.102
2	27	28	5.669	0.704	0.099
3	27	28	6.416[c]	0.706	0.090
4	27	28	5.918	0.698	0.093
5	27	28	5.971	0.699	0.098
6	27	28	5.759	0.701	0.099
7	27	28	5.860	0.734[c]	0.084[c]
8	27	28	5.562	0.717	0.097
9	27	28	5.800	0.702	0.106
10	27	28	5.529	0.718	0.093
11	27	28	6.044	0.728	0.095
12	27	28	5.808	0.702	0.107
13	27	28	5.976	0.700	0.100
14	27	28	5.764	0.702	0.103
15	27	28	6.021	0.720	0.097
16	27	28	7.688[a]	0.737	0.076[b]
17	27	28	5.636	0.710	0.101
18	27	28	5.707	0.707	0.097
21	27	28	5.799	0.723	0.090
24	27	28	5.591	0.702	0.096
26	27	28	6.725[b]	0.766[a]	0.072[a]
27	28	29	5.725	0.703	0.113
27	28	30	6.392	0.739[b]	0.085
27	28	31	5.614	0.696	0.110
27	28	32	5.926	0.708	0.109
27	28	33	5.701	0.703	0.112
27	28	34	5.630	0.697	0.109

[a]Indicates the value with the largest departure.
[b]The second largest.
[c]The third largest.

show that deletion of cases 26, 27 and 28 causes the largest change. In particular, all the combinations including deletion of cases 27 and 28 have large departures. Because of the very significant effect of case 27 (see the results of Shih and Weisberg), Table 6 shows the quantities when deleting two cases from the data without case 27. The deletion of cases 26 and 28 has the largest departures of all three quantities.

The different conclusion reached by Shih and Weisberg is due to the single deletion diagnostic which has a masking effect.

5. Comments

The imputation method plays an important role in the problems of missing values. The current imputation methods for missing values tend to move cases to the centre of the data since the imputed values are conditional expectations. Such methods may

Table 6
(Clinical trial data without case 27) Some multiple deletion diagnostics of cases with $D_I(\hat{\boldsymbol{W}})$ $(I = 2)$ larger than 0.45

Deleted cases		$D_I(\hat{\boldsymbol{W}})$ (14)	R_I (15)	$Q_I(\hat{\boldsymbol{W}})$ (16)
3	28	0.541	0.252	0.471
7	28	0.585	0.323	0.438
11	28	0.489	0.310	0.493
11	30	0.488	0.190	0.628
15	28	0.469	0.289	0.503
15	30	0.523	0.177	0.627
16	28	0.904[c]	0.330	0.397[b]
16	30	0.518	0.176	0.544
17	20	0.527	0.270	0.519
18	20	0.516	0.257	0.506
18	30	0.504	0.169	0.609
19	20	0.772	0.343[b]	0.441
20	24	0.598	0.250	0.497
21	30	1.015[b]	0.220	0.536
25	28	0.465	0.330	0.429[c]
26	28	1.141[a]	0.405[a]	0.373[a]
26	30	0.704	0.222	0.550
28	30	0.903	0.338[c]	0.443

[a]Indicates the value with the largest departure.
[b]The second largest.
[c]The third largest.

impute extra outliers if existing outliers are ignored. Therefore, the swamping and masking effects are more serious in incomplete data. Our studies show that the combination of the forward search algorithm with the EM algorithm or MI is successful in detecting outliers from linear regression models with an appreciable proportion of missing values and also provides a very robust estimation procedure. Multiple deletion diagnostics also provide some useful information on potential outliers, despite the computational problem.

It is well known that high breakdown point estimators may have low efficiency. Moreover, missing values may further reduce the efficiency of the estimates. It is worth determining how much efficiency has been lost by the algorithm. We hope that our current large-scale simulation study will clarify these problems.

References

Atkinson, A.C., 1985. Plots, Transformations and Regression. Oxford University Press, Oxford.

Atkinson, A.C., 1994. Fast very robust methods for the detection of multiple outliers. J. Amer. Statist. Assoc. 89, 1329–1339.

Atkinson, A.C., Cheng, T.-C., 1999. Computing least trimmed squares regression with the forward search. Statist. Comput. 9, 251–263.

Beale, E.M.L., Little, R.J.A., 1975. Missing values in multivariate analysis. J. Roy. Statist. Soc. Ser. B 37, 129–146.

Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression Diagnostics. Wiley, New York.
Cook, R.D., Weisberg, S., 1980. Characterizations of an empirical influence function for detecting influential cases in regression. Technometrics 22, 495–508.
Dempster, A.P., Laird, M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Statist. Soc. Ser. B 39, 1–38.
Glynn, R.J., Laird, N.M., Rubin, D.B., 1993. Multiple imputation in mixture models for nonignorable nonresponse with follow-ups. J. Amer. Statist. Assoc. 88, 984–993.
Hawkins, D.M., 1994. The feasible solution algorithm for least trimmed squares regression. Comput. Statist. Data Anal. 17, 185–196.
Little, R.J.A., 1979. Maximum likelihood inference for multiple regression with missing values: a simulation study. J. Roy. Statist. Soc. Ser. B 44, 226–233.
Little, R.J.A., 1992. Regression with missing X's: a review. J. Amer. Statist. Assoc. 87, 1227–1237.
Little, R.J.A., Rubin, D.B., 1987. Statistical Analysis with Missing Data. Wiley, New York.
Liu, C.H., 1995. Missing data imputation using the multivariate $\boldsymbol{t}$ distribution. J Multivariate Anal. 53, 139–158.
Rousseeuw, P.J., 1984. Least median of squares regression. J Amer. Statist. Assoc. 79, 871–880.
Rousseeuw, P.J., Leroy, A.M., 1987. Robust regression and outlier detection. John Wiley, New York.
Rousseeuw, P.J., Van Driessen, K., 1998. A fast algorithm for the minimum covariance determinant estimator. Technical Report, University of Antwerp.
Rousseeuw, P.J., van Zomeren, B.C., 1990. Unmasking multivariate outliers and leverage points (with discussion), J. Amer. Statist. Assoc. 85, 633–651.
Rubin, D.B., 1976. Inference and missing data. Biometrika 63, 581–592.
Rubin, D.B., 1987. Multiple Imputations for Nonresponse in Surveys. Wiley, New York.
Rubin, D.B., 1996. Multiple imputation after 18+ years. J. Amer. Statist. Assoc. 91, 473–489.
Rubin, D.B., Schafer, J.L., 1990. Efficiently creating multiple imputations for incomplete multivariate normal data in Proceedings of the Survey Research Methods Section, American Statistical Association, pp. 83–88.
Rubin, D.B., Schenker, N., 1986. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. J. Amer. Statist. Assoc. 81, 366–374.
Schafer, J.L., 1997. Analysis of Incomplete Multivariate Data. Chapman & Hall, London.
Shih, W.J., Weisberg, S., 1986. Assessing influence in multiple linear regression with incomplete data. Technometrics 28, 231–239.