Mimeo Series #1906

A COMPARISON OF SOME CLASSICAL APPROACHES TO OUTLIER DETECTION IN LINEAR REGRESSION AND AN APPROACH BASED ON ADAPTIVELY-ORDERED RECURSIVE RESIDUALS

Farid Kianifard and William H. Swallow

Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203

ABSTRACT

Five procedures for detecting outliers in linear regression are compared: sequential testing of the maximum studentized residual, Marasinghe's (1985) multistage procedure, and three procedures based on recursive residuals, calculated on adaptively-ordered observations. Four of these procedures test a no-outliers hypothesis. This hypothesis requires a rather stringent test criterion, meaning that moderate outliers often go undetected. The fifth procedure, the basic recursive-residual method, tests individual observations to identify outliers. This procedure is effective in identifying modest outliers, while by construction reducing masking of one outlier by another. When the data analyst's goal is to identify outliers/influential observations for further scrutiny and perhaps special treatment, not merely to justify their deletion, this procedure has advantages over the others.

KEY WORDS: Influential observations, Marasinghe's multistage procedure, Maximum studentized residual, Regression diagnostics.

1. INTRODUCTION

Data collected by research workers commonly contain outliers, and it is important that these outliers be identified in the course of a thorough and correct statistical analysis. A number of procedures have been proposed in recent years for detecting outliers in linear regression, yet their detection still may be difficult, especially when there are multiple outliers in the data. Two common problems in multiple-outlier detection are masking and swamping. Masking occurs when one outlier obscures the existence of another, swamping when a non-outlier is wrongly included in a group of observations thought to be outliers. A comprehensive text on the study of outliers is that of Barnett and Lewis (1984). A recent review article by Chatterjee and Hadi (1986) describes many of the well-known outlier-detection procedures and model diagnostics and their interrelationships in the context of linear regression.

Along with the proliferation of outlier-detection methodologies over the years, philosophical changes have occurred. In the early years, many statisticians and practitioners viewed outlier-identification methodologies largely as ways to legitimize deleting observations which, though not necessarily erroneous, fell outside the pattern seen in the bulk of the data and were perhaps troublesome in the analysis. Nowadays, outlier identification is viewed more broadly. It is widely recognized that, in some applications, outliers are of interest in their own right and may be the most important observations in the data set; identifying them may help chart future research. And the literature on influential observations has expanded our understanding of the need to identify certain points as candidates for special treatment (perhaps

downweighting or deletion), lest they warp our impression of relationships in the body of the data. The focus has moved away from viewing these procedures as providing support for automatic deletion of points, and toward seeing them as aids in identifying points for more careful scrutiny. The latter perspective suggests interest in identifying moderate as well as extreme outliers.

Data analysts routinely encounter data sets which potentially contain one or more outliers. When, as is usually the case, there is no *a priori* reason to suspect that particular observations are the outliers, an outlier test based on the sequential (perhaps better called "repeated") application of a single-outlier test statistic is commonly used. This sequential approach is especially prone to masking in the presence of multiple outliers.

Gentleman and Wilk (1975) presented a method for detecting the "k most likely outlier subset" in a linear model, based on comparing the effects of deleting each possible subset of k observations in turn. Drawbacks to their procedure are (i) the need to specify k in advance, and (ii) the substantial computational effort involved. Marasinghe (1985) suggested a simpler multistage procedure for identifying up to k outliers in regression based on studentized residuals.

Kianifard and Swallow (1987) introduced a procedure for detecting outliers in linear regression based on recursive residuals, calculated from adaptively-ordered observations. In this article, we compare the performance of the sequential method, Marasinghe's multistage procedure, and several variations of Kianifard and Swallow's "recursive method" in a simulation study. The methods are described in Sections 2 and 3, and compared in Section 4.

We compare the three methods under the familiar linear regression model:

$$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \ , \tag{1.1}$$

where $\underset{\sim}{Y} = (y_1,\ldots,y_n)'$ is an $n \times 1$ vector of values of the response variable, $\underset{\sim}{\beta} = (\beta_1,\ldots,\beta_p)'$ is a $p \times 1$ vector of unknown parameters, $\underset{\sim}{X} = (\underset{\sim}{x}_1',\ldots,\underset{\sim}{x}_n')'$ is an $n \times p$ matrix of explanatory variables with $\text{rank}(\underset{\sim}{X}) = p$ and $\underset{\sim}{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_n)'$ is an $n \times 1$ vector of independent normal random variables with mean 0 and (unknown) variance σ^2. For $\hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}$ being the ordinary least squares estimator of $\underset{\sim}{\beta}$, the residual vector can be written:

$$\underset{\sim}{e} = \underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}} = (\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{Y} \ , \tag{1.2}$$

where $\underset{\sim}{H} = (h_{ij}) = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'$. The residual mean square estimate of σ^2 is then

$$s^2 = \frac{\underset{\sim}{e}'\underset{\sim}{e}}{(n-p)} \tag{1.3}$$

The e_i have several deficiencies which limit their usefulness in outlier detection: they have unequal variance, are correlated, and have a joint distribution which depends on $\underset{\sim}{X}$.

A scaled version of the e_i can be defined as

$$e_{si} = \frac{e_i}{s(1-h_{ii})^{\frac{1}{2}}} \ , \qquad i = 1,\ldots,n \ . \tag{1.4}$$

The e_{si} are usually, but not always, called (internally) studentized residuals. A trivial order-preserving transformation of the studentized residuals gives the so-called adjusted residuals:

$$a_i = \frac{e_i}{(1-h_{ii})^{\frac{1}{2}}}, \qquad i = 1,\ldots,n . \tag{1.5}$$

Because the adjusted residuals are not scaled by any estimate of σ, they are seldom used. They do appear in (2.1) below, however, and are therefore included here.

Another scaled version of the e_i, advocated by Belsley, Kuh, and Welsch (1980) and often called the jackknifed or externally studentized residuals, is

$$t_i = \frac{e_i}{s_{(i)}(1-h_{ii})^{\frac{1}{2}}}, \qquad i = 1,\ldots,n , \tag{1.6}$$

where $s^2_{(i)}$ is the residual mean square estimate of σ^2 obtained with the i^{th} observation omitted. Atkinson (1981) calls t_i the "cross-validatory" residual and, using the relation $t_i = e_{si}\{(n-p-1)/(n-p-e^2_{si})\}$, notes that the t_i are just a monotone transformation of the e_{si}. While the above scaled versions of the residuals e_i have approximately constant variance, they are still correlated. Evidence that a particular observation is an outlier will not appear solely in its own e_i or e_{si} or t_i, but will, to some extent, be smeared over all of them. This lessens their usefulness in outlier detection.

2. THE SEQUENTIAL AND MULTISTAGE PROCEDURES

The sequential method uses $R_n = \max|e_{si}|$ to test the no-outliers hypothesis at each stage. If R_n exceeds the critical value, the i^{th} observation is judged to be an outlier. It is then removed from the data set and the procedure is repeated on the remaining n - 1

observations. This process is repeated until $\max|e_{s_i}|$ fails to exceed the critical value. Lund (1975) and Prescott (1975) employed Bonferroni inequalities to obtain upper bounds to critical values. Tietjen, Moore, and Beckman (1973) gave simulation-based point estimates of the critical values for simple linear regression ($p = 2$).

The multistage procedure of Marasinghe (1985) collects a subset of prespecified size k as follows: Fit the regression model to the complete data and include the observation with $\max|e_{s_i}|$ as the first one in the subset. This observation is then deleted and the model is fitted again to the remaining $n - 1$ observations. As before, the observation with the largest absolute studentized residual from this fit is added to the subset. This procedure is continued until the subset of size k observations is determined.

The reduction in the residual sum of squares resulting from fitting the same model after deleting k observations (numbered $i_1,\ldots,i_k$) is denoted by Q_k and can be computed as (Gentleman and Wilk 1975)

$$Q_k = \sum_{r=1}^{k} a_{i_r}^2 , \tag{2.1}$$

where a_{i_r} is the adjusted residual of (1.5). Marasinghe's test statistic is defined as

$$F_k = \frac{(\underset{\sim}{e}'\underset{\sim}{e} - Q_k)}{\underset{\sim}{e}'\underset{\sim}{e}} . \tag{2.2}$$

The no-outlier hypothesis is rejected in favor of the alternative hypothesis of at most k outliers when F_k is smaller than a specified critical value. If the hypothesis of no-outliers is rejected, the observation corresponding to $\max|e_{s_i}|$ (i.e., the first observation added to the subset) is removed and the test statistic is recomputed with

both n and k reduced by 1. The number of outliers reported cannot exceed k, and is determined when a null hypothesis fails to be rejected. Estimated percentage points of F_k for k = 2,3,4, and 5 are given by Marasinghe (1985) for p = 2. He recommends using a slight overestimate for k to minimize possible masking effects by non-outlying observations. Fung (1987), however, gives two examples to show that, contrary to intuition, taking k too large may cause fewer outliers to be detected. Thus, the choice of the starting value for k may at times be more critical than Marasinghe's results suggested.

3. THE RECURSIVE METHOD

Consider the regression model (1.1) with independent identically-distributed (iid) normal errors $\underset{\sim}{\varepsilon}$. Let $\underset{\sim}{X}_{j-1}$ denote the (j-1) × p matrix consisting of the first j - 1 rows (observations) of $\underset{\sim}{X}$. Provided (j-1) $\geq$ p and assuming $(\underset{\sim}{X}'_{j-1}\underset{\sim}{X}_{j-1})$ to be nonsingular, $\underset{\sim}{\beta}$ can be estimated by:

$$\hat{\underset{\sim}{\beta}}_{j-1} = (\underset{\sim}{X}'_{j-1}\underset{\sim}{X}_{j-1})^{-1}\underset{\sim}{X}'_{j-1}\underset{\sim}{Y}_{j-1} , \qquad (3.1)$$

where $\underset{\sim}{Y}_{j-1}$ denotes the subvector consisting of the first j - 1 elements of $\underset{\sim}{Y}$. Using $\underset{\sim}{\beta}_{j-1}$, one may "forecast" y_j. The forecast error is $(y_j - \underset{\sim}{x}'_j\hat{\underset{\sim}{\beta}}_{j-1})$ and the variance of the forecast error is $\sigma^2[1 + \underset{\sim}{x}'_j(\underset{\sim}{X}'_{j-1}\underset{\sim}{X}_{j-1})^{-1}\underset{\sim}{x}_j]$. The n - p recursive residuals are defined as (Brown, Durbin, and Evans 1975):

$$w_j = \frac{(y_j - \underset{\sim}{x}'_j\hat{\underset{\sim}{\beta}}_{j-1})}{\{1 + \underset{\sim}{x}'_j(\underset{\sim}{X}'_{j-1}\underset{\sim}{X}_{j-1})^{-1}\underset{\sim}{x}_j\}^{\frac{1}{2}}} \qquad j = p + 1,\ldots,n . \qquad (3.2)$$

Recursive residuals cannot be calculated for the first p observations. Under the model, $w_{p+1},\ldots,w_n$ are independent $N(0,\sigma^2)$. BLUS residuals

(Theil 1971) are also independent $N(0,\sigma^2)$ under the model. But, unlike the BLUS residuals, the recursive residuals are in one-to-one correspondence with the n - p observations for which they are calculated, suggesting their greater usefulness in outlier detection. Hedayat and Robson (1970) defined the recursive residuals in an alternative form and called them "stepwise residuals".

The w_j could be calculated using the conventional least squares formula repeatedly to compute each $\hat{\underset{\sim}{\beta}}$ vector in the sequence $\hat{\underset{\sim}{\beta}}_p,\ldots,\hat{\underset{\sim}{\beta}}_{n-1}$. However, the computations are made much more efficient using the following updating formulae (Plackett 1950, Phillips and Harvey 1974, Brown et al. 1975):

$$\hat{\underset{\sim}{\beta}}_j = \hat{\underset{\sim}{\beta}}_{j-1} + \frac{(\underset{\sim}{X}_{j-1}'\underset{\sim}{X}_{j-1})^{-1}\underset{\sim}{x}_j(y_j - \underset{\sim}{x}_j'\hat{\underset{\sim}{\beta}}_{j-1})}{1 + \underset{\sim}{x}_j'(\underset{\sim}{X}_{j-1}'\underset{\sim}{X}_{j-1})^{-1}\underset{\sim}{x}_j} \tag{3.3}$$

$$(\underset{\sim}{X}_j'\underset{\sim}{X}_j)^{-1} = (\underset{\sim}{X}_{j-1}'\underset{\sim}{X}_{j-1})^{-1} - \frac{(\underset{\sim}{X}_{j-1}'\underset{\sim}{X}_{j-1})^{-1}\underset{\sim}{x}_j\underset{\sim}{x}_j'(\underset{\sim}{X}_{j-1}'\underset{\sim}{X}_{j-1})^{-1}}{1 + \underset{\sim}{x}_j'(\underset{\sim}{X}_{j-1}'\underset{\sim}{X}_{j-1})^{-1}\underset{\sim}{x}_j} \tag{3.4}$$

$$S_j = S_{j-1} + w_j^2 \quad . \tag{3.5}$$

Given that the w_j are iid $N(0,\sigma^2)$ random variables, if one were to estimate σ^2 by $\hat{\sigma}_j^2$, an estimate independent of w_j, then $w_j/\hat{\sigma}_j$ would have an exact t distribution with the degrees of freedom (df) of $\hat{\sigma}_j^2$. The null hypothesis that the j^{th} observation is not an outlier could be tested by comparing each $w_j/\hat{\sigma}_j$ to percentiles of the appropriate t distribution, and rejecting the null hypothesis whenever $|w_j/\hat{\sigma}_j|$ was sufficiently large.

However, for the procedure described below, Kianifard and Swallow (1987) recommend instead using $s^2_{(i)}$ [see (1.6)] as the estimate of σ^2 in scaling the recursive residuals. $s_{(i)}$ is not independent of w_j, but is robust to gross errors in the i^{th} observation, and behaves well when the observations are adaptively ordered prior to the calculation of the recursive residuals, as we recommend below. The scaled recursive residuals $w_j/s_{(i)}$ will have approximate t distributions under the null hypotheses.

We propose the following strategy for labelling or ordering the observations, and calculating recursive residuals and test statistics for the null hypotheses that the j^{th} observation is not an outlier, $j = p + 1,\ldots,n$ (Kianifard and Swallow 1987):

1) Fit the regression model to the data.
2) Compute values of an appropriate regression diagnostic (e.g., the studentized residual) for each of the n observations.
3) Order the observations according to the chosen regression diagnostic.
4) Use the first p observations in the ordered data set to form the "basis" for computing recursive residuals.
5) Compute the recursive residuals, w_j, for the remaining n - p ordered observations.
6) Calculate the statistics $w_j/s_{(i)}$, $j = p + 1,\ldots,n$, comparing the computed values against values of Student's t with n - p - 1 df.

We call this procedure the "recursive method". When an appropriate diagnostic measure is used to order the observations, outliers and/or influential observations can be expected to appear late

in the sequence of recursive residuals. The w_j for data points which precede them in the ordered sequence, by construction, will not be affected by these outliers and/or influential observations.

In the simulation study summarized in Section 4, we considered three diagnostic measures according to which the observations could be ordered, arranging them in ascending order of $|e_{s_i}|$ or Cook's influence measure D_i (Cook 1977) or in descending order of $COVRATIO_i$, (Belsley et al. 1980, p. 22). The studentized residual and Cook's D are widely available in packaged programs; COVRATIO is less available, but represents a different class of diagnostics. Other diagnostic measures (some of which are monotone transformations of the above three) could also be used for reordering the observations.

Kianifard and Swallow (1987) showed that the recursive method has good statistical properties. Under the model, the level of the test is very close to the nominal level, with or without ordering by $|e_{s_i}|$, D_i, or $COVRATIO_i$, provided $s_{(i)}$ is used to scale w_j. The test has good power, which is improved by the adaptive ordering recommended above. The use of recursive residuals with adaptive ordering by construction reduces masking. This paper focuses on comparing the recursive method with some of its competitors.

The recursive method, as formulated above, tests the null hypotheses that the j^{th} observation is not an outlier, $j = p+1,\ldots,n$. Of course, this is not the same as the no-outliers hypothesis tested at each stage by the sequential or multistage procedures. Using a variation of the recursive method described above one can test the null hypothesis H_0 of no outliers versus the alternative of one or more outliers. Since each $w_j/s_{(i)}$, $j = p+1,\ldots,n$, has an approximate t

distribution, one can use Bonferroni inequalities to obtain critical values for $\max_{p+1\le j\le n} |w_j/s_{(i)}|$. Hence, the no-outliers hypothesis is rejected at nominal level α whenever

$$\max_{p+1\le j\le n} |w_j/s_{(i)}| > t(\frac{\alpha}{2n}, n-p-1) , \tag{3.6}$$

where $t(\alpha,\nu)$ denotes the $(1 - \alpha)$ percentile of the Student's t distribution with ν df. The nominal α is an upper bound to the actual significance level of the test. This procedure can be implemented in the way the sequential method (Section 2) was. If H_o is rejected, the observation corresponding to $\max_{p+1\le j\le n} |w_j/s_{(i)}|$ is declared an outlier. This observation is then removed from the data set, recursive residuals are recomputed for the remaining n - 1 observations, the test of hypothesis performed as before, and so on. The number of outliers declared is determined when a null hypothesis is not rejected. We call this procedure the "sequential recursive method".

Another variation of recursive method, referred to hereafter as the "modified recursive method", can be described as follows. Use (3.6) to test for the presence of at least one outlier in the data. If the test indicates the presence of at least one outlier, then compare the remaining n - p - 1 of the $|w_j/s_{(i)}|$ to $t(\alpha/2, n-p-1)$, as under the original formulation of the recursive method, to see if additional outliers are to be declared. Thus, the same test criterion is used by this method and the sequential recursive method for the initial test of the no-outliers hypothesis, but this method uses less stringent criteria in subsequent tests (if any) for additional outliers. The computational burden of this method is lighter than for the sequential recursive method. This method requires computing only one set of $|w_j/s_{(i)}|$,

whereas the sequential recursive method recomputes a new set of $|w_j/s_{(i)}|$ each time a null hypothesis is rejected and an observation deleted.

4. MONTE CARLO COMPARISON OF METHODS

We compared the performance of the sequential method, the multistage procedure, the sequential recursive method, the modified recursive method, and the recursive method under the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ with $n = 25$ observations per sample. The residuals are unaffected by the particular values of β_0 and β_1; we set $\beta_0 = 0$, $\beta_1 = 1$. The x's were generated as uniform (0,1) random variables multiplied by 15. The ε's were generated as N(0,1) random variables. Computations were programmed in SAS and used SAS's routines for generating uniform and normal random variables. Up to 3 outliers per sample were introduced by adding appropriate δ's in place of ε's to (effectively) randomly-selected y values. The cases we explored will be identified by the triplet $(\delta_1,\delta_2,\delta_3)$. The case designated by (9,5,0), for example, has two outliers with the δ's being 9 and 5. The results are based on 1000 simulated samples of $n = 25$ for each case, with the x's regenerated after every 100 samples. All procedures used the same simulated samples. The nominal level of the test was $\alpha = .05$ throughout.

Tables 1 and 2 summarize the results of the study. The entries in Table 1 are the proportions of correctly identified outliers (NOCORR) and the proportions of "good" observations incorrectly identified as outliers (NOINC). Each entry in Table 2 is the proportion of 1000 samples in which the indicated number of outliers was found.

The multistage procedure requires that k, the upper limit for the number of outliers that can be declared, be specified in advance. In practice, k will presumably be unknown, so the effect of the value specified on the properties of the procedure is of interest; we show results for k = 2,3,4, and 5. The sequential recursive method, the modified recursive method, and the recursive method require ordering the observations by some diagnostic measure before calculating recursive residuals. We show only the results for ordering by $|e_{s_i}|$; ordering by D_i or descending $COVRATIO_i$ gave very similar results.

In discussing the entries in the tables, we will initially exclude the recursive method, which tests very different hypotheses, and confine our remarks to a comparison of the other four methods. The entries for the no-outliers case (0,0,0) in Table 2 show that all four procedures have Type I error rates close to the nominal level of $\alpha = .05$.

For a single outlier of modest size ($\delta = 4$), the multistage procedure performed worse than the other three methods. Whereas the other methods identified 70% of outliers of this type, the multistage procedure found 35-60% (Table 1). The performance of the multistage procedure was worse for larger, that is, more overspecified, k. The modified recursive method found more than one outlier in about 14% of the samples (Table 2), consistent with its less stringent criteria for declaring additional outliers beyond the first. When two small outliers [(3,3,0), (3,-3,0), or (4,3,0)] were introduced, the multistage procedure performed somewhat better than the other three procedures, but all had low power.

For a pair of large outliers [(7,7,0), (7,-7,0), (9,5,0), or (9,-5,0)], the four methods all performed quite well, with the exception of the

sequential method for the outlier pattern (7,7,0). The particular outlier configuration (7,7,0) leads to masking under the sequential method. All but the modified recursive method found three or more outliers in about 5% of the samples (Table 2). Unlike the other three procedures, the modified recursive method does no deleting and recomputing as outliers are declared, and is thus more vulnerable to masking, as seen under (9,5,0) and (9,-5,0) in Table 2. For the outlier patterns (9,5,0) and (9,-5,0), the multistage procedure suffered when k was overspecified.

When we introduced three outliers, the multistage procedure was clearly the winner, provided k was not underspecified. The other three methods all suffered from masking. Of course, three outliers in a sample of 25 constitutes a very high proportion of outliers and exaggerates any masking problems. The pattern of positive and negative δ's in the outlier pattern made substantial differences in the masking problems of all but the multistage procedure. The pattern (7,7,7) caused more serious masking than did (7,7,-7). Pattern (9,5,-5) generally caused less masking than did (9,5,5) or (-9,5,5), although for the modified recursive method, (9,5,-5) and (9,5,5) differed little.

The four procedures compared above all test a no-outliers hypothesis, that is, a joint hypothesis that every one of the n observations is a non-outlier. This requires a rather stringent criterion, increasingly so as n becomes larger. As a consequence, only extreme outliers are likely to be detected. Some readers may have been surprised to see us exploring the ability of these procedures to detect observations with "errors" as large as $\delta = 7 = 7\sigma$ or $\delta = 9 = 9\sigma$. When the data analyst's goal is to identify

outliers/influential observations for scrutiny, not merely to justify deleting them from the data set, the above procedures may seem very unsatisfactory. The user may prefer a procedure which directs attention to modest outliers (δ's of, say, 3 to 5). The original formulation of the recursive method was intended to do just that.

The recursive method tests one-by-one the null hypotheses that the j^{th} (ordered) observation is not an outlier, $j = p + 1,\ldots,n$. The observed level of this testing procedure is shown as NOINC for the no-outliers case in Table 1; the level .048 is close to the nominal $\alpha = .05$. For this procedure, unlike the procedures discussed earlier, the level is the proportion of observations incorrectly identified as outliers when none are present. Although only $n - p$ recursive residuals are available for testing, with ordering all n observations are effectively tested. Thus, the value .048 was calculated as the proportion of all 25,000 simulated observations incorrectly identified as outliers by the procedure. Similarly, when one outlier is introduced into each sample, NOINC is the proportion of all 24,000 non-outliers incorrectly thought to be outliers, although fewer than 24,000 are actually tested. And so on.

The recursive method had power .999 to detect a single outlier having $\delta = 4$ (Table 1). In contrast, the other procedures had powers .7 or less.

When two small to moderate outliers were introduced [(3,3,0), (3,-3,0), or (4,3,0)], the recursive method identified 75-85% of the outliers (NOCORR), whereas the other procedures detected only 0-18% of them. The number of non-outliers incorrectly identified as outliers (NOINC) by the recursive method quickly dropped well below $\alpha = .05$ as

even these small outliers were introduced. This is an attractive feature of the procedure, and reflects inflation by the outlier(s) of the denominators, $s_{(i)}$, of the test statistics. For outlier patterns featuring two extreme outliers in samples of n = 25 [(7,7,0), (7,-7,0), (9,5,0), or (9,-5,0)] the recursive method identified most of the outliers, as did the other methods.

For the three-outlier patterns (7,7,7) and (7,7,-7), the recursive method performed flawlessly (NOCORR = 1, NOINC = 0). The multistage method also handled these patterns well as long as $k \geq 3$, that is, provided the user had the vision to specify that there might be three or more outliers among the n = 25 observations. The sequential, sequential recursive, and modified recursive procedures all experienced severe masking problems under these outlier patterns.

The outlier patterns (9,5,5), (9,5,-5), and (-9,5,5) created moderate masking problems for the recursive method. The most extreme outlier ($\delta = \pm 9$) was easily identified, but it grossly inflated the $s_{(i)}$ divisors used in testing the other two outliers, sometimes masking their presence. For those patterns that included an extreme outlier ($\delta = \pm 9$), the recursive and modified recursive methods performed identically, and about as well as the sequential or sequential recursive methods. As noted earlier, having three large outliers, one with $\delta = \pm 9$, in a sample of only n = 25, creates more serious masking than is likely in practice.

5. SUMMARY

1. All of the procedures had levels close to the nominal $\alpha = .05$.
2. The stringent test criteria for the four procedures testing the no-outliers hypothesis, meant that a single moderate outlier with "error" $\delta = 4 = 4\sigma$ in a sample of $n = 25$ often went undetected. The recursive method virtually always identified an outlier of that magnitude. When two small to moderate outliers (δ's of 3 or 4) were present, the procedures testing the no-outliers hypothesis rarely identified either, the recursive method usually identified both.
3. When there were two large outliers (δ's of 5,7, or 9) in the samples of $n = 25$, all procedures identified the outliers with high probability.
4. When three outliers were introduced, only the multistage procedure performed consistently well, and it, of course, only when $k \geq 3$ was specified. The recursive method performed flawlessly under some three-outliers situations, but suffered from some degree of masking under others. The sequential, sequential recursive, and modified recursive methods suffered from drastic masking under patterns (7,7,7) and (7,7,-7).
5. The sequential recursive method was never enough better than the modified recursive method to justify the extra computational effort it requires. Except under the outlier pattern (7,7,0), the even-simpler sequential procedure was as good as the sequential recursive method, or nearly so.

6. The multistage procedure requires prespecifying k, the maximum number of outliers that can then be declared. This poses a problem in practice, as the user will not, in general, have any advance knowledge of the number of outliers present. If k is underspecified, some outliers will necessarily be left unidentified. If k is overspecified, the power of the procedure to identify the existing outliers may fall, as seen in our simulation results. When one or two outliers were present, using k = 5 rather than k = 2 caused a 25% loss in power under several outlier patterns.
7. Whereas the sequential and multistage procedures require special tables for critical values, the recursive method and its variations use value of Student's t.

ACKNOWLEDGEMENTS

The programming assistance of Sandra B. Donaghy was greatly appreciated by the authors. This research was partially supported by the North Carolina Agricultural Foundation.

REFERENCES

Atkinson, A. C. (1981), " Two Graphical Displays for Outlying and Influential Observations in Regression," *Biometrika*, 68, 13-20.

Barnett, V., and Lewis, T. (1984), *Outliers in Statistical Data*, Second Edition, New York: Wiley.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, New York: Wiley.

Brown, R. L., Durbin, J., and Evans, J. M. (1975), "Techniques for Testing the Constancy of Regression Relationships Over Time," *Journal of the Royal Statistical Society*, Series B, 37, 149-192.

Chatterjee, S., and Hadi, A. S. (1986), "Influential Observations, High Leverage Points, and Outliers in Linear Regression," *Statistical Science*, 1, 379-416.

Cook, R. D. (1977), "Detection of Influential Observations in Linear Regression, *Technometrics*, 19, 15-18.

Fung, W. K. (1987), Letter to the Editor, *Technometrics*, 29, 125.

Gentleman, J. F., and Wilk, M. B. (1975), "Detecting Outliers II: Supplementing the Direct Analysis of Residuals," *Biometrics*, 31, 387-410.

Hedayat, A., and Robson, D. S. (1970), "Independent Stepwise Residuals for Testing Homoscedasticity," *Journal of the American Statistical Association*, 65, 1573-1581.

Kianifard, F., and Swallow, W. H. (1987), "Using Recursive Residuals, Calculated on Adaptively-Ordered Observations, to Identify Outliers in Linear Regression," North Carolina State University, Institute of Statistics Mimeo Series No. 1905, 1987.

Lund, R. E. (1975), "Tables for an Approximate Test for Outliers in Linear Models," *Technometrics*, 17, 473-476.

Marasinghe, M. G. (1985), "A Multistage Procedure for Detecting Several Outliers in Linear Regression," *Technometrics*, 27, 395-399.

Phillips, G. D. A., and Harvey, A. C. (1974), "A Simple Test for Serial Correlation in Regression Analysis," *Journal of the American Statistical Association*, 69, 935-939.

Plackett, R. L. (1950), "Some Theorems in Least Squares," *Biometrika*, 37, 149-157.

Prescott, P. (1975), "An Approximate Test for Outliers in Linear Models," *Technometrics*, 17, 129-132.

Theil, H. (1971), *Principles of Econometrics*, New York: Wiley.

Tietjen, G. L., Moore, R. H., and Beckman, R. J. (1973), "Testing for a Single Outlier in Simple Linear Regression," *Technometrics*, 15, 717-721.

Table 1. Proportions of outliers correctly identified (NOCORR) and non-outliers incorrectly declared to be outliers (NOINC) by five procedures under various outlier patterns (n = 25).

Outlier Pattern ($\delta_1, \delta_2, \delta_3$)		Outlier-Identification Procedure							
		Sequential	Multistage				Sequential Recursive (Ordered by $\lvert e_{si} \rvert$)	Modified Recursive (Ordered by $\lvert e_{si} \rvert$)	Recursive (Ordered by $\lvert e_{si} \rvert$)
			k=2	k=3	k=4	k=5			
(0,0,0)	NOCORR	–	–	–	–	–	–	–	–
	NOINC	.002	.003	.003	.004	.003	.002	.003	.048
(4,0,0)	NOCORR	.704	.599	.499	.405	.357	.704	.704	.999
	NOINC	.002	.002	.003	.003	.004	.002	.006	.011
(3,3,0)	NOCORR	.002	.128	.148	.151	.151	.002	.004	.767
	NOINC	.000	.000	.002	.003	.004	.000	.000	.015
(3,-3,0)	NOCORR	.078	.153	.177	.174	.164	.078	.110	.847
	NOINC	.000	.001	.002	.003	.003	.000	.000	.008
(4,3,0)	NOCORR	.146	.281	.290	.275	.245	.146	.175	.832
	NOINC	.001	.001	.002	.003	.003	.001	.001	.008
(7,7,0)	NOCORR	.766	.999	1	.997	.992	.932	.932	1
	NOINC	.002	.000	.002	.003	.003	.002	.000	.000
(7,-7,0)	NOCORR	.999	1	1	.998	.995	.999	.999	1
	NOINC	.002	.000	.002	.003	.003	.002	.000	.000
(9,5,0)	NOCORR	.987	.986	.962	.925	.875	.987	.951	.951
	NOINC	.002	.000	.002	.003	.003	.002	.000	.000
(9,-5,0)	NOCORR	.990	.990	.967	.922	.879	.990	.925	.925
	NOINC	.002	.000	.002	.003	.003	.002	.000	.000
(7,7,7)	NOCORR	.026	.479	1	1	.999	.028	.030	1
	NOINC	.000	.000	.000	.003	.003	.000	.000	.000
(7,7,-7)	NOCORR	.281	.557	1	1	.999	.324	.347	1
	NOINC	.001	.000	.000	.003	.003	.001	.000	.000
(9,5,5)	NOCORR	.587	.460	.984	.951	.904	.606	.714	.714
	NOINC	.001	.000	.000	.003	.003	.001	.000	.000
(9,5,-5)	NOCORR	.794	.565	.985	.956	.911	.794	.733	.733
	NOINC	.002	.000	.000	.003	.003	.002	.000	.000
(-9,5,5)	NOCORR	.587	.460	.984	.951	.904	.606	.570	.570
	NOINC	.001	.000	.000	.003	.003	.001	.000	.000

Table 2. Distributions of the number of observations declared to be outliers by five procedures under various outlier patterns (n = 25).

Outlier Pattern ($\delta_1, \delta_2, \delta_3$)	Number of Outliers Declared	Outlier-Identification Procedure							
		Sequential	Multistage				Sequential Recursive (Ordered by $\lvert e_{si} \rvert$)	Modified Recursive (Ordered by $\lvert e_{si} \rvert$)	Recursive (Ordered by $\lvert e_{si} \rvert$)
			k=2	k=3	k=4	k=5			
(0,0,0)	0	.944	.941	.936	.942	.943	.944	.944	.146
	≥1	.056	.059	.064	.058	.057	.056	.056	.854
(4,0,0)	0	.296	.400	.498	.593	.639	.296	.296	.000
	1	.660	.545	.446	.347	.297	.660	.563	.740
	≥2	.044	.055	.056	.060	.064	.044	.141	.260
(3,3,0)	0	.991	.858	.823	.811	.811	.991	.991	.044
	1	.007	.021	.045	.052	.050	.007	.005	.167
	2	.001	.121	.097	.103	.094	.001	.002	.646
	≥3	.001	-	.035	.034	.045	.001	.002	.143
(3,-3,0)	0	.877	.745	.722	.729	.747	.877	.877	.000
	1	.089	.190	.184	.161	.151	.089	.023	.229
	2	.030	.065	.066	.087	.066	.030	.093	.654
	≥3	.004	-	.028	.023	.036	.004	.007	.117
(4,3,0)	0	.819	.586	.564	.602	.646	.819	.819	.000
	1	.065	.256	.274	.223	.198	.065	.006	.230
	2	.111	.158	.127	.138	.111	.111	.169	.686
	≥3	.005	-	.035	.037	.045	.005	.006	.084
(7,7,0)	0	.234	.000	.000	.001	.003	.068	.068	.000
	1	.000	.001	.001	.004	.011	.000	.000	.000
	2	.729	.999	.947	.942	.933	.887	.932	.998
	≥3	.037	.000	.052	.053	.053	.045	.000	.002
(7,-7,0)	0	.001	.000	.000	.000	.000	.001	.001	.000
	1	.000	.000	.000	.004	.011	.000	.000	.000
	2	.947	1	.948	.943	.936	.947	.999	1
	≥3	.052	-	.052	.053	.053	.052	.000	.000
(9,5,0)	0	.000	.000	.000	.000	.001	.000	.000	.000
	1	.026	.026	.076	.150	.247	.026	.094	.094
	2	.923	.974	.872	.797	.700	.923	.906	.906
	≥3	.051	-	.052	.053	.052	.051	.000	.000
(9,-5,0)	0	.000	.000	.000	.000	.000	.000	.000	.000
	1	.020	.020	.066	.157	.242	.020	.151	.151
	2	.930	.980	.882	.790	.705	.930	.849	.849
	≥3	.050	-	.052	.053	.053	.050	.000	.000
(7,7,7)	0	.970	.153	.000	.000	.000	.970	.970	.000
	1	.006	.256	.000	.000	.000	.003	.000	.000
	2	.000	.591	.000	.000	.002	.000	.000	.000
	3	.022	-	1	.938	.945	.025	.030	1
	≥4	.002	-	-	.062	.053	.002	.000	.000

Table 2 (continued)

Outlier Pattern ($\delta_1, \delta_2, \delta_3$)	Number of Outliers Declared	Outlier-Identification Procedure							
		Sequential	Multistage				Sequential Recursive (Ordered by $\lvert e_{si} \rvert$)	Modified Recursive (Ordered by $\lvert e_{si} \rvert$)	Recursive (Ordered by $\lvert e_{si} \rvert$)
			k=2	k=3	k=4	k=5			
(7,7,-7)	0	.654	.015	.000	.000	.000	.653	.653	.000
	1	.097	.298	.000	.000	.000	.035	.000	.000
	2	.000	.687	.000	.000	.003	.000	.000	.000
	3	.234	-	1	.938	.944	.294	.347	1
	≥4	.015	-	-	.062	.053	.018	.000	.000
(9,5,5)	0	.000	.000	.000	.000	.000	.000	.000	.000
	1	.619	.619	.012	.026	.064	.591	.178	.178
	2	.000	.381	.023	.096	.161	.000	.501	.501
	3	.352	-	.965	.816	.722	.380	.321	.321
	≥4	.029	-	-	.062	.053	.029	.000	.000
(9,5,-5)	0	.000	.000	.000	.000	.000	.000	.000	.000
	1	.304	.304	.003	.008	.034	.304	.099	.099
	2	.009	.696	.040	.117	.200	.009	.602	.602
	3	.645	-	.957	.813	.713	.645	.299	.299
	≥4	.042	-	-	.062	.053	.042	.000	.000
(-9,5,5)	0	.000	.000	.000	.000	.000	.000	.000	.000
	1	.619	.619	.012	.026	.064	.591	.468	.468
	2	.000	.381	.023	.096	.161	.000	.354	.354
	3	.352	-	.965	.816	.722	.380	.178	.178
	≥4	.029	-	-	.062	.053	.029	.000	.000