Identification of outliers in exponential samples with stepwise procedures

V. Schultze and J. Pawlitschko*

Department of Statistics, University of Dortmund, D-44221 Dortmund

Abstract

In this paper we consider the problem of identifying outliers in exponential samples with stepwise procedures, namely inward and outward testing procedures. We treat outliers in the spirit of Davies and Gather (1993) as points which for given $\alpha > 0$ lie in a certain α-outlier region and focus especially on the worst-case behaviour of the identification rules. Best results yield stepwise procedures which use test statistics based on a standardized version of the sample median.

KEY WORDS: outlier identification, inward testing, outward testing, breakdown points, robust statistics

1 Introduction

In samples taken from some target population one often observes some data points which seem to differ strongly from the main body of the data. Such seemingly aberrant data points are usually called "outliers". However, there exists no formal definition of what constitutes an outlier that has been widely accepted.

In this paper we focus on outlying observations in exponential samples. The exponential distribution $Exp(\nu)$ with unknown scale parameter $\nu > 0$ is commonly used as a simple but nevertheless quite useful model distribution for lifetime data. The corresponding

*pawl@statistik.uni-dortmund.de

distribution and density function are given by $F_\nu(x) = 1 - \exp(-x/\nu)$, $x > 0$, and $f_\nu(x) = 1/\nu \, \exp(-x/\nu)$, $x > 0$, respectively.

Beginning with Cochran (1941), there is a rich literature on the problem of detecting outliers in exponential samples, see e.g. Gather (1995) for a comprehensive (yet not exhaustive) account on contributions to this topic. In most of this work the problem of outlier detection is seen as a testing problem with null hypothesis that all observed lifetimes come from the same exponential distribution – the null model – and alternative that at least one lifetime comes from another distribution permitted by a sometimes only implicitly assumed outlier-generating model.

An other approach recently pursued by Schultze and Pawlitschko (2000) is based on the so called α-outlier region of a distribution F which for $\alpha \ll 1$ can be described as the maximal part of the support of F which carries only a certain small probability mass. In case of $F = F_\nu = Exp(\nu)$, the corresponding α-outlier region is given by

$$out(\alpha, F_\nu) \; = \; \{x > 0 : \; x \; > \; -\nu \; \ln \alpha\}.$$

To take into account the size N of a given sample one may choose

$$\alpha \; = \; \alpha_N \; = \; 1 - \left(1 - \tilde{\alpha}\right)^{1/N}$$

for a given $\tilde{\alpha}$.

In Schultze and Pawlitschko (2000), the identification of outliers in a given sample $\underline{x}_N = (x_1, \ldots, x_N)$ is achieved by constructing an empirical version $OR(\underline{x}_N, \alpha_N)$ of $out(\alpha_N, F_\nu)$ – a so-called one-step outlier identifier – and classifying all observations as α_N-outliers which lie in OR. Typically, in the exponential case such a one-step outlier identifier has the form

$$OR_S(\underline{x}_N, \alpha_N) \; = \; \{x > 0 \colon \; x > S_N(\underline{x}_N) \, g(N, \alpha_N)\} \qquad (1)$$

where $S_N(\underline{x}_N)$ is an estimator of scale and $g(N, \alpha_N)$ an appropriate normalizing constant. In the set-up of a normal sample, Carey et al. (1997) call a rule of this type a “resistant detection rule” if it is based on robust estimators of the parameters of the null distribution. In the exponential case there are a variety of robust estimators of the scale parameter ν, and it turns out that a standardized version of the sample median yields the best results.

Other approaches for detecting α_N-outliers are not necessarily based on an explicit empirical version of $out(\alpha_N, F_\nu)$ but merely consist in a rule that classifies the observations with respect to their “outlyingness”. Especially stepwise procedures, namely consecutive inward and outward testing procedures can be interpreted in this way. The main topic of this paper is the investigation of some of these procedures with respect to their worst-case behaviour. Therefore, in Section 2 we recall some worst-case criteria for outlier identification rules, namely masking and swamping breakdown points and the size of the largest nonidentifiable outlier, and adapt them for stepwise procedures. Section 3 discusses some inward testing procedures, a classical one using Cochran-statistics at each step and three alternatives which are based on robust estimators of scale. The fourth section contains corresponding results for some outward testing procedures: again the classical version with Cochran-statistics, two procedures based on spacings and a new proposal that makes use of the standardized sample median. Clearly, worst-case analysis only sheds light on the behaviour of outlier identification rules for extreme data situations. Therefore, Section 5 contains the results of a small simulation study where we compare their power of correctly identifying the outlying observations in a contaminated sample, generated from a scale-slippage model of Ferguson-type. A real data example and a short discussion in Section 6 conclude this paper.

2 Worst-case behaviour of outlier identification rules

In the literature, comparisons of different rules for outlier identification are mostly based on simulations since exact results on e.g. power are seldom available. A field, however, where also theoretical results can be obtained is the analysis of the worst-case behaviour of the identification rules.

When identifying outliers two possible mistakes can occur: (1) a genuine outlier is not recognized as such, and (2) a non-outlying observation is identified as outlier. If these errors are caused by the outliers themselves, the phenomena are called (1) masking and (2) swamping. The worst-case concerning these mistakes occurs if either badly placed outlying observations in the sample cause an identification rule to be unable to recognize an arbitrarily large outlier as such or to identify a non-outlying observation as arbitrarily large outlier. The smallest fraction of outliers needed in a sample to achieve these extreme results are called the masking and swamping breakdown point of the identification rule, respectively (cf. Davies and Gather, 1993).

More formally, given a sequence $\underline{\alpha} = (\alpha_N)_{N\in\mathbb{N}}$ with $\alpha_N \in (0,1)$, $\delta \in (0,1)$, and a sample with n arbitrary observations $\underline{x}_n$, the masking breakdown point of an outlier identification rule OI may be defined as

$$\epsilon^M(OI, \underline{\alpha}, \underline{x}_n, \delta) = \frac{k^M}{n + k^M},$$

with $k^M = \min\{k : \beta^M(\alpha_{n+k}, \underline{x}_n, k, \delta) = 0\}$ and

$$\beta^M(\alpha_{n+k},\, \underline{x}_n,\, k,\, \delta) \;=\; \inf\big\{\beta > 0\text{: there exist } \delta\text{-outliers } \underline{x}_k^o = (x_1^o, \ldots, x_k^o) \text{ such that } OI \text{ fails to identify some } x_i^o \text{ also in } out(\beta, F_\nu) \text{ as an } \alpha_{n+k}\text{-outlier in the combined sample } (\underline{x}_n, \underline{x}_k^o)\big\}.$$

Further, the swamping breakdown point of OI may be defined as

$$\epsilon^S(OI,\, \underline{\alpha},\, \underline{x}_n,\, \delta) \;=\; \frac{k^S}{n + k^S},$$

with $k^S = \min\{k : \beta^S(\alpha_{n+k}, \underline{x}_n, k, \delta) = 0\}$ and

$$\beta^S(\alpha_{n+k},\, \underline{x}_n,\, k,\, \delta) \;=\; \inf\big\{\beta > 0\text{: there exist } \delta\text{-outliers } \underline{x}_k^o \text{ such that } OI \text{ identifies some non-}\alpha_{n+k}\text{-outlier in } \underline{x}_n \text{ as } \beta\text{-outlier in the combined sample}\big\}.$$

Here, the definitions in Davies and Gather, 1993, have been altered slightly to cover the situation that with stepwise procedures only the observations of a given sample are classified with respect to their “outlyingness”.

A high masking and swamping breakdown point are desirable features of an identification rule and there are indeed many rules which achieve the optimal value 1/2. Note that theoretically larger values are possible but this is irrelevant in practical applications since a sample with a fraction of outliers of more than 50% makes no sense. For comparisons between those optimal rules further criteria are needed. For instance, if the fraction of outlying observations in the sample is smaller than the masking breakdown point then one may still ask for the size of the largest nonidentifiable outlier which is then finite. Formally this quantity can be defined as follows. Let again OI denote an identification rule and $\tilde{\alpha} > 0$ and δ and $\underline{x}_n$ be as in the definition of the breakdown points. If for some $k < n$

$$\frac{k}{n+k} \;<\; \epsilon^M(OI,\, \underline{\alpha},\, \underline{x}_n,\, \delta),$$

then set

$$LO(k,\, OI,\, \tilde{\alpha},\, \underline{x}_n,\, \delta) \;=\; \sup\big\{x \in out(\alpha_{n+k}, F_\nu)\text{: there exists } \underline{x}_k^o \in out(\delta, F_\nu) \text{ with } x \in \underline{x}_k^o \text{ such that } OI \text{ fails to identify } x \text{ as } \alpha_{n+k}\text{-outlier in the combined sample } (\underline{x}_n, \underline{x}_k^o)\big\}.$$

The value of LO will in general depend on the given sample $\underline{x}_n$, further it may even not be well defined if the points in $\underline{x}_n$ are located such that for all possible choices of $\underline{x}_k^o$ each α_{n+k}-outlier in the combined sample will be correctly identified. To work with this worst-case criterion, one may e.g. calculate the expectation of the largest observation that is not identified as α_{n+k}-outlier under the condition that the random sample $\underline{X}_n$ comes i.i.d. from an $Exp(\nu)$ distribution. Indeed, this expectation will often be a point in $out(\alpha_{n+k}, F_\nu)$ and we will use the notation LO also in this case. Another way to work with this worst-case criterion is to give an asymptotic approximation of LO.

3 Inward testing procedures

The idea of applying stepwise procedures for detecting multiple outliers if there is no a priori information about their number immediately suggests itself and the first procedure of this type has indeed been proposed as early as in 1936 by Pearson and Chandra Sekar. With an inward testing procedure, in the first step the “most extreme observation” of a given sample is considered by means of a discordancy test. If the test decides that this is an outlier it is removed from the sample and in the next step the “most extreme” of the remaining observations is checked. The procedure terminates if for the first time such an observation is not identified as an outlier or if the largest sensible number $k^* = \lfloor (N-1)/2 \rfloor$ of possible outliers is reached.

In the exponential case, the most extreme observation of a (sub-) sample is simply its maximum. Let $x_{(1)} \leq \cdots \leq x_{(N)}$ denote the ordered observations in the complete sample and $\underline{x}_{N-i+1,N} = (x_{(1)}, \ldots, x_{(N-i+1)})$ the sample investigated in the i-th step of the inward testing procedure. There are many possible ways for choosing the test statistics for the discordancy tests in each step. Appealing are statistics of the form

$$T^S_{N-i+1}(\underline{x}_{N-i+1,N}) = \frac{x_{(N-i+1)}}{S_{N-i+1}(\underline{x}_{N-i+1,N})}, \qquad i = 1, \ldots, k^*, \tag{2}$$

where S_{N-i+1} denotes an estimator of the scale parameter ν that is based on the first $N-i+1$ ordered observations only. We consider the following choices for S_{N-i+1}:

i) Standardized median

$$SM_{N-i+1}(\underline{x}_{N-i+1,N}) = 1.4427 \, \mathsf{Med}(x_{(1)}, \ldots, x_{(N-i+1)}).$$

ii) RCS-estimator

$$RCS_{N-i+1}(\underline{x}_{N-i+1,N}) = 1.6982 \, \mathsf{Med}_k \left\{ \mathsf{Med}_j \{ |x_{(j)} - x_{(k)}|;\ j,k = 1, \ldots, N-i+1 \} \right\}.$$

iii) RCQ-estimator

$$RCQ_{N-i+1}(\underline{x}_{N-i+1,N}) = 3.476 \{ |x_{(j)} - x_{(k)}|;\ j,k = 1, \ldots, N-i+1,\ j < k \}_{(l)}$$

where

$$l = \left\lceil \frac{(N-i+1)(N-i)}{8} \right\rceil.$$

iv) Mean of the subsample (Maximum-Likelihood- (ML-) estimator)

$$ML_{N-i+1}(\underline{x}_{N-i+1,N}) = \frac{1}{N-i+1} \sum_{j=1}^{N-i+1} x_{(j)}.$$

For samples from an exponential distribution, the standardized median has been proposed as robust alternative to the sample mean by Gather and Schultze (1999). The RCS- and RCQ-estimators were proposed by Rousseeuw and Croux (1993) in the general context of robust estimation of a scale parameter. Their properties in the exponential case have been studied further by Gather and Schultze. Their usefulness in the construction of one-step outlier identifiers has been shown in Schultze and Pawlitschko (2000). Choice iv) corresponds to the well known Cochran-statistic which has been discussed e.g. by Cochran (1941), Kimber (1982), Chikkagoudar and Kunchur (1987), Likeš (1987), and Tse and Balasooryia (1991). In the following, we denote the corresponding inward testing procedures as SM-IT, RCS-IT, RCQ-IT, and Cochran-IT, respectively.

It remains to specify the critical values for the discordancy tests. A common requirement for identification rules is that for given $\tilde{\alpha} \in (0,1)$ under the null model H_0 – all X_i come from i.i.d. $Exp(\nu)$-distributed random variables – one has

$$P_{H_0}\big(\text{no observation is identified as } \alpha_N\text{-outlier}\big) \;\geq\; 1-\tilde{\alpha}. \tag{3}$$

For an inward testing procedure, condition (3) is already fulfilled if the critical value in the first step, say $t_N^S(\tilde{\alpha})$, is chosen such that the corresponding discordancy test keeps the level $\tilde{\alpha}$. The critical values $t_{N-i+1}^S(\tilde{\alpha})$, $i = 2, \ldots, k^*$, can be chosen arbitrarily. Here they are determined such that every discordancy test keeps the niveau $\tilde{\alpha}$ under H_0.

For the three inward testing procedures based on robust estimators of scale the null distribution of T_{N-i+1}^S is quite difficult to handle. For SM-IT, expressions for the survival function of T_{N-i+1}^{SM} have been calculated in Pawlitschko (2000). These expressions are helpful for deriving critical values if N is not too large. Otherwise and for RCS- and RCQ-IT in general, the most appropriate way to obtain critical values is via simulations. For Cochran-IT a quite good approximation is given by choosing

$$t_{N-i+1}^{ML}(\tilde{\alpha}) \;=\; (N-i+1)\,\frac{1-a_{N-i+1}}{1+(i-1)\,a_{N-i+1}} \tag{4}$$

where for $i = 1, \ldots, k^*$

$$a_{N-i+1} \;=\; \left(\frac{\tilde{\alpha}}{\binom{N}{i-1}}\right)^{1/(i-1)} \tag{5}$$

see e.g. Likeš (1987). This approximation is exact if $t_{N-i+1}^{ML}(\tilde{\alpha}) > (N-i+1)/2$, otherwise it makes the Cochran-tests slightly conservative.

We now investigate masking and swamping breakdown points of the four inward testing procedures.

Theorem 3.1 Let $\underline{\alpha} = (\alpha_N)_{N\in\mathbb{N}}$ be a sequence with $\alpha_N \in (0,1)$, $\delta \in (0,1)$, and $\underline{x}_n$ be a sample of arbitrary observations assumed to come from $Exp(\nu)$. If $\tilde{\alpha}$ in (3) is chosen reasonably small, then

(i) $\epsilon^M(\text{SM-IT}, \underline{\alpha}, \underline{x}_n, \delta) \;=\; \epsilon^M(\text{RCS-IT}, \underline{\alpha}, \underline{x}_n, \delta) \;=\; \epsilon^M(\text{RCQ-IT}, \underline{\alpha}, \underline{x}_n, \delta) \;=\; \frac{1}{2}$,

(ii) $$\epsilon^M(\text{Cochran-IT}, \underline{\alpha}, \underline{x}_n, \delta) \;\le\; \frac{k^*}{n+k^*},$$

where $k^* \in \{1, \ldots, n\}$ is the smallest k such that with $N = n + k$

$$\frac{N}{k} \;\le\; t_N^{ML}(\tilde{\alpha}).$$

Proof.

(i) Suppose that for an identification rule with test statistics of type (2) there exists $k \in \{1, \ldots, n-1\}$ such that

$$\beta^M(\alpha_{n+k},\, \underline{x}_n,\, k,\, \delta) \;=\; 0 \;\;.$$

That is, for any $\beta > 0$ we can find k δ-outliers $\underline{x}_k^o = (x_1^o, \ldots, x_k^o)$ such that at least one arbitrarily large β-outlier contained in the combined sample $\underline{x}_N = (\underline{x}_n, \underline{x}_k^o)$ of size $N = n+k$ is not identified as α_N-outlier. If $\beta \to 0$, the lower border of the corresponding β-outlier region moves to infinity. Now assume that for some $r \le k$ we have that $x_{(N-r+1)} \in out(\beta, F_\nu)$ – then it also holds that $x_{(N-i+1)} \in out(\beta, F_\nu)$ for $i < r$ – and that it is not correctly identified as α_N-outlier. This implies the existence of some $j \le r$ such that

$$T^S_{N-j+1}(\underline{x}_{N-j+1,N}) \;\le\; t^S_{N-j+1}(\tilde{\alpha}). \tag{6}$$

However, if S is chosen as an estimator of scale with explosion breakdown point

$$\epsilon^+(S, \underline{x}_n) \;=\; \frac{1}{2} \;>\; \frac{i}{n+i}$$

the denominator of T^S_{N-i+1}, $i = 1, \ldots, n-1$, is bounded. This holds especially for j, hence for $\beta \to 0$ we have that $T^S_{N-j+1}(\underline{x}_{N-j+1,N})$ moves to infinity which contradicts (6). As a simple consequence

$$\epsilon^M(OI, \underline{\alpha}, \underline{x}_n, \delta) \;\ge\; \frac{n-1}{2\,n-1} \tag{7}$$

for SM-, RCS-, and RCQ-IT, since the explosion breakdown point of the standardized median, the RCS-, and the RCQ-estimator equals 1/2 (cf. Gather and Schultze, 1999).

If $k = n$, for each of those three inward testing procedures we have for reasonable small $\tilde{\alpha}$ that

$$\inf_{\underline{x}_n^o \in out(\delta, F_\nu)} T^S_N(\underline{x}_N) \;<\; t^S_N(\tilde{\alpha}),$$

where this infimum is obtained if some of the δ-outliers move to infinity and therefore are also β-outliers for each $\beta > 0$. For example, consider the case of $S = SM$. The least favourable position of the δ-outliers is given if they all are placed at the same point x^o. In this case

$$\inf_{\underline{x}_n^o \in out(\delta, F_\nu)} T_N^{SM}(\underline{x}_N) \;=\; \lim_{x^o \to \infty} T_N^{SM}(\underline{x}_N) \;=\; 2\,\ln 2 \;\approx\; 1.386$$

which is smaller than $t_N^{SM}(\tilde{\alpha})$ for all reasonable choices of $\tilde{\alpha}$. For the other two procedures similar arguments apply. Together with (7), these findings establish part (i) of the theorem.

(ii) For Cochran-IT, the least favourable position of k δ-outliers is given if they all are placed at the same point x^o. This placement gives

$$\inf_{\underline{x}_k^o \in out(\delta, F_\nu)} T_N^{ML}(\underline{x}_N) \;=\; \lim_{x^o \to \infty} T_N^{ML}(\underline{x}_N) \;=\; \frac{n+k}{k},$$

note again that for $x^o \to \infty$ these δ-outliers are also β-outliers for any $\beta > 0$. $\square$

If an inward testing procedure is build with test statistics of type (2) and the critical value in the first step is chosen according to (3), then the most extreme observation $x_{(N)}$ is identified as α_N-outlier if and only if it is located in the corresponding empirical α_N-outlier region (1) based on the scale estimator S_N. However, as the proof of Theorem 3.1 (ii) shows, this fact does not allow the conclusion that the masking breakdown points of these two identification rules are equal. The reason is the slightly different definition of the breakdown point for the two classes of identification rules.

Theorem 3.2 Given an arbitray sample $\underline{x}_n$, $\delta \in (0,1)$, and a sequence $\underline{\alpha} = (\alpha_N)_{N\in\mathbb{N}}$, the swamping breakdown point of SM-IT, RCS-IT, RCQ-IT, and Cochran-IT is not smaller than 1/2.

Proof. Denote with OI any of the four inward testing procedures. Suppose there exists $k \in \{\,1,\ldots,n-1\}$ with

$$\beta^S(OI, \alpha_{n+k}, \underline{x}_n, k, \delta) \;=\; 0 \;\;.$$

Denote with x^r the largest non-α_{n+k}-outlier with respect to F_ν that is contained in $\underline{x}_n$. The condition for β^S is then fulfilled if and only if there exist k δ-outliers $\underline{x}_k^o \in out(\delta, F_\nu)$ such that

$$T_{N-i+1}^S(\underline{x}_{N-i+1,N}) \;>\; t_{N-i+1}^S(\beta), \quad i = 1\;,\ldots,k^*, \tag{8}$$

for all $\beta > 0$, where k^* denotes the rank of x^r in the combined sample $(\underline{x}_n, \underline{x}_k^o)$. However, if $\beta \to 0$, it follows generally for all $i \in \{\,1,\ldots,k^*\}$ that $t_{N-i+1}^S(\beta) \to \infty$. Since$x^r < -\nu\;\ln(\alpha_N) < \infty$, at least at stage k^* of the inward testing procedure the corresponding test statistic becomes bounded. This is obvious for Cochran-IT, in case of SM-, RCS-, and RCQ-IT the boundedness is due to the implosion breakdown point 1/2 of the scale estimators involved (cf. Gather and Schultze, 1999). This finding contradicts (8) and hence proves the theorem. $\square$

Concerning the largest nonidentifiable outlier in the presence of k δ-outliers, the calculation of its expectation would be a quite involved task. However, for SM-IT and Cochran-IT it is possible to give an upper bound which is quite accurate where in case of Cochran-IT one has to assume that k/N is smaller than the masking breakdown point.

Theorem 3.3 Let $\underline{X}_n = (X_1, \ldots, X_n)$ be a sample of size n coming i.i.d. from a $Exp(\nu)$-distribution and $X_{(1)} \leq \cdots \leq X_{(n)}$ denote the ordered sample. For $k < n$ and appropriate $\delta \in (0,1)$ let $\underline{X}_k^o = (X_1^o, \ldots, X_k^o)$ be a random sample of δ-outliers (possibly dependent of $\underline{X}_n$) and set $N = n + k$. Then conditional on $\min(\underline{X}_k^o) > X_{(n)}$, for all reasonable $\tilde{\alpha} > 0$ we have the following results for the expected value of the largest observation that is not identified as α_N-outlier.

(i) For SM-IT one has

$$
\begin{aligned}
&\mathsf{E}\big(LO(k,\, \text{SM-IT},\, \underline{X}_n,\, \tilde{\alpha},\, \delta)\big) \\
&= \nu\, \frac{t_N^{SM}(\tilde{\alpha})}{\ln 2} \begin{cases} \displaystyle\sum_{i=1}^{(N+1)/2} \frac{1}{N-k-i+1}, & N \text{ odd}, \\ \displaystyle\left(\sum_{i=1}^{N/2} \frac{1}{N-k-i+1} + \frac{1}{N-2k}\right), & N \text{ even}. \end{cases}
\end{aligned}
$$

(ii) If

$$
\frac{k}{N} \leq \frac{1}{t_N^{ML}(\tilde{\alpha})} \tag{9}
$$

then for Cochran-IT one has

$$
\mathsf{E}\big(LO(k, \text{Cochran-IT}, \underline{X}_n, \tilde{\alpha}, \delta)\big) = \nu\, \frac{t_N^{ML}(\tilde{\alpha})}{N - k\, t_N^{ML}(\tilde{\alpha})}\,(N-k).
$$

Proof. W.l.o.g. let $\nu = 1$. For both inward testing procedures the least favourable constellation of k δ-outliers is given if they are located at one point, say X^o, such that in the first step the corresponding discordancy test fails to reject.

(i) In case of SM-IT, let X^o be defined as the solution of

$$
X^o = t_N^{SM}(\tilde{\alpha})\, \frac{\mathsf{Med}(\underline{X}_n, \underline{X}_k^o)}{\ln 2}
$$

where now $\underline{X}_k^o$ denotes the random vector with k components equal to X^o. Provided that this solution is larger than $X_{(n)}$, we have that $\mathsf{Med}(\underline{X}_n, \underline{X}_k^o)$ does depend on $\underline{X}_k^o$ only through k and its distribution is equal to that of the $(N+1)/2$-th order statistic out of $N-k$ observations from an $Exp(\nu)$-distribution if N is odd and to that of the mean of the $N/2$-th and $(N+1)/2$-th order statistic if N is even. The result now follows from the well known relation

$$\mathsf{E}(X_{(r)}) \;=\; \nu \sum_{i=1}^{r} \frac{1}{N-k-i+1} \tag{10}$$

which holds for the r-th order statistic out of $N-k$ observations from $Exp(\nu)$.

(ii) In case of Cochran-IT, consider the equation

$$X^o \;=\; t_N^{ML}(\tilde{\alpha})\,\frac{1}{N}\,\Big(\sum_{i=1}^{n} X_i + k\,X^o\Big).$$

Under condition (9) this equation has a positive finite solution

$$X^o \;=\; \frac{t_N^{ML}(\tilde{\alpha})}{N-k\,t_N^{ML}(\tilde{\alpha})}\,\sum_{i=1}^{n} X_i.$$

Taking expectation gives the expression stated in the theorem. □

Simulations show that in both cases the probability of the event $X^o \le X_{(n)}$ is quite small even if k is small, hence the unconditional expectation – which is smaller – will not be very different from the conditional one. Further calculations have shown that for reasonable $\tilde{\alpha}$ the expectations given in the theorem are indeed α_N-outliers.

For the three inward testing procedures based on robust estimators of scale the following asymptotic result can be derived.

Theorem 3.4 Let $X_1, X_2, \ldots$ be a sequence of observations coming i.i.d. from an $Exp(\nu)$-distribution and $\underline{X}_n$ the sample consisting of the first n observations of this sequence. For $\eta \in (0,1)$ set $k = \lfloor \eta\, n \rfloor$, and $N = k+n$. Further, let $\underline{\delta} = (\delta_N)_{N\in\mathbb{N}}$ denote an appropriate sequence with elements in $(0,1)$, $\underline{X}_k^o$ a sample of size k containing δ_N-outliers (possibly dependent on $\underline{X}_n$), and $\underline{X}_N = (\underline{X}_n, \underline{X}_k^o)$ the combined sample. Then for all $\tilde{\alpha} > 0$, with $n \to \infty$ the probability that SM-, RCS-, and RCQ-IT identify all α_N-outliers in $out(\alpha_N^o(\eta), F_\nu)$ converges to one, where

$$\alpha_N^o(\eta) \;=\; \alpha_N^{\exp(b(S,\,\eta,\,\underline{\delta}))}$$

and

$$b(S,\,\eta,\,\underline{\delta}) \;=\; \limsup_{n\to\infty}\left(\sup_{\underline{X}_k^o \in out(\delta_N, F_\nu)} \ln\Big(\frac{S_N(\underline{X}_N)}{\nu}\Big)\right)$$

denotes the maximum asymptotic bias of the respective scale estimator S.

Before giving the proof of Theorem 3.4 some remarks are in order: The maximum asymptotic bias for scale estimators has been introduced in Davies and Gather (1993). For the estimators considered in Theorem 3.4 the corresponding expressions can be found in Schultze and Pawlitschko (2000). An approximation of the largest nonidentifiable outlier in a sample of size N with a fraction $\eta/(1+\eta)$ of δ_N-outliers is then given as

$$ALO(\eta,\, OI,\, N,\, \underline{\alpha},\, \underline{\delta}) \;=\; -\nu\; \ln \alpha_N^o(\eta),$$

where OI stands for one of the three inward testing procedures considered in the theorem. Note that these approximations coincide with those for the corresponding one-step identifiers (1) based on the robust scale estimators. Hence the comparisons in Schultze and Pawlitschko (2000) carry over directly to the inward testing scheme. For all values of N and k considered there it turned out that ALO is smallest for SM-IT and largest for RCQ-IT.

Proof of Theorem 3.4. Under the conditions stated in the theorem, the largest possible observation $X_{(N)}^o$ in $\underline{X}_N$ which is not identified as an α_N-outlier by the inward testing procedure based on S_N is determined from

$$X_{(N)}^o \;=\; t_N^S(\tilde{\alpha}) \sup_{\underline{X}_k^o \in out(\delta_N, F_\nu)} S_N(\underline{X}_N). \tag{11}$$

Consider the critical value: if S_N is a root-N-consistent estimator of ν, under the null model of no outliers we can asymptotically approximate the distribution of $T_N^S(\underline{X}_N)$ with that of $X_{(N)}/\nu$. That is, for any $c > 0$

$$\lim_{N\to\infty} \left(P_{H_0}\big(T_N^S(\underline{X}_N) \;\le\; c\big) - \left(1 - \frac{\exp(-c)}{N}\right)^N \right) \;=\; 0 \;. \tag{12}$$

Therefore, the critical values in the first step of the three identification rules considered here fulfill

$$\lim_{N\to\infty} \big(t_N^S(\tilde{\alpha}) + \ln \alpha_N\big) \;=\; 0$$

and from (11) one obtains that in probability

$$\lim_{N\to\infty} \Big(X_{(N)}^o + \nu\; \ln \alpha_N\; \exp\big(b(S,\, \eta,\, \underline{\delta})\big) \Big) \;=\; 0 \;.$$

Hence, asymptotically observations larger than $-\nu\; \ln \alpha_N\; \exp\big(b(S, \eta, \underline{\delta})\big)$ are always identified as α_N-outliers. Setting

$$\nu\; \ln(\alpha_N^o) \;=\; \nu\; \ln \alpha_N\; \exp\big(b(S, \eta, \underline{\delta})\big)$$

and solving for α_N^o gives the assertion of the theorem. □

For Cochran-IT, we can make no asymptotic statement as in Theorem 3.4, since the maximum asymptotic bias of the ML-estimator is infinity. Further, notice from Theorem 3.1 that for $n \to \infty$ the masking breakdown point of Cochran-IT tends to zero. However, in those cases where it exceeds $1/(n+1)$ it is possible to give an approximation based on asymptotic arguments that works quite well if N is sufficiently large. Given the assumptions of Theorem 3.4 and provided that

$$\eta \; < \; -\frac{1}{1+\ln \alpha_N} \tag{13}$$

we can approximate the largest nonidentifiable outlier by

$$ALO(\eta, \text{Cochran-IT}, N, \underline{\alpha}, \underline{\delta}) \; = \; \frac{\ln \alpha_N}{1+\eta\,(1+\ln \alpha_N)}. \tag{14}$$

Condition (13) is derived from the masking breakdown point of Cochran-IT given in Theorem 3.1 (ii), the argument leading to (14) is essentially the same as in the proof of Theorem 3.3.

Simulations show that all approximations work well in samples of size $N \geq 50$, see Schultze and Pawlitschko (2000) where some numerical results for the corresponding one-step identifiers are presented.

4 Outward testing procedures

The proneness of classical inward testing procedures like Cochran-IT to masking has led to the development of alternative stepwise rules. Rosner (1975), in the context of normal samples, suggested a method that inverts the concept of inward testing and, therefore, is often denoted as outward testing. In an outward testing procedure, at first the k^* "most extreme" observations of a sample $\underline{x}_N$ are removed. In the following we always set $k^* = \lfloor (N-1)/2 \rfloor$, the maximal reasonable number of outliers. Of course smaller values for k^* are possible, e.g. if one has a rough idea about the fraction of irregular observations in the sample. Then, beginning with the least extreme of the selected observations, these are tested with an appropriate discordancy test whether they are indeed outlying. If at one stage the test rejects, the corresponding observation and all that are more extreme are identified as outliers. If the test rejects not, the corresponding observation is reunited with the reduced sample and the next observation is considered. In general, in the first step a problem is the need for a criterion to judge the extremeness of an observation. In case of an exponential sample, however, this problem does not occur since naturally the k^* largest observations stick out as the most extreme ones. There is a rich literature on outward testing procedures in exponential samples, contributions to this topic have been made by Kimber (1982), Sweeting (1983), Chikkagoudar and Kunchur (1987), Likeš (1987), Balasooriya (1989), Tse and Balasooriya (1991), and Balasooriya and Gadag (1994). To our knowledge, however, the worst-case behaviour of these procedures has not been considered yet.

For each step of an outward testing procedure principally the same discordancy tests can be used as with the inward testing scheme. However, now also application of discordancy tests which in the latter case suffered from masking yields satisfactory results and therefore deserves a closer investigation. We consider the worst-case behavior of the following four outward testing procedures based on

i) Cochran-statistics (Cochran-OT),

ii) Dixon-statistics (Dixon-OT), that is in the j-th step we use

$$T^D_{N-k^*+j}(\underline{x}_{N-k^*+j,N}) \;=\; \frac{x_{(N-k^*+j)} - x_{(N-k^*+j-1)}}{x_{(N-k^*+j)}}, \qquad j = 1\,,\ldots,k^*$$

(see Dixon, 1950, Likeš, 1967, Chikkagoudar and Kunchur, 1987),

iii) test statistics proposed by Balasooriya (B-OT),

$$T^B_{N-k^*+j}(\underline{x}_{N-k^*+j,N}) \;=\; \frac{x_{(N-k^*+j)} - x_{(N-k^*+j-1)}}{W_j}, \qquad j = 1\,,\ldots,k^*,$$

with

$$W_j \;=\; \frac{\sum_{i=1}^{N-k^*+j-1} x_{(i)} + (\,N - k^* + j - 1)\; x_{(N-k^*+j-1)}}{(N - k^* + j - 1)\;(k^* - j + 1)}$$

(see Balasooriya, 1989, Tse and Balasooriya, 1991, Balasooriya and Gadag, 1994),

iv) test statistics of type (2) with the standardized median as scale estimator (SM-OT).

The first three outward testing procedures are well known. The Balasooriya-statistics have the interesting property that they are mutually independent under the null model (cf. Sweeting, 1983). In this case, the denominator of T^B_r is equal to the best linear predictor of the r-th order statistic given the $r-1$ smallest observations. The procedure based on the standardized median is new and is included in the following investigation to see how an outward testing procedure based on a robust estimator of scale compets with classical methods.

As in the previous section, we require that condition (3) is fulfilled. In case of an outward testing procedure with test statistics T_{N-k^*+j}, $j = 1\,,\ldots,k^*$, this requirement is equivalent to

$$P_{H_0}\Big(\bigcup_{j=1}^{k^*}\{\, T_{N-k^*+j}(\underline{X}_{N-k^*+j,N}) \;>t_{N-k^*+j}(\tilde{\alpha})\,\}\Big) \;\leq\; \tilde{\alpha}. \tag{15}$$

Condition (15) does not uniquely determine the critical values. This can be achieved by the additional requirement that

$$P_{H_0}\big(T_{N-k^*+j}(\underline{X}_{N-k^*+j,N}) \;>\; t_{N-k^*+j}(\tilde{\alpha})\big) \;=\; \alpha_{k^*}, \qquad j = 1\,,\ldots,k^*. \tag{16}$$

If the test statistics for the individual steps are mutually independent then α_{k^*} can be chosen to $1-(1-\tilde{\alpha})^{1/k^*}$. In case of B-OT we get the following simple expression for the corresponding critical values:

$$t^B_{N-k^*+j}(\tilde{\alpha}) \;=\; (\;N-k^*+j-1)\left(\alpha_{k^*}^{-1/(N-k^*+j-1)}-1\right)$$

(cf. Tse and Balasooriya, 1983). Otherwise, the Bonferroni-inequality allows for the slightly more conservative choice $\alpha_{k^*} = \tilde{\alpha}/k^*$. In case of Cochran-OT the critical values then can be approximated as in (4) with $\tilde{\alpha}$ in (5) now replaced with $\tilde{\alpha}/k^*$. In case of Dixon-OT we find them as the solutions of

$$\tilde{\alpha}/k^* \;=\; \big(1-t^D_{N-k^*+j}(\tilde{\alpha})\big)^{N-k^*+j-1}\prod_{i=1}^{N-k^*+j-1}\frac{N-i+1}{N-i-t^D_{N-k^*+j}(\tilde{\alpha})\;(\;N-k^*+j-i)}. \tag{17}$$

This equation can be deduced from the more general expressions in Likeš (1967) and Kabe (1970) or by direct calculation.

We now determine masking and swamping breakdown points for the four outward testing procedures introduced above.

Theorem 4.1 Given an arbitrary sample $\underline{x}_n$, $\delta \in (0,1)$, and a sequence $\underline{\alpha} = (\alpha_N)_{N\in\mathbb{N}}$, all four outward testing procedures have masking breakdown point

$$\epsilon^M(OI,\,\underline{\alpha},\,\underline{x}_n,\,\delta) \;\geq\; \frac{1}{2}.$$

Proof. We first consider the case of Cochran-OT and assume that its masking breakdown point is smaller than 1/2. With similar arguments as in the proof of Theorem 3.1 this would imply the existence of $k<n$ such that for the combined sample $(\underline{x}_n, \underline{x}^o_k)$ none of the discordancy tests with test statistics T^{ML}_{n+j}, $j = 1\;,\ldots,k$,rejects as x^o moves to infinity. Here $\underline{x}^o_k$ denotes a vector with k components equal to some $x^o \in out(\delta, F_\nu)$. That is for all $j \leq k$ we must have

$$\lim_{x^o\to\infty} T^{ML}_{n+j}(\underline{x}_{n+j,n+k}) \;\leq\; t^{ML}_{n+j}(\tilde{\alpha}/k^*). \tag{18}$$

However, already in the first of these steps

$$\lim_{x^o \to \infty} T^{ML}_{n+1}(\underline{x}_{n+1,n+k}) \;=\; n+1$$

whereas

$$t^{ML}_{n+1}(\tilde{\alpha}/k^*) \;<\; n+1$$

and this contradicts (18).

The proof for the other three outward testing procedures is quite similar and therefore omitted. $\square$

Theorem 4.2 Let the assumptions of Theorem 4.1 be fulfilled and assume further that all $x_i \in \underline{x}_n$ are positive. Then all four outward testing procedures have swamping breakdown point

$$\epsilon^S(OI, \underline{\alpha}, \underline{x}_n, \delta) \;\geq\; \frac{1}{2}.$$

Proof. Assume that the swamping breakdown point is smaller than $1/2$. As in the proof of Theorem 3.2 this implies the existence of $k < n$ such that for some $x^r \in \underline{x}_n$ which is not in $out(\alpha_{n+k}, F_\nu)$ we can find a sample $\underline{x}^o_k \in out(\delta, F_\nu)$ for each $\beta > 0$ such that x^r is falsely identified as β-outlier. As in the proof of Theorem 4.1 we look closer at Cochran-OT. Set $k^* = \lfloor (n+k-1)/2 \rfloor$. Then x^r is identified as β-outlier if for some $j \in \{\, 1, \ldots, k^* \}$ it is not smaller than the maximum of the subsample $\underline{x}_{n+k-k^*+j,n+k}$ of the combined sample $(\underline{x}_n, \underline{x}^o_k)$ and if

$$T^{ML}_{n+k-k^*+j}(\underline{x}_{n+k-k^*+j,n+k}) \;>\; t^{ML}_{n+k-k^*+j}(\beta). \tag{19}$$

Since x^r is no α_{n+k}-outlier it is bounded:

$$x^r \;\leq\; -\;\nu \ln \alpha_{n+k} \;=\; c \;<\; \infty.$$

Hence, also the test statistic of the discordancy test is bounded by

$$T^{ML}_{n+k-k^*+j}(\underline{x}_{n+k-k^*+j,n+k}) \;\leq\; (n+k-k^*+j)\,\frac{c}{\sum\limits_{i=1}^{n+k-k^*+j-1} x_{(i)} + c} \;<\; n+k-k^*+j$$

where the last inequality is strict since the regular observations are assumed to be positive. However, for $j = 1\,, \ldots, k^*$ we have

$$\lim_{\beta \to 0} t^{ML}_{n+k-k^*+j}(\beta) \;=\; n+k-k^*+j$$

and this contradicts (19).

Again, for the other three outward testing procedures the assertion of the theorem follows quite similarly. $\square$

Theorems 4.1 and 4.2 show that concerning their breakdown points there is no difference between the three "classical" outward testing procedures and the new one based on the standardized median: all procedures have optimal high masking and swamping breakdown points. Hence, a finer investigation of their worst-case behaviour based on the size of the largest nonidentifiable outlier is necessary.

As in case of the EDR-ESD identifier for normal samples which is thoroughly discussed in Davies and Gather (1993), the most infavourable constellation of outliers is given if to a regular sample of size n a "string" of k δ-outliers is added. Such a string is constructed by placing in each step of the outward testing procedure a δ-outlier at that point at which the corresponding discordancy test just fails to reject. For the four identifiers discussed here, this leads to the following results.

Theorem 4.3 Let $\underline{X}_n = (X_1, \ldots, X_n)$ be a sample of size n coming i.i.d. from an $Exp(\nu)$-distribution and with $X_{(n)}$ denote the maximum of $\underline{X}_n$. For $k < n$ and $\delta \in (0,1)$ let $\underline{X}_k^o = (X_1^o, \ldots, X_k^o)$ be a random sample of δ-outliers (possibly dependent of $\underline{X}_n$) and set $N = n + k$. Then for all reasonable $\tilde{\alpha} > 0$, conditional on the event that no regular observation is identified as α_N-outlier, the expected value of the largest observation that is not identified as α_N-outlier is given by

(i) for Cochran-OT

$$\mathsf{E}\big(LO(k, \text{Cochran-OT}, \underline{X}_n, \tilde{\alpha}, \delta)\big) \;=\; \nu\, c_N^{ML}(\tilde{\alpha}) \prod_{j=1}^{k-1} \big(1 + c_{N-k+j}^{ML}(\tilde{\alpha})\big)\,(N-k)$$

with

$$c_{N-k+j}^{ML}(\tilde{\alpha}) \;=\; \frac{t_{N-k+j}^{ML}(\tilde{\alpha}/k^*)}{(N-k+j) + t_{N-k+j}^{ML}(\tilde{\alpha}/k^*)}, \qquad j = 1, \ldots, k,$$

(ii) for Dixon-OT

$$\mathsf{E}\big(LO(k, \text{Dixon-OT}, \underline{X}_n, \tilde{\alpha}, \delta)\big) \;=\; \nu \prod_{j=1}^{k} \frac{1}{1 - t_{N-k+j}^{D}(\tilde{\alpha})} \sum_{i=1}^{k} \frac{1}{N-i+1},$$

(iii) for outward testing with Balasooriya-statistics

$$\mathsf{E}\big(LO(k, \text{B-OT}, \underline{X}_n, \tilde{\alpha}, \delta)\big) \;=\; \mathsf{E}(X_k^o),$$

with X_k^o recursively defined by

$$X_1^o = X_{(n)} + \left(\alpha_{k^*}^{-1/(N-k)} - 1\right) \left(\frac{1}{k} \sum_{i=1}^{n} X_i + X_{(n)}\right),$$

$$X_j^o = X_{j-1}^o + \left(\alpha_{k^*}^{-1/(N-k+j-1)} - 1\right) \left(\frac{1}{k-j+1} \left(\sum_{i=1}^{n} X_i + \sum_{i=1}^{j-1} X_i^o\right) + X_{j-1}^o\right),$$

for $j = 2, \ldots, k$,

(iv) for SM-OT

$$\begin{aligned} &\mathsf{E}\big(LO(k, \text{SM-OT}, \underline{X}_n, \tilde{\alpha}, \delta)\big) \\ &= \nu \, \frac{t_N^{SM}(\tilde{\alpha}/k^*)}{\ln 2} \begin{cases} \displaystyle\sum_{i=1}^{(N+1)/2} \frac{1}{N-k-i+1} & N \text{ odd}, \\ \displaystyle\left(\sum_{i=1}^{N/2} \frac{1}{N-k-i+1} + \frac{1}{N-2\,k}\right) & N \text{ even}. \end{cases} \end{aligned}$$

Proof. Since the proofs of parts (i) – (iv) are quite similar, we only look closer at Cochran-OT. Assume that for given $\tilde{\alpha}$ no regular observation is classified as α_N-outlier. To achieve that no further discordancy test rejects, choose

$$X_1^o = \frac{t_{N-k+1}^{ML}(\tilde{\alpha}/k^*)}{(N-k+1) + t_{N-k+1}^{ML}(\tilde{\alpha}/k^*)} \sum_{i=1}^{n} X_i$$

which is a δ-outlier for appropriately chosen δ, and then subsequently

$$\begin{aligned} X_j^o &= \frac{t_{N-k+j}^{ML}(\tilde{\alpha}/k^*)}{(N-k+j) + t_{N-k+j}^{ML}(\tilde{\alpha}/k^*)} \left(\sum_{i=1}^{n} X_i + \sum_{r=1}^{j-1} X_r^o\right) \\ &= c_{N-k+j}^{ML}(\tilde{\alpha}) \, \frac{1 + c_{N-k+j-1}^{ML}(\tilde{\alpha})}{c_{N-k+j-1}^{ML}(\tilde{\alpha})} \, X_{j-1}^o \end{aligned}$$

for $j = 2, \ldots, k$ Consecutive application of this recurrence relation and making use of $\mathsf{E}\big(\sum_{i=1}^n X_i\big) = \nu\,(N-k)$ yields the result in (i). For (ii) and (iv), relation (10) can be applied in the same way as in the proof of Theorem 3.3 since the smallest δ-outlier is placed to the right of $X_{(n)}$. □

Note that under the assumptions of Theorem 4.3 for SM-OT the expected size of LO only depends on the critical value used in the last step and on the expected values of certain order statistics of the regular X_i. Further, since

$$t_N^{ML}(\tilde{\alpha}/k^*) \ \geq \ t_N^{ML}(\tilde{\alpha})$$

it follows from Theorem 3.4 that the expected size of LO is always larger for the outward testing procedure based on the standardized median than for the corresponding inward testing procedure. A similar result would be obtained when using one of the other robust estimators of scale suggested in Section 3.

Figure 1 illustrates the results from Theorem 4.3 for $N = 20$ and all possible values of k. From plot a) it can be seen that B-OT and especially Dixon-OT perform very poorly if the fraction of irregular observations approaches 1/2. Therefore, plot b) contains only the values for Cochran-OT and SM-OT. Generally it can be said that the latter performs best with respect to this worst-case criterion with the exception of the case $k = 1$, where Cochran-OT has the edge. It is worth noting that the two procedures where the nominator of the test statistics used in the discordancy tests is a difference of consecutive order statistics are clearly outperformed by those where only a single order statistic appears.

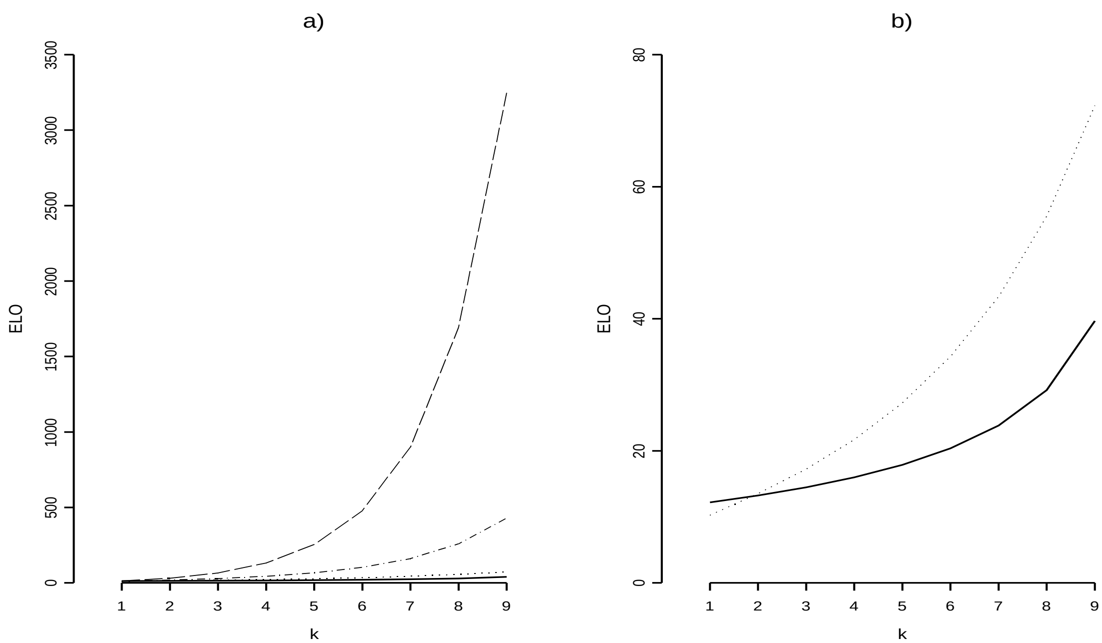

Figure 1: Expected size of LO for $\tilde{\alpha} = 0\ .05, N = 20$, with ——— : SM-OT, – – – – – : Dixon-OT, – · – · – : B-OT, and ········· : Cochran-OT

5 Some remarks on power

In the previous sections we have mainly been concerned with the worst-case behaviour of stepwise outlier identification rules. If no distributional assumptions are made for the outlying observations apart from being located in a certain outlier region, it is not possible to give general results for the performance of these identifiers in non worst-case situations. Therefore, we have also studied the fraction of correctly identified outliers and inliers (that are the sample elements not located in $out(\alpha_N, F_\nu)$) via a small simulation study under a certain popular outlier-generating model, namely a slippage model of Ferguson-type (cf. Gather, 1995). Under this model, the distribution of a random sample $\underline{X}_N = (X_1, \ldots, X_N)$ of size N is given by $F_{\underline{X}_N} = \prod_{i=1}^N F_{s(i)}$ where $s(i) \in \{1, b\}$for some $b \ll 1$ and $\sum_{i=1}^N I[s(i) = b] = k$ for some given $k \leq \lfloor N/2 \rfloor$. To guarantee that the number of outliers generated by this model is not too small we set $b = 1/13$. In Schultze and Pawlitschko (2000) the expected number of α_N-outliers for $\tilde{\alpha} = 0.05$ and some selected values for N and k has been calculated. For these values the expectation always exceeds $k/2$ but is nearly always smaller than $2/3\,k$. For the simulation study we set $N = 20, 50$, and $\tilde{\alpha} = 0.05$. The corresponding α_N-outlier regions are given by the intervals (5.97, ∞) and $(6.88, \infty)$, respectively. Then 5000 random samples from the Ferguson-model were drawn for each combination of N and some selected values of $k \leq k^* = \lfloor (N-1)/2 \rfloor$. The average fractions of correctly classified observations are displayed in Figures 2 to 5. To make differences in the height of the blocks more visible, their shading becomes more dense with growing fraction.

Concerning the fraction of correctly identified outliers, Figures 2 and 4 show that SM-IT yields the best results of the four inward testing procedures while the other two procedures based on robust estimators of scale are nearly as good. Cochran-IT performs poorly if k gets large and therefore is not recommendable. The highest power of the four outward testing procedures is obtained with Cochran-OT. If k is not to small, SM-OT is nearly as good for $N = 50$, however, it performs not very satisfactory for $N = 20$. For outward testing procedures that use test statistics of type (2) there seems to be a close connection between the finite sample efficiency of the scale estimator and the power of the identification rule. Some further simulations with outward testing procedures that are based on the RCS- and RCQ-estimator showed that they had smaller power than Cochran-OT too, although they did not perform better than SM-OT. Dixon-OT in general shows the least convincing results, whereas B-OT performs quite well as long as k is not too large. When comparing inward and outward testing procedures one finds that the latter are always inferior to their competitors if k is small and have similar power for large k only. This comparison, however, is somewhat misleading, which becomes clear from Figures 3 and 5 where the fraction of correctly classified inliers is shown. With few exceptions for all outward testing procedures this fraction is at least as great as 90% irrespective of the value of k, whereas for the inward testing procedures it may be smaller than 85% if k

is large. The reason can be seen in the choice of the critical values from the second to k^*-th step of the inward testing procedures. These values are determined under the null model of no outliers which leads to discordancy tests that are too liberal if indeed outliers are contained in the sample. There is a nice analogy in the field of multiple testing: The probability of declaring at least one regular observation as outlier corresponds in some way to the probability of falsely rejecting at least one true null hypothesis in a given family of hypotheses, the so-called familywise error rate (FWE). An outlier identification rule that is standardized according to (3) corresponds to a multiple test that controls FWE only in the weak sense that is if all hypotheses in the family are true. Neither outward nor inward testing procedures provide an equivalent to strong control of FWE. However the former seem to have a smaller probability of classifying at least one regular observation as outlying if the null model does not hold.

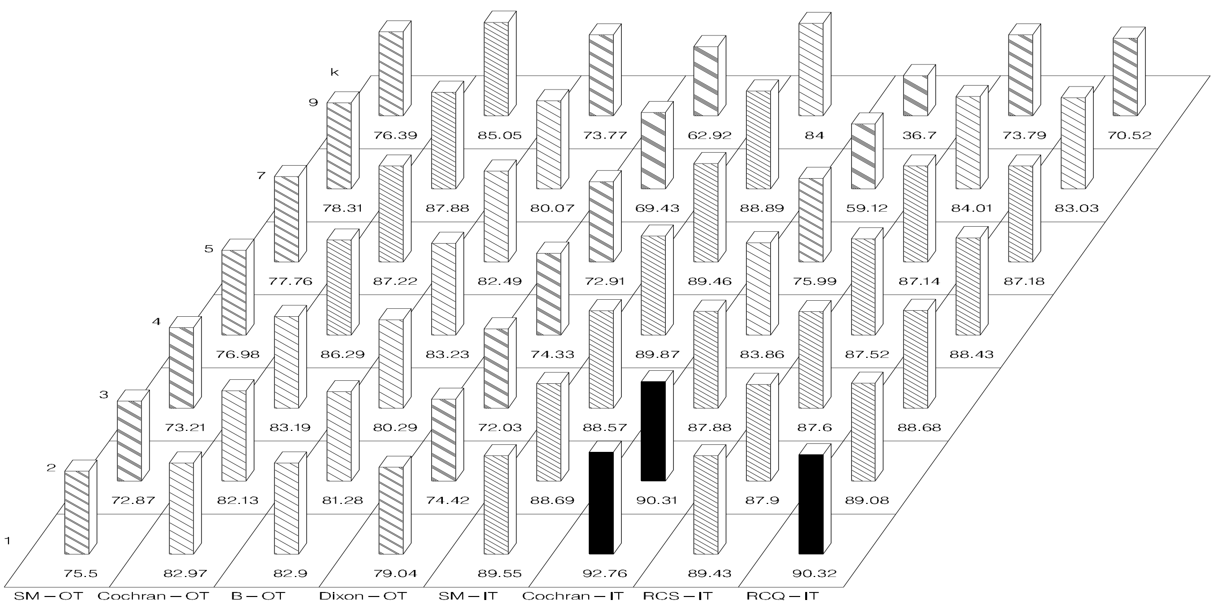

Figure 2: Fraction of correctly identified outliers in the Ferguson-model, $\alpha = 0\ .05$, $N = 20$

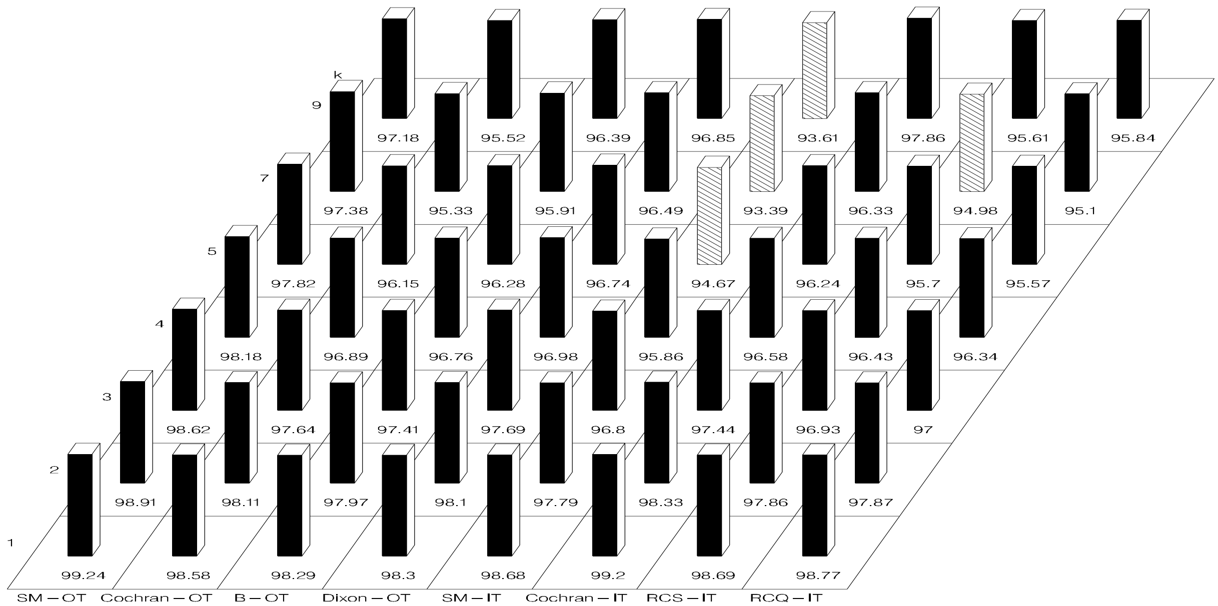

Figure 3: Fraction of correctly identified inliers in the Ferguson-model, $\alpha = 0\ .05$, $N\!= 20$

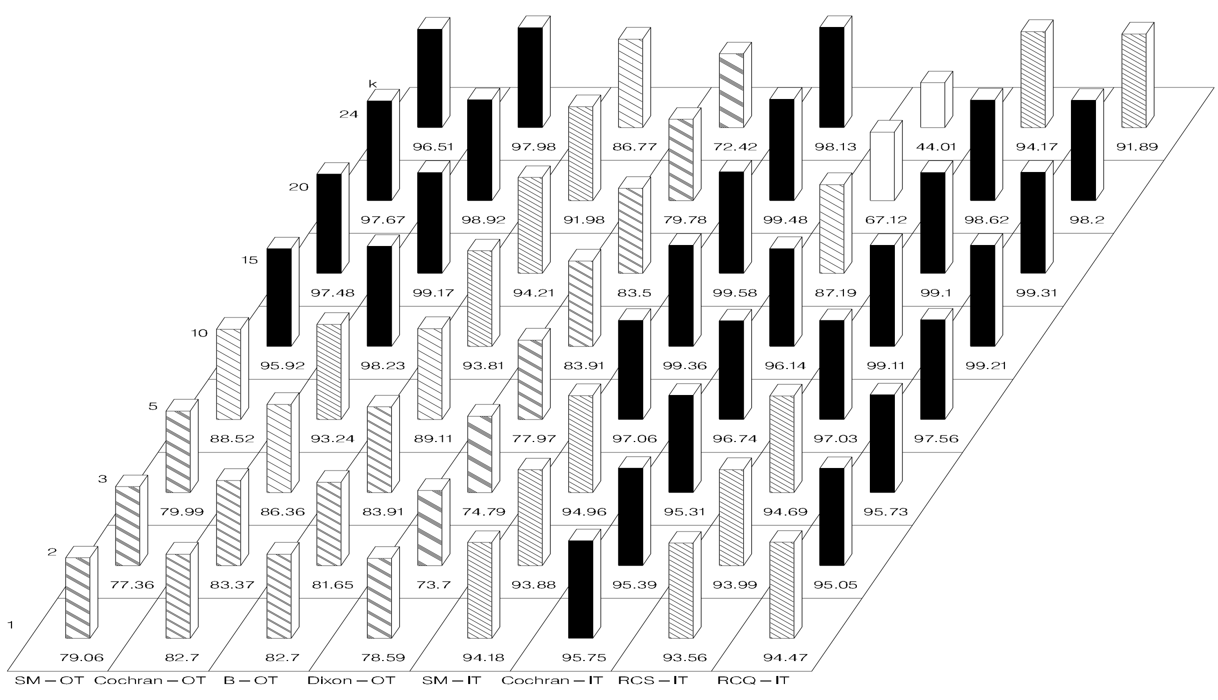

Figure 4: Fraction of correctly identified outliers in the Ferguson-model,
$\alpha = 0\ .05$, $N\!= 50$

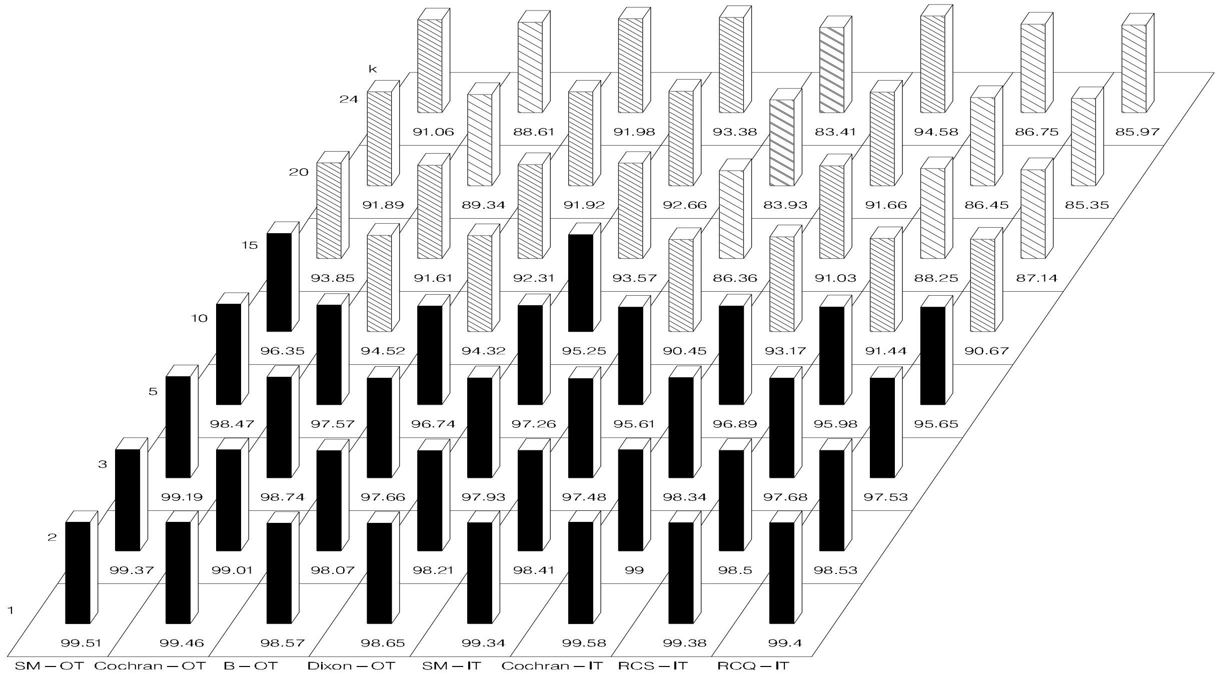

Figure 5: Fraction of correctly identified inliers in the Ferguson-model, $\alpha = 0\ .05$, N= 50

6 Example and conclusion

As an example for the application of the stepward outlier identification rules discussed in this paper, we consider a data set taken from Nelson (1982, p. 104). The data are the times to breakdown of an insulating fluid between two electrodes, recorded at a voltage of 32 kV. The recorded breakdown times in ascending order are 0.27, 0.40, 0.69, 0.79, 2.75, 3.91, 9.88, 13.95, 15.93, 27.80, 53.24, 82.85, 89.29, 100.58, and 215.10. We suppose that the breakdown times follow a one-parameter exponential distribution and seek to find out if the sample contains any α_N-outliers where $N = 15$ and $\tilde{\alpha} = 0\ .05$. Tables 1 and 2 contain a comparison of the results given by the eight identification rules discussed so far in this paper. For each rule the test statistics and corresponding critical values are listed up to the terminal step.

The results can be summarized as follows: The outward testing procedures based on spacings do not classify any observation as outlying whereas Cochran-OT identifies the five largest ones and SM-OT even one more. Inward testing with Cochran-statistics is less successful: Cochran-IT fails to identify any outlier. This result is not due to masking since also RCQ-IT declares no observation as outlying. RCS-IT identifies only the largest observation whereas the inward testing procedure based on the standardized median flaggs the maximal reasonable number $k^* = 7$ of observations as outliers. In this example we have assumed for purpose of demonstration that under the null modell the data come

i	$x_{(16-i)}$	SM-IT T_{16-i}^{SM}	SM-IT t_{16-i}^{SM}	Cochran-IT T_{16-i}^{ML}	Cochran-IT t_{16-i}^{ML}	RCS-IT T_{16-i}^{RCS}	RCS-IT t_{16-i}^{RCS}	RCQ-IT T_{16-i}^{RCQ}	RCQ-IT t_{16-i}^{RCQ}
1	215.10	10.6879	7.0437	5.2257	6.3146	9.2590	8.0307	5.5251	5.9838
2	100.58	5.8512	4.9587			4.7287	5.6065		
3	89.29	6.2643	4.3398						
4	82.85	8.3288	3.7570						
5	53.24	9.4381	3.6342						
6	27.80	5.7866	3.2748						
7	15.93	4.0152	3.3284						

Table 1: Inward testing procedures for the example

j	$x_{(8+j)}$	SM-OT T_{8+j}^{SM}	SM-OT t_{8+j}^{SM}	Cochran-OT T_{8+j}^{ML}	Cochran-OT t_{8+j}^{ML}	B-OT T_{8+j}^{B}	B-OT t_{8+j}^{B}	Dixon-OT T_{8+j}^{D}	Dixon-OT t_{8+j}^{D}
1	15.93	4.0152	5.4701	2.9518	3.5457	0.8510	6.7968	0.1243	0.5763
2	27.80	5.7866	5.0888	3.6402	3.6587	4.4466	6.5470	0.4270	0.5615
3	53.24			4.5185	3.8111	5.9061	6.3555	0.4778	0.5569
4	82.85					3.8031	6.2041	0.3574	0.5637
5	89.29					0.5029	6.0815	0.0721	0.5861
6	100.58					0.6111	5.9801	0.1122	0.9873
7	215.10					3.1880	5.8950	0.5324	0.7324

Table 2: Outward testing procedures for the example

from a one-parameter exponential distribution. The results obtained with Cochran-OT and both procedures based on the standardized median may however also be seen as indication that this assumption is questionable. It is indeed more properly to analyze the breakdown times under a Weibull model.

To come to a final conclusion: with respect to their breakdown properties there are many competing optimal stepwise procedures for outlier identification in exponential samples, among them the well known outward testing procedures based on discordancy tests with Cochran-, Dixon-, and Balasooryia-statistics. A finer worst-case analysis with respect to the size of the largest nonidentifiable outlier LO reveals that these classical procedures are outperformed by an outward testing procedure SM-OT that relies on a standardized version of the sample median. Especially the two procedures based on spacings perform very poorly whereas Cochran-OT is close second and seems also to have slightly better power than SM-OT in identifying outliers in non worst-case situations.

An alternative to these outward testing procedures is given by inward testing procedures that are based on robust estimators of the scale parameter. Whereas inward testing procedures with classical discordancy tests suffer heavily from masking and are therefore not recommendable, these new methods have optimal breakdown properties and lead to a smaller size of LO than competing inward testing procedures. A minor drawback is their tendency to identify too many observations as outlying if the null model does not hold. This tendency is due to the fact that the usual choice of critical values after the first step seems to be too liberal. How a better choice could be made is an interesting topic for further investigations.

Acknowledgement

Support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

Balasooriya, U. (1989). Detection of outliers in the exponential distribution based on prediction. Commun. Statist. – Theory Meth., 18, 711–720.

Balasooriya, U. and Gadag, V. (1994). Tests for upper outliers in the two-parameter exponential distribution. J. Statist. Comput. Simul., 50, 249–259.

Chikkagoudar, M. S. and Kunchur, S. H. (1987). Comparison of many outlier procedures for exponential samples. Commun. Statist. – Theory Meth., 16, 627–645.

Carey, V. J., Rosner, B. A., Wager, C. G. and Walters, E. E. (1997). Resistant and test-based outlier rejection: Effects on gaussian one- and two-sample inference. Technometrics, 39, 320–330.

Cochran, W. G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. Ann. Eugen., 11, 47–52.

Davies, L. and Gather, U. (1993). The identification of multiple outliers. J. Amer. Statist. Assoc., 88, 782–792.

Dixon, W. J. (1950). Analysis of extreme values. Ann. Math. Statist., 21, 488–506.

Gather, U. (1995). Outlier models and some related inferential issues, in: N. Balakrishnan and A.P. Basu (eds.): The exponential distribution: Theory, methods and applications, Gordon and Breach Publishers, Amsterdam, 221-239.

Gather, U. and V. Schultze (1999). Robust estimation of scale of an exponential distribution. Statistica Neerlandica, 53, 327–341.

Kabe, D. G. (1970). Testing outliers from an exponential population. Metrika, 15, 15–18.

Kimber, A. C. (1982). Tests for many outliers in a exponential sample. Applied Statistics, 31, 263–271.

Likeš, J. (1967). Distribution of Dixon's Statistics in the case of an exponential population. Metrika, 11, 46–54.

Likeš, J. (1987). Some tests for $k \geq 2$ upper outliers in a exponential sample. Biom. J. 29, 313–324.

Nelson, W. (1982). Applied life data analysis. John Wiley and Sons, New York.

Pawlitschko, J. (2000). On the distribution of a test statistic for outlier identification in exponential samples. Universität Dortmund, preprint.

Pearson, E. S. and Chandra Sekar, C. (1936). The efficiency of statistical tools and a criterion for the rejection of outlying observations. Biometrika, 28, 308–320.

Rosner, B. (1975). On the detection of many outliers. Technometrics, 17, 221–227.

Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute deviation. J. Amer. Statist. Assoc., 88, 1273–1283.

Schultze, V. and Pawlitschko, J. (2000). The identification of outliers in exponential samples. Statistica Neerlandica, to appear.

Sweeting, T. J. (1983). Independent scale-free spacings for the exponential and uniform distributions. Statist. Prob. Letters, 1, 115–119.

Tse, Y. K. and Balasooriya, U. (1991). Tests for multiple outliers in an exponential sample. Sankhyā Ser. B, 53, 56–63.